Harvard University
Harvard University Biostatistics Working Paper Series

Year 2008 *Paper* 91

Evaluating Subject-level Incremental Values of New Markers for Risk Classification Rule

Tianxi Cai* Lu Tian†
Donald M. Lloyd-Jones‡ L. J. Wei**

*Harvard University, tcai@hsph.harvard.edu
†Northwestern University
‡Northwestern University
**Harvard University, wei@hsph.harvard.edu

EVALUATING SUBJECT-LEVEL INCREMENTAL VALUES OF NEW MARKERS FOR RISK CLASSIFICATION RULE

T. Cai[1], L. Tian[2], Donald M. Lloyd-Jones[2], and L.J. Wei[1]

[1]*Department of Biostatistics, Harvard University, Boston, MA 02115, USA*

[2]*Department of Preventive Medicine, Northwestern University, Chicago, IL 60611, USA*

Summary

Suppose that we need to classify a population of subjects into several well-defined ordered risk categories for disease prevention or management with their "baseline" risk factors/markers. In this article, we present a systematic approach to identify subjects using their conventional risk factors/markers, who would benefit from a new set of risk markers for more accurate classification. Specifically for each subgroup of individuals with the same conventional risk estimate, we present inference procedures for the reclassification and the corresponding correct re-categorization rates with the new markers. We then applied these new tools to analyze the data from the Cardiovascular Health Study sponsored by the US National Heart, Lung, and Blood Institute (NHLBI). We used Framingham risk factors plus the information of baseline anti-hypertensive drug usage to identify adult American women who may benefit from the measurement of a new blood biomarker, CRP, for better risk classification in order to intensify prevention of coronary heart disease for the next ten years.

Keywords: *Coronary Heart Disease, Nonparametric Functional Estimation, Risk Factors/Markers, Pointwise and Simultaneous Confidence Interval, Subgroup Analysis.*

1. INTRODUCTION

An integral part of evidence-based guidelines for disease prevention or management is a well-defined risk classification rule, which assigns each subject from a population of interest to one of several ordered risk groups. The assignment is based on the individual risk, the chance that the subject will experience a specific type of events during a given time period. Appropriate interventions will then be offered to subjects in each risk category. The risk is estimated using a scoring system with the subject's "conventional" baseline risk factors/markers. For example, in the United States more than a half million women die of cardiovascular diseases (CVD) each year and the majority of such deaths are due to coronary heart disease (CHD). Recently, the American Heart Association (AHA) issued guidelines for prevention of CHD in adult women (Mosca et al., 2004; 2007). Specifically, risk categories are defined by the individual's predicted risk of having a CHD event in the next ten years. Adult American women are classified as being low risk ($< 10\%$), intermediate risk (between 10% and 20%), or high risk ($\geq 20\%$). Such risk threshold values have also been employed by the Adult Treatment Panel III (ATP III) of the National Cholesterol Education Program to develop an evidence-based set of guidelines on cholesterol management (Grundy et al., 2004). For subjects with intermediate or high risk, certain life style and pharmacologic interventions are recommended. The risk is estimated using a modified version of the Framingham Risk Score (FRS), which is a multivariate risk equation based on traditional risk factors such as age, blood cholesterol, HDL cholesterol, blood pressure, smoking status and diabetes mellitus (Wilson et al., 1998).

Now, suppose that new risk markers for such future events are available, and they may potentially improve the conventional risk estimation. Measuring these markers, however, may be invasive or costly. Under various settings, novel procedures have been proposed to quantify the *overall* incremental benefit from the new markers for the entire population of interest (Pepe et al., 2004; Cook, 2007; Tian et al., 2007; Uno et al., 2007; Pepe et al., 2008; Pencina et al., 2008). In a recent paper, Wang et al. (2006) concluded that almost all new contemporary biomarkers for prevention of CHD added rather moderate *overall* predictive value to FRS.

In general, for subjects whose conventional risk estimates are either low or high, clinical practitioners would not require new markers for their decision making. Unfortunately, rather little research has been done for developing a systematic procedure to identify a subset of individuals who would benefit significantly from measurement new risk markers (D'Agostino, 2006). With respect to the aforementioned risk classification rule recommended by the AHA and ATP III, recently Cook et al. (2006) and Ridker et al. (2007) examined the incremental value of a new blood biomarker, the C-Reactive protein (CRP) measured by a highly sensitive assay, with a cohort of subjects from the Women's Health Study (Buring & Hennekens, 1992). They concluded that subjects whose traditional risk scores are between 5% and 20% would benefit from this marker. Their claims, however, were based on the reclassification rates with CRP among four highly discretized subgroups of the entire cohort based on conventional risk estimates. Moreover, it is not clear how to quantify the gain of reclassification solely based on the observed event rates for the risk categories.

In this article, for each subject in the study we construct two individual risk estimates, one based on the conventional risk markers and the other based on the conventional and new markers. A subject may be reclassified into a different risk category with the new risk estimate. For the subset $\mathcal{D}$ of subjects with the same conventional risk estimate, we obtain consistent nonparametric functional estimates for the reclassification rates and their corresponding standard error estimates. Note that large observed reclassification rates do not automatically imply that the new markers are important for subjects in $\mathcal{D}$. Consistent estimates for the proportions of subjects who are reclassified correctly are also needed for cost-benefit decision making. In this paper, we develop procedures for making inference about such proportions. Furthermore, we propose a test procedure to examine whether reclassifications by the new markers are better than "random allocations" of subjects in $\mathcal{D}$.

As an illustration, consider the set $\mathcal{D}$ consisting of individuals who have the same conventional FRS estimate of 9% for experiencing CHD events within ten years. Based on the aforementioned prevention guidelines, individuals with this predicted risk would not be recommended for lipid-

lowering drug therapy at present. Now, suppose that with the new markers, the estimate for the proportion of subjects in $\mathcal{D}$ reclassified to the next higher risk category is 25%, and the corresponding rate estimate for subjects in $\mathcal{D}$, who are reclassified and had CHD events, is 3%. The potential benefit with the new markers would be preventing 3% of subjects in $\mathcal{D}$ from having future CHD events. However, the costs consist of measuring new marker values for every subject in this subset, and giving possibly long-term, potentially toxic interventions to 25% of the individuals. Even if we decide that for this scenario, the benefit outweighs the cost, we still need to know whether this reclassification scheme is better than a "random allocation" process. If we randomly move 25% of subjects in $\mathcal{D}$ to the next higher risk class, on average, 9% of these subjects would have CHD events. The question is whether the observed 12% (3% out of 25%) is significantly different from this null value of 9%.

