ARIC Manuscript Proposal #1376

PC Reviewed: 06/10/08 **Status: A** **Priority: 2**
SC Reviewed: _____ **Status: _____** **Priority: ____**

1. a. Full Title: Optimal predictors of incident hospitalized heart failure: the ARIC cohort study.

b. Abbreviated Title (Length 26 characters): HF risk score

2. Writing Group: Sunil Kumar Agarwal, Lloyd E. Chambless, A. Folsom, A. Bertoni, H. Ni, S. Russel, W. Rosamond, G. Heiss

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

First author: **Sunil Kumar Agarwal**
Address: Department of Epidemiology
School of Public Health
University of North Carolina at Chapel Hill
E-mail: sunilagarwal@unc.edu
Phone: (919) 265-4727 Fax: (919) 966-9800

Corresponding/senior author: Gerardo Heiss
E-mail: gerardo_heiss@unc.edu
Phone: (919) 962-3253 Fax: (919) 966-9800

3. Timeline: Approval of this manuscript by the ARIC Publications Committee will then enable work on this manuscript. Once started, this work will lead to manuscript(s) within 15 months.

4. Rationale: Heart Failure (HF) is a major public health problem in the economically developed nations(Garg, Packer et al. 1993). The population burden of HF in economically less developed nations is expected to increase. Despite improvements in medical management and advances in therapy, HF incidence and prevalence have increased unabated(Roger, Weston et al. 2004). HF also is the leading indication for hospitalization in the United States among patients older than 65 years. Hospital discharges rose from 399,000 in 1979 to 1,009,000 in 2004(Rosamond, Flegal et al. 2007). Treatment costs for HF exceed those for both coronary artery disease and cancer, requiring 5.4% of total health care cost(O'Connell and Bristow 1994). In 2001, $4.0 billion ($5912 per discharge) was paid to Medicare beneficiaries for CHF(Rosamond, Flegal et al. 2007).

Most patients with HF present for the first time to general practitioners and are mostly managed by them(Fowler 1997). Results of a qualitative study involving focus groups of 30 GPs in four primary care settings in the UK showed that lack of confidence in establishing diagnosis, rapidly changing knowledge and doubts about applicability of research to primary care settings were barriers to providing standard care to patients with HF in the community(Fuat, Hungin et al. 2003). In contrast, results of the SOLVD trial show the importance of early diagnosis and treatment in individuals with LV systolic dysfunction(Ahn, Jong et al. 2006). There is also evidence of misdiagnosis when HF is assessed by physicians using objective criteria(Remes,

Miettinen et al. 1991). Thus, accurate and early diagnosis of HF remains a cornerstone of patient management, which is known to be associated with multiple challenges in primary care settings.
It has been noted that most of the symptoms and signs commonly associated with HF are sensitive but non-specific, while the less prevalent or subtle ones occurring in moderate or severe disease are specific but less sensitive. A normal EKG is highly specific on the other hand, virtually ruling out HF or LV dysfunction(Rihal, Davis et al. 1995). Further, abnormal EKGs have been shown to have good sensitivity, i.e., 73-94% in a meta-analysis(Khunti, Squire et al. 2004) and 81% in the community based EPICA study(Fonseca, Mota et al. 2004). Thus, a combined use of patient's history, physical exam, basic laboratory investigations and EKG may provide an optimum tool for patient's risk stratification.

Only one published report has evaluated the utility of selected symptoms, signs and tests commonly available to the general practitioners in predicting HF(Kannel, D'Agostino et al. 1999). This report identified 486 heart failure cases during 38 years of follow-up using Framingham criteria. Similarly, little updated information is available on the most effective and parsimonious set of diagnostic elements readily available to the primary care physician to predict risk of incident HF for purposes of risk stratification, referral for cardiac imaging, or proactive intervention and scheduling. To this date information of this kind is completely lacking for African Americans, a population group that manifests a heavy burden of HF and its associated mortality.

Two recently published studies have examined the predictive value for incident hospitalized HF of ECG parameters(Rautaharju, Prineas et al. 2007) and kidney function(Kottgen, Russell et al. 2007), respectively Their results support the research proposed here since they show the usefulness of each of these parameters for HF risk stratification.

The goals of this proposal are to define the most parsimonious set of information readily available in the primary care setting that is optimally predictive of incident HF in African American and white middle-aged men and women, to compare their performance to that of the Framingham risk score for HF and to that of the Gothenburg score, and to examine whether a simple, updated HF risk score equation has merit in clinical and public health settings. We propose to perform these analyses in extant data from the Atherosclerosis Risk in Communities (ARIC) cohort, i.e., 15,972 men and women aged 45-64 years at baseline, drawn as a sample from four U.S. communities and followed from 1987 through 2007 for hospitalization and mortality attributed to HF.

Risk score functions that can provide general practitioners with a simple, reliable and cost effective tool for HF risk stratification of individuals without overt heart failure are lacking at present. If validated, such optimal risk score functions would be useful in several ways, as follows:

a). It would help to stratify individuals into risk categories before they develop overt HF requiring hospital admission. Such stratification may facilitate early recognition of asymptomatic individuals with ventricular dysfunction who are at a high risk of developing HF and sudden cardiac death. Early recognition may promote further investigations, referral to specialists, proactive intervention, and frequent follow-up appointments. If early diagnosis and treatment are facilitated by this tool, overt HF and initial hospitalization may be delayed.

b). An effective and user friendly risk calculator may promote adherence to guidelines among general practitioners.

c) Quantification of risk of HF may facilitate translation of epidemiologic research to clinical research and practice.

5. Main Hypothesis/Study Questions:

Aim 1: Identify a minimum set of variables that optimally predict risk of first hospitalized HF during 13 years of follow up for each gender and race subset.

> Derive the minimum set of variables from an individual's medical history, physical exam, biochemistry panel, an electrocardiogram (ECG) and co-morbid conditions to optimize the prediction of hospitalized incident HF over an average 17 years of follow up. The study variables are chosen on the basis of the literature and their availability to general practitioners.

Aim 2: Assess the predictive performance of the most predictive variables according to calibration, discrimination, and clinical usefulness by gender, race and comorbidity.

Aim 3: Estimate the risk score that optimally predicts the 5 and 10-year risks of HF by gender and race in the ARIC population.

Aim 4: Compare the performance and practical applicability of the ARIC HF risk score to that of the Framingham Heart Study HF risk score (in whites) and to that of the Gothenburg score by gender and for African Americans and whites.

6. Design and analysis (study design, inclusion/exclusion, outcome and other variables of interest with specific reference to the time of their collection, summary of data analysis, and any anticipated methodological limitations or challenges if present).

6.1 Data and Measurements

6.1.1 Outcome assessment: Incident hospitalized Heart Failure:

Incident HF will be defined as the first HF hospitalization or HF coded as the underlying cause of death and was identified through hospital records identified from discharge lists that showed an HF code in any position, and from death certificate codes. Hospitalizations were coded as heart failure (428) using the International Classification of Diseases Code, Ninth Revision (ICD-9), and deaths were coded as HF (428 and I50) using the ICD-9 and ICD-10. All cohort hospitalizations and deaths that occurred before January 1, 2003, will be included. Study participants with evidence of prevalent HF (n=752) at baseline will be excluded from analyses. Prevalent HF will be defined as the reported current intake of HF medication at the baseline examination (n = 83) or evidence of manifest HF as defined by the Gothenburg criteria stage 3 (n = 669), which require the presence of specific cardiac and pulmonary symptoms as well as medical treatment of HF(Eriksson, Caidahl et al. 1987). By January 1st, 2003, 1193 study participants met these incident HF criteria. Of these, 1187 (99.5%) were identified through a hospitalization. The overall incidence rate of HF was 6.1 per 1000 person-years. Ascertainment of fatal events in the cohort, which is virtually complete, indicates that 2079 (14.0%) study participants died by January 1st, 2003. Only 6 of these deaths were attributed to HF according to an underlying cause of death coded as ICD-9 428 or ICD-10 I50. Analyses for this study will include additional incident HF events (1987 through 2004), to be released by the ARIC Coordinating Center in March of 2008.

6.1.2 Baseline predictor variables:

The following predictor variables will be included in the analysis based on review of literature, their availability and meaningful interpretation by general practitioners in their daily practice. Standardized protocol and definitions were used for these variables.