The new proposal is illustrated with a data set from the Cardiovascular Health Study (CHS) sponsored by the US National Heart, Lung and Blood Institute (Fried et al., 1991). This study is a prospective, population-based, long term follow-up cohort study to determine risk factors for predicting coronary future heart disease and stroke in adults 65 years or older. There were 5888 study subjects recruited between 1988 and 1993. For our analysis, we only considered data from 3393 female participants. For each subject, we utilized the risk factors/markers values at her entry time to the study and her CHD event time by year 2003. The binary response variable is whether the subject had experienced a CHD event (non-fatal MI, angina or CHD-related death) during ten year follow-up. For this data set, the median age at the baseline is 72.5 and there is no loss of follow-up for these endpoints. During the first 10 year follow-up, 19.5% of female participants experienced non-fatal MI, angina or CHD-related death. On the other hand, 37% of them died from non-CHD-related diseases. Among 3393 subjects, 52% of them survived by year 2003. In our analysis, the conventional risk factors consist of the Framingham risk factors and the anti-hypertensive drug usage. The new marker considered here is CRP. For the risk classification rule recommended by the ATP III, the CRP provides *pointwise* significant incremental values for subjects whose conventional risk estimates are around 10% and 20% . On the other hand,

we cannot find any subgroup of individuals who would benefit from the new marker under the *simultaneous* inference setting when controlling the overall family-wise type I error rate of 0.05.

In the next section, we describe our proposed procedures for quantifying the incremental value of new markers. Procedures for making inference about the incremental values were proposed in Section 3. The results from analyzing the CHS data are detailed in Section 4. Concluding remarks are given in Section 5.

2. QUANTIFYING INCREMENTAL VALUES FROM NEW MARKERS

Consider a subject randomly drawn from the study population. Let Y be its binary response variable, U be the set of conventional markers and V be the set of new markers. Let $p(U) = \text{pr}(Y = 1|\ U)$ and $p(U,V) = \text{pr}(Y = 1|\ U,V)$, the risks of this subject conditional on U and $\{U,V\}$, respectively. Assume that the classification rule assigns each subject to one of L ordered risk categories $\{C_1, \cdots, C_L\}$. A subject is classified to C_l, if its risk for $Y = 1$ is in the interval $[\nu_{l-1}, \nu_l)$, where $l = 1, \cdots, L$, and $0 = \nu_0 < \cdots < \nu_L = 1$. Based on $p(U,V)$, the subject may be reclassified into a higher or lower risk category.

Now, suppose that the data $\{(Y_i, U_i, V_i), i = 1, \cdots, n\}$ consist of n independent copies of (Y,U,V). In theory, for any given (u,v), the probabilities $p(u)$ and $p(u,v)$ may be consistently estimated nonparametrically. In practice, however, such nonparametric functional estimates do not behave well even when the dimension of U or V is not small. A practical alternative is to consider a working model for approximating $p(U)$ by a parametric model

$$g_1(\beta' X), \tag{2.1}$$

where X, a $p \times 1$ vector, is a function of U, β is unknown vector of parameters and $g_1(\cdot)$ is a known, smooth, increasing function. Now, suppose that β is estimated by $\widehat{\beta}$ via an estimating function $S_1(\beta)$. The risk for a subject with $U = u$, whose $X = x$, is estimated by $\widehat{p}_u = g_1(\widehat{\beta}' x)$.

With the additional marker set V, consider the working model for approximating $p(U,V)$ by

$$g_2(\gamma' W), \tag{2.2}$$

where W, a $r \times 1$ vector, is a function of U and V, $g_2(\cdot)$ is a known, smooth, increasing function, and γ is a vector of unknown parameters. Let γ be estimated by $\widehat{\gamma}$ via an estimating function $S_2(\gamma)$. For a subject with $(U, V) = (u, v)$ whose $W = w$, the estimated risk is $\widehat{p}_{\{u,v\}} = g_2(\widehat{\gamma}'w)$. Note that when Models (2.1) and (2.2) are correctly specified, with reasonable estimating functions $S_1(\cdot)$ and $S_2(\cdot)$, $\widehat{p}_u$ and $\widehat{p}_{\{u,v\}}$ are consistent estimators for $p(u)$ and $p(u, v)$, respectively.

Now, consider a random future subject with $(Y, U, V) = (Y^0, U^0, V^0)$. For a given $0 < s < 1$, let $\mathcal{D}_s$ be the group of future subjects whose $\widehat{p}_{U^0} = s$. For subjects in $\mathcal{D}_s$ with additional new marker information, the probability of a subject being classified to C_l with $Y^0 = q$, is

$$\eta_l^{(q)}(s) = \text{pr}\left\{ \widehat{p}_{\{U^0,V^0\}} \in [\nu_{l-1}, \nu_l), Y^0 = q \;\middle|\; \widehat{p}_{U^0} = s \right\}, \quad \text{for } q = 0, 1; l = 1, ..., L. \tag{2.3}$$

Here, the conditional probabilities are with respect to the data $\{(Y_i, U_i, V_i)\}$, Y^0, U^0 and V^0. The probability of a subject being classified to C_l is $\eta_l(s) = \eta_l^{(0)}(s) + \eta_l^{(1)}(s)$. It is important to note that large reclassification rates $\eta_l(s)$ do not automatically imply that the new markers are valuable for $\mathcal{D}_s$. The (2.3) also plays an important role in cost-benefit decision making. For subjects who are re-assigned to a higher risk category, $\{\eta_l^{(1)}(s), l > l_s\}$ provides us the proportion of subjects in $\mathcal{D}_s$ who would benefit from the new markers, where $l_s = \sum_{l'=1}^{L} I(\nu_{l'} \leq s)$. On the other hand, for subjects who are moved down to a lower risk group, $\{\eta_l^{(0)}(s), l < l_s\}$ would reflect the incremental gain.

Moreover, it is important to know whether the above reclassification is more than a purely random allocation process. Specifically, if the new markers contribute nothing to the classification rule, one would expect that conditional on $\widehat{p}_{U^0} = s$, $\widehat{p}_{\{U^0,V^0\}}$ is independent of Y^0. That is,

$$\eta_l^{(1)}(s) = \eta_l(s) \times \xi(s), \quad l = 1 \cdots, L, \tag{2.4}$$

where $\xi(s) = \text{pr}(Y^0 = 1 | \ \widehat{p}_{U^0} = s) = \sum_{l=1}^{L} \eta_l^{(1)}(s)$. The differences

$$\{D_l(s) = \eta_l^{(1)}(s) - \eta_l(s)\xi(s), \ l = 1, ..., L\} \tag{2.5}$$

also play important roles for quantifying the incremental value of the new markers for future subjects in $\mathcal{D}_s$. In fact, if the ratios, $\{\eta_l^{(1)}(s)/\eta_l(s)\}$, increase in $l, l = 1, \cdots, L$, this indicates that the markers indeed provide extra information about risk classifications of these subjects.