Socio-demographic variables: age, gender, race, socio-economic status, smoking and alcohol excess.

Self reported morbid conditions: Coronary heart disease, angina, hypertension, type 2 diabetes, pulmonary disease (bronchitis, asthma).

Symptoms: Dyspnea on exertion, orthopnea, paroxysmal nocturnal dyspnea, fatigue, coughing and wheezing,

Physical exam parameters: Heart rate, blood pressure, pulse pressure, anthropometric variables (Body mass index, waist circumference, waist-hip-ratio), HDL cholesterol, LDL cholesterol, ankle edema, rales.

Basic lab parameters: estimated glomerular filtration rate based on the abbreviated Modification of Diet in Renal Disease (MDRD) Study equation), hemoglobin, blood glucose, serum LDL and HDL cholesterol, urine albumin.

EKG parameters: The 12 lead electrocardiograms measured at baseline, using standardized protocol and identical electrocardiograms at each of the four study site will used. ECGs were initially processed in a central laboratory at the EPICORE Center (University of Alberta, Edmonton, Alberta, Canada). A repeat processing was done recently using 2001 version of the GE Marquette 12-SL program (GE Marquette, Milwaukee, Wisconsin). Minnesota Code criteria were used in the classification of the parameters. A total of 20 variables representative of abnormalities of rhythm, atria, conduction, morphology and duration of waves, and ventricular hypertrophy, left ventricular strain will be included.

Current use of selected therapeutic agents: diuretics, b-blockers and ACE inhibitors, anti-arrhythmic.

Analyses

The data will be checked for completeness, plausibility and logical consistency of values by exploring univariate and bivariate distributions for each gender*race stratum, by outcome. The continuous ECG variables will be categorized according to established clinical relevance or convention. Analysis will be done using the following software packages: SAS 9.1.3, SAS Institute, Cary; S–Plus 8, Insightful Corporation, Seattle, WA, and CART, Salinger, Salford Systems', San Diego, CA.

Aim 1 - Identify a minimum set of variables that optimally predict risk of first hospitalized HF during 13 years of follow up for each gender and race subset.

We will examine the rate of HF over the course of follow up using Kaplan-Meier survival curves by race and sex. Univariate Cox regression models will be fitted and subsequently graphical and residuals statistics (Martingale and Schoenfeld (Sasieni and Winnett 2003)) will be used to test the proportional hazard assumption for each variable. We will then fit a parsimonious Cox proportional hazard regression model with selected variables that predict time to first hospitalized HF.The log-rank test will be used to examine equality across strata. Under the assumption of no violation of proportional hazards (PH) the variables that are significant at a 10% level will be retained in the multivariable Cox regression model. On violation of PH assumption, an appropriate extended Cox model will be fitted (i.e. by incorporating splines or interactions with time). Thus, these analyses will yield sex- and race-specific hazard ratio estimates with 90% confidence intervals, area under receiver operating (ROC) curve, sensitivity, specificity, positive predictive value (PPV) and negative predictive value (NPV). For continuous variables cut-points tha maximize predictive accuracy will be considered if threshold values are not provided by clinical guidelines.

A multivariable Cox regression model will be fitted for each gender*race subgroup. Clinically motivated two-way interaction terms will be examined. Since we have selected potential predictor variables on the basis of the extant literature and our best judgment regarding their availability to primary care physicians, statistical significance be de-emphasized in identifying the most parsimonious set of predictors while considering model fit and the discriminatory ability of the risk functions. As described in the literature, removal of variables to achieve parsimony will have two aims: a). gain maximum calibration without over-fitting the available data, and b). gain maximum discrimination (to give a score or classify individuals into appropriate risk groups). A threshold value of alpha = 0.1 (or chi-square > 2*number of variable removed) instead of traditional 0.05, may suit the purpose of building a model to risk stratify individuals better, as many marginal risks may still add up to predict high risk in a covariate stratum.

Calibration: Calibration refers to the agreement between predicted and actual outcome. To check the fit of data and also existence of a predictor variable combination with poor fit, we will use Hosmer and Lemeshow delta-chi square influence statistic (the decrease in the Pearson goodness of fit statistic that results from deleting the set of observations that share a specific covariate pattern) and the Pregibon delta-beta statistic, which results from the Pearson residual and the "hat matrix"(Hosmer and Lemeshow 1999). Further, the overall function's predicted risks will be used to divide the observations into deciles of predicted, compared to observed risk. Plots will be constructed showing predicted and actual event rates for each decile And a chi-square statistic will be used to compare the differences between predicted and actual event rates (with values exceeding 20 indicative of significant lack of calibration). Since these measures do not consider censoring we will explore the use of the integrated version of Brier score or logarithmic score to measure inaccuracy (Graf, Schmoor et al. 1999).

Discrimination: Discrimination is the ability of the predictive model to rank those who experienced outcome higher than those who didn't. A simple measure of this property is the proportion of HF cases in the highest decile of predicted risk, or the ratio of cases in the highest decile to lowest decile or lowest quintile of predicted risk. Other commonly used approaches are based on the change in area under the curve (AUC) of a ROC curve, or its proxy c-index for classification(Harrell, Lee et al. 1996). Methods for deriving time dependent AUC for risk prediction has been described in literature{Chambless, 2006 #204}. The c-index value represents an estimate of the probability that a model assigns a higher risk to those who develop HF within the follow-up period compared to those who do not. Since closely related, the c-index has limitations similar to those of the ROC curves(Graf, Schmoor et al. 1999),(Pencina, D' Agostino RB et al. 2008). Among them, the AUC is the full area estimated giving equal weights to all false positive rates. This does not take into account the shape of the ROC curve and thus neglects the clinical need of knowing the partial AUC under low false positive rates (such as 0-0.2)(McClish 1989) Other measures of separation like SPEP i.e. predicted probability of HF for the group with worst risk score – predicted probability of HF for the group with best risk score(Altman and Royston 2000), and SEP i.e., weighted geometric mean of absolute relative risk between a strata and baseline has been described(Sauerbrei, Hubner et al. 1997). Recent developments consider the net reclassification improvement for meaningful cut-offs and integrated discrimination improvement if no risk cut-off for decision making exists(Pencina, D' Agostino RB et al. 2008). We propose to use the partial change in AUC (false positive rate = 0 – 0.4) and a cut off of at least 0.005 improvement to retain a variable. Further, as part of the learning opportunities provided by this research plan, several of the methods outlined above will be explored for a methodological review and potential use in the in the analyses for Aim 1. Thus, using the above criteria, we expect to arrive at and report a risk score function for each gender

and race subgroup. For internal validation, v fold validation (dividing data into 10 sets, nine learning and one test during each iteration) technique will be used. Linearity assumptions of continuous and nominal predictor variables will be used using linear and quadratic splines (Greenland 1995), and incremental hazard, respectively. Finally, sensitivity analysis for time varying covariates will be included.

Validation of each model's calibration and discrimination ability will be based on bootstrapping (Harrell, Lee et al. 1996), as a means of internal validation. This permits to use the entire dataset for model development and allows for estimation of the error rates or for the reduction of bias of effect estimates. However, external validation of the optimal predictive model for each gender and race group in an independent test data set is a requisite before the use of such models can be recommended and measures of inaccuracy are known (Ripley 1996). External validation is Framingham risk function using ARIC cohort is described in aim 4. However, further attempts to validate the risk function derived in ARIC cohort using other NHLBI cohorts can be atempted in the future. We may also explore the derivation of a parisimonious set of diagnostic parameters predictive of HF with the use of Classification and Regression Tree (CART) analysis.

Aim 2 - Assess the performance of the most predictive variables according to calibration, discrimination, and clinical usefulness by gender, race and comorbidity.

Comparison of the risk functions for each gender and race group will consider differences in the variables selected, in the regression coefficients, and differences in the functional form of the risk model. Given the marked differences in the morbidity profile by gender and race, we will also consider whether comorbidity importantly modifies these risk models. Comparisons of goodness of fit, discrimination ability and calibration will follow the process outlined above. Clinical usefulness will consider the parsimony of the minimal set of predictive variables, sensitivity, specificity, positive and negative predictive values (PPV, NPV). Barring important differences between the optimal risk functions for each gender-race subset a common risk score will be developed as described above, for subsequent external validation (see Aim 4).