3. ESTIMATING $\eta_l^{(q)}(s)$ WITH POSSIBLY CENSORED EVENT TIME OBSERVATIONS

Oftentimes the binary response Y indicates that either the event time of interest is greater or less than a specific time point via a long term follow-up study. To this end, Let T_0 be the random event time and t_0 be a prespecified time point. Here, the binary variable $Y = 1$, if $T_0 < t_0$; 0, otherwise. The "response" Y may not be observed directly due to censoring. That is, T_0 may be censored by a random variable C. Let $G(\cdot)$ be the survival function of C. In this article, we assume that $G(\cdot)$ is independent of T_0, U and V. For T_0, one can only observe $T = \min(T_0, C)$ and $\Delta = I(T_0 \leq C)$, where $I(\cdot)$ is the indicator function. Our data $\{(T_i, \Delta_i, U_i, V_i), i = 1, \cdots, n\}$ consist of n independent copies of (T, Δ, U, V). Note that if there is no censoring involved, $G(\cdot) = 1$ and $\{Y_i, i = 1, \cdots, n\}$ can be observed completely.

To obtain the estimates $\widehat{p}_u$ and $\widehat{p}_{\{u,v\}}$, one may assume proportional hazards *working* models for T_0 with U, and then for T_0 with U and V. With the maximum partial likelihood estimates for the regression coefficients and the estimates for the underlying cumulative hazard functions, we can estimate $p(u)$ and $p(u, v)$. However, if a working proportional model is not correctly specified, the regression coefficient estimator converge to a constant vector, as $n \to \infty$, which may depend on the censoring distribution. Moreover, a good prediction model for short term survivors may not work well for predicting long term survivors. In this article, instead of modeling the entire hazard function of T_0, we use the approach taken by Uno et al. (2007) to model the conditional risks of experiencing an event by t_0 directly via (2.1) and (2.2).

To estimate β in (2.1), we let the estimating function $S_1(\beta)$ be

$$\sum_{i=1}^{n} \widehat{\omega}_i X_i \left\{ I(T_i < t_0) - g_1(\beta' X_i) \right\} = \sum_{i=1}^{n} \widehat{\omega}_i X_i \left\{ Y_i - g_1(\beta' X_i) \right\}, \tag{3.1}$$

where the weighting $\widehat{\omega}_i = \Delta_i I(T_i < t_0)/\widehat{G}(T_i) + I(T_i \geq t_0)/\widehat{G}(t_0)$, and $\widehat{G}(\cdot)$ is the Kaplan-Meier estimator of $G(\cdot)$ Here, the weighting takes care of the problem due to censoring and $I(T_i < t_0) = Y_i$ when $\widehat{\omega}_i \neq 0$. It is shown in Uno et al. (2007) that as $n \to \infty$, the resulting estimator $\widehat{\beta}$ converges to a constant which is free of $G(\cdot)$ even when the model (2.1) is misspecified.

Similarly, with additional V and Model (2.2), one can use the estimating function

$$S_2(\gamma) = \sum_{i=1}^{n} \widehat{\omega}_i W_i \{Y_i - g_2(\gamma' W_i)\} \tag{3.2}$$

to estimate γ.

Now, since $\eta_l^{(q)}(s)$ is between 0 and 1, we use a non-parametric local logistic likelihood estimation procedure to obtain consistent estimator $\widehat{\eta}_l^{(q)}(s)$. Specifically, first consider a kernel-type nonparametric functional estimator based on the local likelihood "score" function for the standard logistic regression which relates the binary response $I\{\widehat{p}_{\{U_i,V_i\}} \in [\nu_{l-1}, \nu_l), Y_i = q\}$ to the regressor $\widehat{\mathcal{E}}_i(s) = \psi(\widehat{p}_{U_i}) - \psi(s)$. Here, we choose a proper transformation $\psi(\widehat{p}_U)$ of $\widehat{p}_U$ to implement the smoothing, where $\psi(\cdot)$ is a known, non-decreasing function (Wand et al., 1991; Park et al., 1997). For any given s, l and q, this results in a score function of the intercept parameter a and slope parameter b:

$$\sum_{i=1}^{n} \begin{bmatrix} 1 \\ h^{-1}\widehat{\mathcal{E}}_i(s) \end{bmatrix} K_h\{\widehat{\mathcal{E}}_i(s)\}\widehat{\omega}_i \Big[I\big\{\widehat{p}_{\{U_i,V_i\}} \in [\nu_{l-1}, \nu_l), Y_i = q\big\} - g_0\big\{a + b\widehat{\mathcal{E}}_i(s)\big\}\Big], \tag{3.3}$$

where $g_0(x) = \exp(x)/\{1+\exp(x)\}$, $K_h(x) = K(x/h)/h$, $K(\cdot)$ is a known smooth symmetric kernel density function with a bounded support, and the bandwidth $h > 0$ is assumed to be $O(n^{-v})$, for $1/5 < v < 1/2$. Let $\widehat{\eta}_l^{(q)} = g_0(\widehat{a})$, where $\widehat{a}$ is the resulting estimator of the intercept by solving the equations: (3.3)=0, for $q = 0, 1$ and $l = 1, \cdots, L$. Then $\eta_l(s)$ and $\xi(s)$ can be estimated by $\widehat{\eta}_l(s) = \widehat{\eta}_l^{(1)}(s) + \widehat{\eta}_l^{(0)}(s)$ and $\widehat{\xi}(s) = \sum_{l=1}^{L} \widehat{\eta}_l^{(1)}(s)$, respectively.

In Appendix A, we show that $\widehat{\eta}_l^{(q)}(s) - \eta_l^{(q)}(s) \to 0$ in probability, uniformly in $s \in \Omega_h \equiv [\psi^{-1}(\rho_l + h), \psi^{-1}(\rho_r - h)]$, where $[\rho_l, \rho_r]$ is a subset of the support of $\psi\{g_1(\beta_0' X)\}$ and β_0 is the limit of $\widehat{\beta}$. Furthermore, we show in Appendix B that for large n, the joint distribution of

$$\{(nh)^{1/2}\{\widehat{\eta}_l^{(q)}(s) - \eta_l^{(q)}(s)\}, q = 0, 1; l = 1, \cdots, L\}$$

can be approximated by the conditional distribution of a mean-zero normal vector

$$\left\{ n^{-1/2} h^{1/2} \sum_{i=1}^{n} K_h\{\widehat{\mathcal{E}}_i(s)\}\widehat{\mathcal{V}}_{li}^{(q)}(s) Z_i,\ q = 0, 1; l = 1, \cdots, L \right\}, \tag{3.4}$$

given the data, where

$$\widehat{\mathcal{V}}_{li}^{(q)}(s) = \frac{\widehat{\omega}_i[I\{\widehat{p}_{\{U_i,V_i\}} \in [\nu_{l-1}, \nu_l), Y_i = q\} - \widehat{\eta}_l^{(q)}(s)]}{\widehat{\tau}\{\psi(\widehat{p}_{U_i})\}},$$

$\widehat{\tau}(s) = \sum_{i=1}^{n} K_h\{\widehat{\mathcal{E}}_i(s)\}/n$ is the estimated density function of $\psi(\widehat{p}_U)$ and $\{Z_i, i = 1, \cdots, n\}$ is a random sample from the standard normal variable and is independent of the data. Note that the $\{Z_i, i = 1, \cdots, n\}$, are the only random quantities in (3.4), whose distribution can be approximated easily by simulating $\{Z_i, i = 1, \cdots, n\}$ repeatedly. Confidence interval estimates for $\{\eta_l^{(q)}(s)\}$ and $\{\eta_l(s)\}$ can be constructed via this large sample approximation.