Various methods for comparison of discrimination and calibration have been described under Aim 1, above. Equality of hazard ratio (HR) will be compared using the regression coefficients for each variable derived for overall and each of gender*race subsets (total of five models). To compare these coefficients a test statistic z will be calculated, where z= (b [model 1] − b [model 2])/SE, and where b (model 1) and b (model 2) are, respectively, the regression coefficients of the model 1 (overall cohort or subset) and each of the other model, while SE is the standard error of the difference in the coefficients. This will computed as the square root of the sum of the squares of the SEs for the 2 coefficients and twice their covariance. Because the HR of a variable is computed by exponentiation its z statistic tests the equality of HRs across overall cohort model and each of the four model using subset analysis restricted to each of the gender*race combination. For each of the subset model 2 discrimination statistics such as c-index over a false positive range will be computed, one applying the overall cohort function to the subset and the other from the subset's own prediction function.

Aim 3 - Estimate the risk score that optimally predicts the 5 and 10-year risks of HF by gender and race in the ARIC population.

The gender and race-specific risk functions, or a common risk function if no important quantitative or qualitative differences are detected that would prevent this, will be derived as described. .From this, a simple point score system can be developed using the variables in the

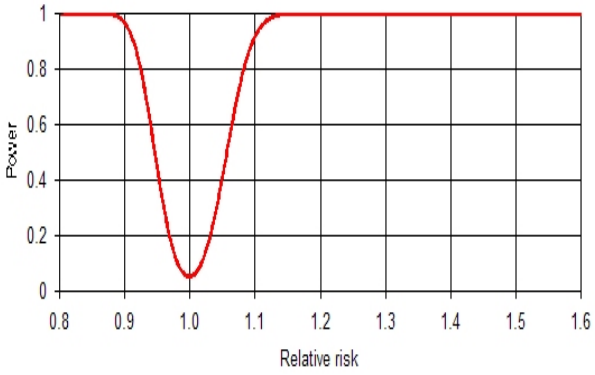
final risk function, as previously done for several risk scores such as coronary heart disease, stroke, diabetes, and others. Prior to this translation step however, external validation of the HF risk function based on the ARIC cohort is advisable.

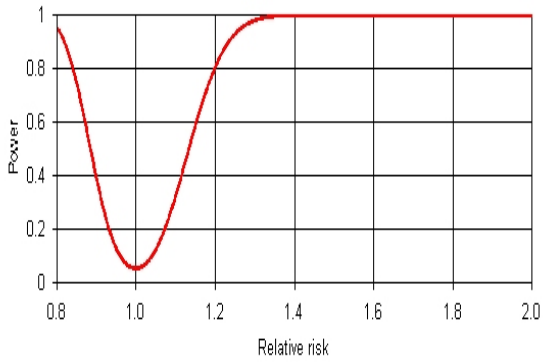
Aim 4 Compare the performance and practical applicability of the ARIC HF risk score to that of the Framingham HF risk score (in whites) and to that of the Gothenburg score.

A risk score function will be derived using the variables used for the model without vital capacity or x-ray film in Framingham study(Kannel, D'Agostino et al. 1999) for the whites subset in ARIC cohort. The coefficients for each variable will be compared to that of coefficients reported in the Framingham risk score function. This comparison will be made using the methods described in aim 2 above i.e. equality of risk ratio, discrimination, and calibration. Also, a comparison of discrimination and calibration between the optimal predictor function derived and validated for whites in ARIC will be made with Framingham small model for whites in ARIC.

Statistical Power

6.3.1 Cox Regression: Power analysis was done using a total sample size at baseline = 14857, follow up period = 13 years, alpha = 0.05 (two-sided), exposure proportion at baseline = 50% and median time to event among unexposed = 11 years (get conservative power). As we can see power reaches >0.8 to detect a hazard of 1.1. Similar, analysis was done by restricting the sample size at baseline to that of African American males (n=1600), other parameters being same. The power of >0.8 is achieved at HR>1.2. Also, the number of variables fitted in the multivariable model * 10 is less than number of HF(Rothman 2002), hence degrees of freedom available for calculation of coefficients should not be a problem. Hence, there seems to be enough power to identify important predictors using Cox regression. However, the power will be less for analysis restricted for HF events during 5 and 10 years of follow up. The above power analysis was done using PS software(Schoenfeld and Richter 1982).





Area under receiver-operator curve:
Minimum difference in AUC detectable at two values of power .

HF	No HF	AUC under null	Min Diff at power = 0.8	Min Diff at power 0.9
1200	13657	0.5	0.022	0.026
1200	13657	0.6	0.023	0.027
1200	13657	0.7	0.023	0.027
1200	13657	0.8	0.024	0.025
144	1632	0.5	0.062	0.064
144	1632	0.6	0.063	0.068
144	1632	0.7	0.062	0.076
144	1632	0.8	0.044	0.085

Above table shows the minimum detectable differences in area under curve on a two sided null hypothesis at alpha = 0.05 and beta = 0.2 and 0.1. The two study sample shown are for the overall cohort and for African American males (smallest subset sample). The estimates make the assumption of a correlation of 0.6 for both false positive and false negative on two risk functions, which is conservative as we expect high correlation. Further, we wish to test one sided hypothesis of whether inclusion of variable improves the AUC, hence these estimates might be conservative. Also, the two issues of partial AUC, and risk-cut off for decision making dictates power. Unlike, coronary artery disease, there are no risk cut-offs for action for HF. Hence, we were unable to currently estimate power for such analysis.

7.a. Will the data be used for non-CVD analysis in this manuscript? No
7.b. NA

8.a. Will the DNA data be used in this manuscript? No
8.b. NA

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.
No overlaps

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)?
Proposals looking at risk factors of HF:
MP#922 Alcohol consumption and risk of congestive heart failure
MP#927 Heart Failure Incidence and Survival: 13 Year Follow up of the ARIC Cohort
MP#1118 Kidney Function as a Risk Factor for Incident Heart Failure
MP#1125 Diabetes, obesity and insulin resistance as risk factors for incident hospitalized HF
MP#1144 The Obesity Paradox in Heart Failure.
MP#1160 Life Course Socioeconomic Exposures and Heart Failure.
MP# 1164 Hemoglobin A1c as a Risk Factor for HF Hospitalization among Persons with Diabetes.
MP#1197 Albuminuria as a Predictor of Incident Heart Failure Hospitalization and Mortality.
MP#1232 ECG Abnormalities Preceding Heart Failure: Estimation and Prediction
MP#1276 Exhaustion and risk for congestive heart failure.

Other Proposals with HF as focus:
MP#855 - Retinal Microvascular Abnormalities and Congestive Heart Failure
MP#617 - Evaluation of ICD Codes to Identify Hospitalized MI Patients with Acute Congestive HF.
MP#890B -Plasma Fatty Acid Composition and Incidence of Heart Failure in Middle Aged Adults
MP#1049 - Prevalence and Prognosis of ALVSD in African Americans
MP#1282 - Outpatient Surveillance of Heart Failure.
MP#1325 - Neighborhood and Individual Socioeconomic Status and Heart Failure Rehospitalization
MP#1325 - Socioeconomic, demographic and clinical predictors of heart failure care.

11. a. Is this manuscript proposal associated with any ARIC ancillary studies or use any ancillary study data? Yes

11.b. If yes, is the proposal
B. primarily based on ARIC data with ancillary data playing a minor role (usually control variables; list number(s)* _____ 2002.02 _____)

12. Manuscript preparation is expected to be completed in one to three years. The authors are aware of this fact.

References:

. "Manuals, Forms, and Data Dictionary, Atherosclerosis Risk in Community Study (ARIC) " Retrieved January 03 2008, from http://www.cscc.unc.edu/aric/pubuse/.

Ahn, S. A., P. Jong, et al. (2006). "Early versus delayed enalapril in patients with left ventricular systolic dysfunction: impact on morbidity and mortality 15 years after the SOLVD trial." J Am Coll Cardiol **47**(9): 1904-5.

Altman, D. G. and P. Royston (2000). "What do we mean by validating a prognostic model?" Stat Med **19**(4): 453-73.

Eriksson, H., K. Caidahl, et al. (1987). "Cardiac and pulmonary causes of dyspnoea--validation of a scoring test for clinical-epidemiological use: the Study of Men Born in 1913." Eur Heart J **8**(9): 1007-14.