Lastly, we need to examine whether the above reclassifications are better than realizations from a random allocation scheme. To this end, we test the hypothesis that $D_l(s) = 0, l = 1, \cdots, L$, defined in (2.5). Since we are interested in an alternative hypothesis that the ratios $\{\eta_l^{(1)}(s)/\eta_l(s), l = 1, ..., L\}$, are non-decreasing in l, we consider a "trend" test statistic

$$\widehat{D}(s) = \sum_{l=1}^{L} \tilde{\mathfrak{w}}_l \widehat{D}_l(s), \tag{3.5}$$

where $\tilde{\mathfrak{w}}_l$ is a set of pre-specified non-negative, increasing constants (Cochran, 1954; Armitage, 1955) and $\widehat{D}_l(s) = \widehat{\eta}_l^{(1)}(s) - \widehat{\eta}_l(s) \times \widehat{\xi}(s)$, for $l = 1, \cdots, L$. Without loss of generality, we let $\sum_{l=1}^{L} \tilde{\mathfrak{w}}_l = 1$. Under the null hypothesis, $(nh)^{1/2}\widehat{D}(s)$ is asymptotically equivalent to

$$(nh)^{1/2} \sum_{l=1}^{L} \tilde{\mathfrak{w}}_l \Bigg[\widehat{\eta}_l^{(1)}(s) - \eta_l^{(1)}(s) - \sum_{q=0}^{1} \left\{ \widehat{\eta}_l^{(q)}(s) - \eta_l^{(q)}(s) \right\} \widehat{\xi}(s) - \sum_{l'=1}^{L} \left\{ \widehat{\eta}_{l'}^{(1)}(s) - \eta_{l'}^{(1)}(s) \right\} \widehat{\eta}_l(s) \Bigg]. \tag{3.6}$$

The null distribution of $\widehat{D}(s)$ can be obtained by approximating all $(nh)^{1/2}\{\widehat{\eta}_l^{(q)}(s) - \eta_l^{(q)}(s)\}$ in (3.6) via (3.4). A large observed $\widehat{D}(s)$ suggests a rejection of the null hypothesis. Note that the choice of the cutoff point for the above test is based on controlling Type I error rate at this specific value s.

To control the overall *family-wise* Type I error rate for $s \in \Omega_h$ under the simultaneous inference framework, one may consider the test statistic $\widehat{\mathcal{S}} = \sup_{s \in \Omega_h} \widehat{D}(s)$. Unfortunately,

$\{\widehat{D}(s), s \in \Omega_h\}$ does not converge in distribution as a process and $\widehat{\mathcal{S}} \to \infty$ in probability as $n \to \infty$. Therefore, we cannot use the standard large sample theory for stochastic processes to obtain a finite sample approximation to the distribution of $\widehat{\mathcal{S}}$. On the other hand, by the strong approximation arguments and extreme value limit theorem (Bickel & Rosenblatt, 1973), in Appendix C, we show that under the null hypothesis, a standardized version of $\widehat{\mathcal{S}}$ converges in distribution to a proper random variable. In practice, for finite sample size n, one can approximate the null distribution of $\widehat{\mathcal{S}}$ by approximating $(nh)^{1/2}\{\widehat{\eta}_l^{(q)}(s) - \eta_l^{(q)}(s)\}$ in (3.6) with (3.4) as we did for the above pointwise large sample approximation, but using the same set of perturbation variables $\{Z_i, i = 1, \cdots, n\}$ for all $s \in \Omega_h$. For each realized set of $\{Z_i\}$, we obtain a realization of the approximation to the distribution of $\widehat{\mathcal{S}}$. The one-sided p value can then be calculated via the observed $\widehat{\mathcal{S}}$ and the empirical distribution of the above realizations obtained by generating the set of $\{Z_i\}$ repeatedly.

As for any nonparametric functional estimation problem, the choice of the smooth parameter h is crucial for making inferences about $\eta_l^{(q)}(s)$, $\eta_l(s)$ and $D_l(s)$. Here, we propose to use the standard K-fold cross validation procedure to obtain an "optimal" h. Specifically, we randomly split the data into K disjoint subsets of about equal sizes denoted by $\{\mathcal{J}_k, k = 1, \cdots, K\}$. For each k, we use all observations which are not in $\mathcal{J}_k$ to estimate $p(u)$ and $p(u, v)$ by fitting the working models (2.1) and (2.2) and then with a given h, to estimate $\eta_l^{(q)}(s)$. Let the resulting estimators be denoted by $\widehat{p}_{(k)u}$, $\widehat{p}_{(k)\{u,v\}}$ and $\widehat{\eta}_{l(k)}^{(q)}(s)$. We then use the observations from $\mathcal{J}_k$ to calculate the prediction error

$$\mathrm{PE}^{(q)}(k, l) = \sum_{j \in \mathcal{J}_k} \widehat{\omega}_j \left[I\{\widehat{p}_{(k)\{U_j, V_j\}} \in [\nu_{l-1}, \nu_l), Y_j = q\} - \widehat{\eta}_{l(k)}^{(q)}\{\widehat{p}_{(k)U_j}\} \right]^2. \tag{3.7}$$

Lastly we sum (3.7) over $k = 1, \cdots, K$, and then choose h by minimizing this sum of K prediction errors. Note that the optimal smooth parameter value may only work for this specific (l, q) of $\widehat{\eta}_l^{(q)}(s)$. Alternatively, one may obtain a uniform bandwidth, which minimizes the sum of (3.7) over k, but also over $q = 0, 1$, and $l = 1, \cdots, L$. The order of such a bandwidth estimator is expected to be $n^{-1/5}$ (Wand & Jones, 1995; Fan & Gijbels, 1996) and thus the final bandwidth for estimation

can be obtained by multiplying the estimated bandwidth by n^{-d_0}, for some $d_0 \in (0, 3/10)$.

4. EXAMPLE

We apply the proposed procedures to analyze the data from female participants in the Cardiovascular Health Study (CHS) with respect to the risk classification rule recommended by AHS. There are two different ways to define the binary response variable Y. First, a study subject died before her ten-year follow-up from a non-CHD cause, we let $Y = 0$ (no CHD event). If the subject had experienced a CHD event (non-fatal MI, angina or CHD-related death), let $Y = 1$. There are no loss-to-follow-ups for these endpoints, therefore, we observe all Y's in this analysis. For the second analysis, we let the time to death of other causes be an independent censoring variable C for the time to the CHD event.

For both analyses, we consider an additive model (2.1) with $g_1(\cdot)$ being the anti-logit function. The vector X consists of the usage of hypertensive medication and all variables used in the model for deriving the FRS given in Wilson et al (1998). Specifically, the FRS model includes various dummy variables indicating levels of blood pressure, total cholesterol, and HDL, as well as age, age^2, present smoking status and diabetic status. Next, we fit the data using an additive model (2.2) with $g_2(\cdot)$ being the anti-logit function and W being the above risk factors/markers X and the log-transformed CRP variable. For estimating functions (3.1) and (3.2), the estimator $\widehat{G}(\cdot)$ is the standard K-M estimator with all death and CHD event times as censored observation for censoring variable C. For the nonparametric function estimation, we let $K(\cdot)$ be the Epanechnikov kernel, and $\psi(\cdot) = g_1^{-1}(\cdot)$. The smoothing parameter $h = 0.13$ was obtained through the 10-fold cross validation scheme. Specifically, we let h be the minimizer of $\sum_{l=1}^{3}\sum_{k=1}^{10}\sum_{q=0}^{1} \mathrm{PE}^{(q)}(k, l)$ in (3.7) for both $\widehat{\eta}_l^{(0)}(s)$ and $\widehat{\eta}_l^{(1)}(s)$ multiplied by a normalizing constant $n^{-1/20}$. To approximate the distributions of these estimators, we used the resampling method (3.4) with 500 independent realized samples of $\{Z_i, i = 1, \cdots, n\}$.

For the case that there is no censoring involved, we present the results of our analysis in Figure 1. Part (a) of the Figure is a smoothed density estimate for $\widehat{p}_U$, which provides useful

information regarding the relative size of the subgroup $\mathcal{D}_s$ of subjects such that $\widehat{p}_{U^0} = s$. Here, we choose the 1th and 99th percentiles of the empirical distribution of $\{\widehat{\beta}' X_i, i = 1, \cdots, n\}$ as the boundary points ρ_l and ρ_r, respectively. Therefore, the results presented here are for $\widehat{p}_U = s \in [0.074, 0.46]$. Part (b) of the Figure gives the estimated re-classification rates over $\widehat{p}_U$. The green, blue and red curves are for the risk categories C_1, C_2 and C_3, respectively. It appears that there are substantial re-allocations with the CRP around the boundary points for the risk classification rule. The solid green curve in Part (c) gives the estimates $\widehat{\eta}_1^{(1)}(s)$. Each estimate is the proportion of subjects in $\mathcal{D}_s$, who were reclassified to C_1 and had CHD events. The dotted green curve is the corresponding set of their "expected" estimates $\widehat{\eta}_l(s) \times \widehat{\xi}(s)$ when the re-categorizations are simply generated via a random allocation process. It appears that the differences between these two curves are rather small. Parts (d) and (e) of the Figure are the corresponding estimates for the risk categories C_2 and C_3. Part (f) is the curve of the p-values based on the test statistic $\widehat{D}(s)$ in (3.5) with weight $\tilde{\mathfrak{w}}_l = l/6$. Part (f) suggests that from the statistical point of view the subjects with conventional risk estimates $\widehat{p}_U \in [0.099, 0.107]$ and $\widehat{p}_U \in [0.161, 0.215]$ may benefit from having CRP measurement at the (pointwise) significance level of 0.05. For simultaneous inference, in Part (f) we also present the p-value curve using $\widehat{\mathcal{S}}$ proposed in Section 3. With an overall Type I error rate of 0.05, there is no subgroup of individuals who would benefit from the additional CRP information. That is, we cannot claim that the reclassifications to risk categories by the new marker are not generated from a random allocation process.

In Table 1(a), we present the results in details with $s = 0.11, 0.18, 0.21$ at which the pointwise p-values based on the test $\widehat{D}(s)$ are less than 0.05. Specifically, for each s, we provide the estimates $\widehat{\eta}_l(s)$, $\widehat{\eta}_l^{(1)}(s)$ and $\widehat{\eta}_l^{(0)}(s)$, along with their sampling standard errors. For example, the subjects with $\widehat{p}_U = 0.11$ would be classified into the intermediate risk category based on the conventional risk factors/markers. By including the additional CRP information, these subjects are further stratified into the two risk categories: 37.9% (s.e. = 2.5%) down to the low risk category and 62.1% (s.e. = 2.5%) remaining in the intermediate risk category. As discussed in Section 2, not

all reclassifications are correct. For the present case, 34.9% of the subjects in $\mathcal{D}_s$ are accurately reclassified downward since these subjects would not develop an event within 10 years.

For the second set of analysis by treating non-CHD death as censoring, 18.8% of subjects are censored with respect to the CHD events of interest. For this case, the smoothing parameter is 0.2 via the 10-fold cross validation by minimizing $\sum_{l=1}^{3}\sum_{k=1}^{10}\sum_{q=0}^{1}\mathrm{PE}^{(q)}(k,l)$ in (3.7) with normalizing constant $n^{-1/20}$. We present the results in Figure 2 and Table 1(b) for $\widehat{p}_U = s \in [0.079, 0.63]$ under the same settings as in Figure 1 and Table 1(a). The regions for which the CRP may be helpful are similar to those from the first set of analysis with the additional possible region near $\widehat{p}_U = 0.30$. However, the magnitude of gain is rather minimal near $\widehat{p}_U = 0.30$ given the small reclassification rate. Again, after controlling for the overall family-wise type I error rate of 0.05, there does not appear to be a region with significant gain from CRP.

5. REMARKS

In this article, we present a systematic approach to quantify the added value from new risk markers for classifying subjects into pre-specified risk categories. At each estimated conventional risk level, we provide point and interval estimates for the reclassification rates along with the corresponding proportions of accurate re-assignment for this subgroup of individuals. These quantities play vital roles for cost-benefit analysis even when the cost of measuring the new markers is not an issue or the re-categorization via the new markers is not generated from a random allocation process. In general, if the new markers improve the risk prediction and change the risk estimates drastically, one may expect the new marker to be helpful in risk reclassification for the entire population. However, in most practical settings, we expect that subjects whose conventional risk estimates are not around the risk threshold values would not benefit much from the new markers with respect to a better assignment of risk category. This is indeed confirmed with the results from extensive analysis of the data from the Cardiovascular Health Study. Thus, to optimize the usage of the new markers for risk classification, one may consider ascertaining the new markers only for subjects with conventional risk estimates in a certain range. Our proposed

procedure provides a systematic approach to identifying the subgroups for which the new marker improves risk classification.