Fonseca, C., T. Mota, et al. (2004). "The value of the electrocardiogram and chest X-ray for confirming or refuting a suspected diagnosis of heart failure in the community." Eur J Heart Fail **6**(6): 807-12, 821-2.

Fowler, P. B. (1997). "Evidence-based diagnosis." J Eval Clin Pract **3**(2): 153-9.

Fuat, A., A. P. Hungin, et al. (2003). "Barriers to accurate diagnosis and effective management of heart failure in primary care: qualitative study." Bmj **326**(7382): 196.

Garg, R., M. Packer, et al. (1993). "Heart failure in the 1990s: evolution of a major public health problem in cardiovascular medicine." J Am Coll Cardiol **22**(4 Suppl A): 3A-5A.

Graf, E., C. Schmoor, et al. (1999). "Assessment and comparison of prognostic classification schemes for survival data." Stat Med **18**(17-18): 2529-45.

Greenland, S. (1995). Dose-Response and Trend Analysis in Epidemiology: Alternatives to Categorical Analysis, JSTOR. **6:** 356-365.

Harrell, F. E., Jr., K. L. Lee, et al. (1996). "Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors." Stat Med **15**(4): 361-87.

Hosmer, D. W. and S. Lemeshow (1999). Applied survival analysis : regression modeling of time to event data. New York, Wiley.

Kannel, W. B., R. B. D'Agostino, et al. (1999). "Profile for estimating risk of heart failure." Arch Intern Med **159**(11): 1197-204.

Khunti, K., I. Squire, et al. (2004). "Accuracy of a 12-lead electrocardiogram in screening patients with suspected heart failure for open access echocardiography: a systematic review and meta-analysis." Eur J Heart Fail **6**(5): 571-6.

Kottgen, A., S. D. Russell, et al. (2007). "Reduced kidney function as a risk factor for incident heart failure: the atherosclerosis risk in communities (ARIC) study." J Am Soc Nephrol **18**(4): 1307-15.

McClish, D. K. (1989). "Analyzing a portion of the ROC curve." Med Decis Making **9**(3): 190-5.

O'Connell, J. B. and M. R. Bristow (1994). Economic impact of heart failure in the United States: time for a different approach. **13:** S107-12.

Pencina, M. J., S. D' Agostino RB, et al. (2008). "Evaluating the added predictive ability of a new marker: From area under the ROC curve to reclassification and beyond." Stat Med **27**(2): 157-72.

Rautaharju, P. M., R. J. Prineas, et al. (2007). "Electrocardiographic predictors of new-onset heart failure in men and in women free of coronary heart disease (from the Atherosclerosis in Communities [ARIC] Study)." Am J Cardiol **100**(9): 1437-41.

Remes, J., H. Miettinen, et al. (1991). "Validity of clinical diagnosis of heart failure in primary health care." Eur Heart J **12**(3): 315-21.

Rihal, C. S., K. B. Davis, et al. (1995). "The utility of clinical, electrocardiographic, and roentgenographic variables in the prediction of left ventricular function." Am J Cardiol **75**(4): 220-3.

Ripley, B. D. (1996). Pattern Recognition and Neural Networks, Cambridge University Press.

Roger, V. L., S. A. Weston, et al. (2004). "Trends in heart failure incidence and survival in a community-based population." Jama **292**(3): 344-50.
Rosamond, W., K. Flegal, et al. (2007). "Heart Disease and Stroke Statistics 2008 Update. A Report From the American Heart Association Statistics Committee and Stroke Statistics Subcommittee." Circulation.
Rothman, K. J. (2002). Epidemiology : an introduction. Oxford ; New York, Oxford University Press.
Sasieni, P. D. and A. Winnett (2003). Martingale difference residuals as a diagnostic tool for the Cox model, Biometrika Trust. **90:** 899-912.
Sauerbrei, W., K. Hubner, et al. (1997). "Validation of existing and development of new prognostic classification schemes in node negative breast cancer. German Breast Cancer Study Group." Breast Cancer Res Treat **42**(2): 149-63.
Schoenfeld, D. A. and J. R. Richter (1982). "Nomograms for calculating the number of patients needed for a clinical trial with survival as an endpoint." Biometrics **38**(1): 163-70.