Appendix

Throughout, unless noted otherwise, we use the notation $\simeq$ to denote equivalence up to $o_p(1)$ uniformly in s, $\lesssim$ to denote being bounded above up to a universal constant, and $\dot{\mathcal{F}}(x)$ to denote $d\mathcal{F}(x)/dx$ for any function $\mathcal{F}$. We use $\mathbb{P}_n$ and $\mathbb{P}$ to denote expectation with respect to the empirical probability measure of $\{(T_i, \Delta_i, X_i, W_i), i = 1, \cdots, n\}$ and the probability measure of (T, Δ, X, W), respectively. Similarly $\mathbb{G}_n = n^{\frac{1}{2}}(\mathbb{P}_n - \mathbb{P})$.

Let β_0 and γ_0 denote the solution to the equations $E[X_i\{I(T_{0i} < t_0) - g_1(\beta' X_i)\}] = 0$ and $E[W_i\{I(T_{0i} < t_0) - g_2(\gamma' W_i)\}] = 0$, respectively. Let $\bar{p}_U = g_1(\beta_0' X)$. We assume that $\tau(\cdot)$, the density function of $\psi(\bar{p}_U)$, is continuously differentiable with bounded derivatives and bounded away from zero on the interval $[\psi^{-1}(\rho_l), \psi^{-1}(\rho_u)] \subset (0, 1)$. We also assume that the marker values are bounded, $(\beta_0', \gamma_0')'$ belongs to a compact set Ω. For the bandwidth h, we assume that $h = O(n^{-v}), 1/5 < v < 1/2$.

We first note that from Uno et al. (2007), we have

$$|\widehat{\beta} - \beta_0| + |\widehat{\gamma} - \gamma_0| = O_p(n^{-\frac{1}{2}}).$$

It follows that $|\eta_l^{(q)}(s) - \bar{\eta}_l^{(q)}(s)| = O_p(1)$, where

$$\eta_l^{(q)}(s) = E(\widehat{M}_{li}^{(q)} \mid \widehat{p}_U = s), \quad \widehat{M}_{li}^{(q)} = I\{\widehat{p}_{\{U_i,V_i\}} \in [\nu_0, \nu_1), Y_i = q\},$$
$$\bar{\eta}_l^{(q)}(s) = E(M_{li}^{(q)} | \bar{p}_U = s), \quad M_{li}^{(q)} = I\{\bar{p}_{\{U_i,V_i\}} \in [\nu_0, \nu_1), Y_i = q\}, \quad \bar{p}_{\{U,V\}} = g_2(\gamma_0' W).$$

It is also important to note that $E\{\omega s(T_0, X, W)\} = E\{s(T_0, X, W)\}$ for any function $s(\cdot, \cdot, \cdot)$, where $\omega = I(T \le t_0)\Delta/G(T) + I(T > t_0)/G(t_0)$.

We next derive the asymptotic theory for $\widehat{\eta}_l^{(q)}(\cdot)$ when $l = 1$ and $q = 1$, but note that the same arguments can be used to establish the asymptotic theory for other quantities. For the ease of notation, in Appendix A and B, we suppress the subscript l and supscript (q) from $\widehat{M}_{li}^{(q)}$ and $M_{li}^{(q)}$.

A Uniform Consistency of $\widehat{\eta}_1^{(1)}(\cdot)$

At any given s, let $\{\widehat{a}(s), \widehat{b}(s)\}'$ be the root of the estimating equation

$$\sum_{i=1}^{n} \begin{bmatrix} 1 \\ h^{-1}\widehat{\mathcal{E}}_i(s) \end{bmatrix} K_h\{\widehat{\mathcal{E}}_i(s)\}\widehat{\omega}_i \left[\widehat{M}_i - g_0\{a + b\widehat{\mathcal{E}}_i(s)\}\right] = 0,$$

$a(s) = g_0^{-1}\{\eta_1^{(1)}(s)\}$, $b(s) = d[g_0^{-1}\{\eta_1^{(1)}(s)\}]/ds = \dot{\eta}_1^{(1)}(s)/\dot{g}_0\{a(s)\}$, and

$$\widehat{\mathbf{d}}(s) = \begin{bmatrix} \widehat{d}_a(s) \\ \widehat{d}_b(s) \end{bmatrix} = \begin{bmatrix} \widehat{a}(s) - a(s) \\ h^{-1}\{\widehat{b}(s) - b(s)\} \end{bmatrix}.$$

Our objective is to show that $\widehat{\mathbf{d}}(s) \to \mathbf{0}$ in probability as $n \to \infty$. To this end, we note that for any given s, $\widehat{\mathbf{d}}(s)$ is the solution to the estimating equation $\widehat{\mathbb{S}}(\mathbf{d}, s) = \mathbf{0}$, where $\mathbf{d} = (d_a, d_b)'$,

$$\widehat{\mathbb{S}}(\mathbf{d}, s) = \begin{bmatrix} \widehat{\mathbb{S}}_1(\mathbf{d}; s) \\ \widehat{\mathbb{S}}_2(\mathbf{d}; s) \end{bmatrix} = n^{-1}\sum_{i=1}^{n} \begin{bmatrix} 1 \\ h^{-1}\widehat{\mathcal{E}}_i(s) \end{bmatrix} K_h\{\widehat{\mathcal{E}}_i(s)\}\omega_i \left[\widehat{M}_i - \mathcal{G}\{\mathbf{d}, s; \psi(\widehat{p}_{U_i}), h\}\right],$$

and

$$\mathcal{G}(\mathbf{d}, s; y, h) = g_0[a(s) + b(s)\{y - \psi(s)\} + d_a + d_b h^{-1}\{y - \psi(s)\}].$$

The first step is to show that $\widehat{\mathbb{S}}(\mathbf{d}; s)$ is uniformly consistent for

$$\mathbb{S}(\mathbf{d}; s) = \begin{bmatrix} \mathbb{S}_1(\mathbf{d}; s) \\ \mathbb{S}_2(\mathbf{d}; s) \end{bmatrix} = \tau\{\psi(s)\} \begin{bmatrix} \eta_1^{(1)}(s) - \int K(t) g_0\{a(s) + d_a + d_b t\} dt \\ -\int tK(t) g_0\{a(s) + d_a + d_b t\} dt \end{bmatrix}$$

We first show that

$$\sup_s \left| n^{-1}\sum_{i=1}^{n} K_h\{\widehat{\mathcal{E}}_i(s)\}\widehat{\omega}_i\widehat{M}_i - \tau\{\psi(s)\}\eta_1^{(1)}(s) \right|$$

and

$$\sup_{\mathbf{d},s} \left| n^{-1}\sum_{i=1}^{n} K_h\{\widehat{\mathcal{E}}_i(s)\}\widehat{\omega}_i\mathcal{G}\{\mathbf{d}, s; \psi(\widehat{p}_{U_i}), h\} - \tau\{\psi(s)\}\int K(t) g_0\{a(s) + d_a + d_b t\} dt \right|$$

are both $O_p\{(nh)^{-\frac{1}{2}}\log(n)\}$. To this end, we note that since

$$\sup_u |\widehat{G}(u) - G(u)| = O_p(n^{-\frac{1}{2}}) \qquad \text{(A·1)}$$

(Kalbfleisch & Prentice, 2002) and $|\widehat{\beta}-\beta_0|=O_p(n^{-\frac{1}{2}})$,

$$\left|n^{-1}\sum_{i=1}^{n}(\widehat{\omega}_i-\omega_i)K_h\{\widehat{\mathcal{E}}_i(s)\}\mathcal{G}\{\mathbf{d},s;\psi(\widehat{p}_{U_i}),h\}\right|\leq n^{-1}\sum_{i=1}^{n}|\widehat{\omega}_i-\omega_i|K_h\{\widehat{\mathcal{E}}_i(s)\}=O_p(n^{-\frac{1}{2}}).$$

This, together with the convergence of $n^{\frac{1}{2}}(\widehat{\beta}-\beta_0)$, implies that

$$\begin{aligned}
&\left|\, n^{-1}\sum_{i=1}^{n}K_h\{\widehat{\mathcal{E}}_i(s)\}\widehat{\omega}_i\mathcal{G}\{\mathbf{d},s;\psi(\widehat{p}_{U_i}),h\}-\tau\{\psi(s)\}\int K(t)g_0\{a(s)+d_a+d_bt\}dt\,\right|\\
\leq&\left|\, n^{-\frac{1}{2}}\int K_h\{y-\psi(s)\}\mathcal{G}(\mathbf{d},s;y,h)d\mathbb{G}_n\left[\omega I\{\psi(\widehat{p}_U)\leq y\}-\omega I\{\psi(\bar{p}_U)\leq y\}\right]\,\right|\\
&+\left|\int K_h\{y-\psi(s)\}\mathcal{G}(\mathbf{d},s;y,h)d\mathbb{P}\left[\omega I\{\psi(\widehat{p}_U)\leq y\}\right]-\tau\{\psi(s)\}\int K(t)g_0\{a(s)+d_a+d_bt\}dt\,\right|\\
&+\left|\, n^{-\frac{1}{2}}\int K_h\{y-\psi(s)\}d\mathbb{G}_n\left[\omega\mathcal{G}\{\mathbf{d},s;\psi(\bar{p}_U),h\}I\{\psi(\bar{p}_U)\leq y\}\right]\,\right|+O_p(n^{-\frac{1}{2}})\\
\lesssim& n^{-\frac{1}{2}}h^{-1}\|\mathbb{G}_n\|_{\mathcal{H}_\delta}+\left|\, n^{-\frac{1}{2}}\int K_h\{y-\psi(s)\}d\mathbb{G}_n\left[\omega\mathcal{G}\{\mathbf{d},s;\psi(\bar{p}_U),h\}I\{\psi(\bar{p}_U)\leq y\}\right]\,\right|+O_p(n^{-\frac{1}{2}}+h^2)
\end{aligned}$$

where $\mathcal{H}_\delta=\{\omega I[\psi\{g_1(\beta' x)\}\leq c]-\omega I[\psi\{g_1(\beta_0' x)\}\leq c]:\|\beta-\beta_0\|\leq\delta,y\}$ is a class of functions indexed by β and c. Furthermore, $\mathcal{H}_\delta$ is uniformly bounded by an envelop function in the order of $\delta^{\frac{1}{2}}$ with respect to L_2 norm. By the maximum inequality of Van der vaart and Wellner (1996), we have

$$\mathrm{E}\|\mathbb{G}_n\|_{\mathcal{H}_\delta}\lesssim\delta^{\frac{1}{2}}(|\log(\delta)|+|\log(h)|)\left(1+\frac{\delta^{\frac{1}{2}}(|\log(\delta)|+|\log(h)|)}{\delta n^{\frac{1}{2}}}\right).$$

This, coupled with the fact that $|\widehat{\beta}-\beta_0|=O_p(n^{-\frac{1}{2}})$, implies that

$$n^{-\frac{1}{2}}h^{-1}\|\mathbb{G}_n\|_{\mathcal{H}_\delta}\lesssim O_p\{(nh)^{-\frac{1}{2}}(nh^2)^{-\frac{1}{4}}\log(n)\}.$$

Secondly, with the standard arguments used in Bickel & Rosenblatt (1973), it can be shown that

$$\left|n^{-\frac{1}{2}}\int K_h\{x-\psi(s)\}d\mathbb{G}_n\left[\omega\mathcal{G}\{\mathbf{d},s;\psi(\bar{p}_U),h\}I\{\psi(\bar{p}_U)\leq x\}\right]\right|=O_p\{(nh)^{-\frac{1}{2}}\log(n)\}.$$

Therefore, for $h=n^{-v},1/5<v<1/2$,

$$\sup_{\mathbf{d},s}\left|\, n^{-1}\sum_{i=1}^{n}K_h\{\widehat{\mathcal{E}}_i(s)\}\widehat{\omega}_i\mathcal{G}\{\mathbf{d},s;\psi(\widehat{p}_{U_i}),h\}-\tau\{\psi(s)\}\int K(t)g_0\{a(s)+d_a+d_bt\}dt\,\right|$$

is $O_p\{(nh)^{-\frac{1}{2}}\log(n)\}$. Following from (A·1), $|\widehat{\gamma}-\gamma_0| = O_p(n^{-\frac{1}{2}})$, and similar arguments as given above, we have

$$\sup_s \left| n^{-1}\sum_{i=1}^{n} K_h\{\widehat{\mathcal{E}}_i(s)\}\widehat{\omega}_i\widehat{M}_i - \tau\{\psi(s)\}\eta_1^{(1)}(s)\right| = O_p\{(nh)^{-\frac{1}{2}}\log(n)\},$$

and hence $\sup_{\mathbf{d},s}|\widehat{\mathbb{S}}_1(\mathbf{d},s)-\mathbb{S}_1(\mathbf{d},s)| = O_p\{(nh)^{-\frac{1}{2}}\log(n)\} = o_p(1)$. It follows from the same arguments as given above that $\sup_{\mathbf{d},s}|\widehat{\mathbb{S}}_2(\mathbf{d},s)-\mathbb{S}_2(\mathbf{d},s)| = O_p\{(nh)^{-\frac{1}{2}}\log(n)+h\} = o_p(1)$. Therefore $\sup_{\mathbf{d},s}|\widehat{\mathbb{S}}(\mathbf{d},s)-\mathbb{S}(\mathbf{d},s)| = o_p(1)$. This uniform convergence, coupled with the fact that $\mathbf{0}$ is the unique solution to the equation $\mathbb{S}(\mathbf{d},s)=0$ with respect to $\mathbf{d}$ and all the eigenvalues of $\mathbb{A}(s) = -\partial\mathbb{S}(\mathbf{d};s)/\partial\mathbf{d}'|_{\mathbf{d}=\mathbf{0}} = \tau\{\psi(s)\}\dot{g}_0\{a(s)\}\mathrm{diag}\{1,\int v^2K(v)dv\}$ are uniformly bounded above zero, suggests that $\sup_s|\widehat{\mathbf{d}}(s)| = O_p\{(nh)^{-\frac{1}{2}}\log(n)+h\} = o_p(1)$, which implies the consistency of $\widehat{\eta}_1^{(1)}(s) = g_0\{\widehat{a}(s)\}$.

B Asymptotic distribution of $\widehat{\eta}_1^{(1)}(\cdot)$

It follows from a Taylor series expansion that

$$(nh)^{\frac{1}{2}}\widehat{d}_a(s) = (nh)^{\frac{1}{2}}\{\widehat{a}(s)-a(s)\} = \widehat{\mathbb{A}}_1(s)'(nh)^{\frac{1}{2}}\widehat{\mathbb{S}}(\mathbf{0};s) + O_p\left\{(nh)^{\frac{1}{2}}(|\widehat{d}_a(s)|^2+|\widehat{d}_b(s)|^2)\right\},$$

where $\widehat{\mathbb{A}}_1(s)$ is the first row of $\widehat{\mathbb{A}}(s) = -\{\partial\widehat{\mathbb{S}}(\mathbf{d};s)/\partial\mathbf{d}'\}^{-1}|_{\mathbf{d}=\mathbf{0}}$. Using the similar arguments in the previous section, one can show that $\widehat{\mathbb{A}}_1(s)$ converges to $\mathbb{A}_1(s)$, the first row of $\mathbb{A}(s)$, uniformly in s. Furthermore, with the convergence rate of $\widehat{\mathbb{S}}(\mathbf{d},s)$, it is not difficult to show that the remainder term is bounded by $O_p\left\{(nh)^{-\frac{1}{2}}\log(n)^2+(nh)^{\frac{1}{2}}h^2\right\}$ uniformly in s. It follows that

$$\begin{aligned}(nh)^{\frac{1}{2}}\widehat{d}_a(s) &= \mathbb{A}_1(s)'(nh)^{\frac{1}{2}}\widehat{\mathbb{S}}(\mathbf{0};s) + O_p\left\{(nh)^{-\frac{1}{2}}\log(n)^2+(nh)^{\frac{1}{2}}h^2\right\})\\ &= \frac{(nh)^{\frac{1}{2}}\widehat{\mathbb{S}}_1\mathbf{0};s)}{\tau\{\psi(s)\}\dot{g}_0\{a(s)\}} + O_p\left\{(nh)^{-\frac{1}{2}}\log(n)^2+(nh)^{\frac{1}{2}}h^2\right\}.\end{aligned}$$

This, together with the convergence rate of $\widehat{\mathbb{S}}_1(\mathbf{0};s)$, implies $\sup_s|\widehat{d}_a(s)| = O_p\{(nh)^{-\frac{1}{2}}\log(n)\}$. It follows that $(nh)^{\frac{1}{2}}\widehat{d}_a(s)$ is asymptotically equivalent to

$$\frac{(nh)^{\frac{1}{2}}\mathbb{P}_n\left(K_h\{\widehat{\mathcal{E}}(s)\}\omega\left[\widehat{M}-\mathcal{G}_0\{s,\psi(\widehat{p}_U)\}\right]\right)}{\tau\{\psi(s)\}\dot{g}_0\{a(s)\}}.$$

where $\mathcal{G}_0(s,y) = g_0[a(s) + b(s)\{y - \psi(s)\}]$. We next show that $(nh)^{\frac{1}{2}}\widehat{d}_a(s)$ is asymptotically equivalent to

$$(nh)^{\frac{1}{2}}\tilde{d}_a(s) = \frac{(nh)^{\frac{1}{2}}\mathbb{P}_n\left(K_h\{\mathcal{E}(s)\}\omega\left[M - \mathcal{G}_0\{s,\psi(p_U)\}\right]\right)}{\tau\{\psi(s)\}\dot{g}_0\{a(s)\}},$$

i.e., $\widehat{p}_U$ and $\widehat{p}_{\{U,V\}}$ can be replaced by their respective $\bar{p}_U$ and $\bar{p}_{\{U,V\}}$ in $\widehat{d}_a(s)$, where $\mathcal{E}(s) = \psi(\bar{p}_U) - \psi(s)$. To this end, noticing the fact that $\tau\{\psi(s)\}\dot{g}_0\{a(s)\}$ is bounded away from zero uniformly in s, we have

$$\begin{aligned}
(nh)^{\frac{1}{2}}\left|\widehat{d}_a(s) - \tilde{d}_a(s)\right| \lesssim\ & h^{\frac{1}{2}}\left|\int K_h\{x-\psi(s)\}d\mathbb{G}_n\left[\omega\widehat{M}I\{\psi(\widehat{p}_U)\le x\} - \omega M I\{\psi(\bar{p}_U)\le x\}\right]\right| \\
& + h^{\frac{1}{2}}\left|\int K_h\{x-\psi(s)\}\mathcal{G}_0(s,x)d\mathbb{G}_n\left[\omega I\{\psi(\widehat{p}_U)\le x\} - \omega I\{\psi(\bar{p}_U)\le x\}\right]\right| \\
& + \left|(nh)^{\frac{1}{2}}\int K_h\{x-\psi(s)\}d\mathbb{P}\left[\omega\{\widehat{M} - \mathcal{G}_0(s,\widehat{p}_U)\}I\{\psi(\widehat{p}_U)\le x\}\right]\right| \\
\lesssim\ & h^{-\frac{1}{2}}\|\mathbb{G}_n\|_{\mathcal{F}_\delta} + h^{-\frac{1}{2}}\|\mathbb{G}_n\|_{\mathcal{H}_\delta} + O_p\{(nh)^{\frac{1}{2}}|\widehat{\beta}-\beta_0| + |\widehat{\gamma}-\gamma_0| + h^2\}
\end{aligned}$$

where $\mathcal{F}_\delta = \big\{\omega y I\{g_2(\gamma' w)\in[\nu_0,\nu_1)\}I[\psi\{g_1(\beta' x)\}\le c] - \omega y I\{g_2(\gamma_0' w)\in[\nu_0,\nu_1)\}I[\psi\{g_1(\beta_0' x)\}\le c]\big\} : |\gamma-\gamma_0|+|\beta-\beta_0|\le\delta, c\}$ is the class of functions indexed by γ, β and c. By the maximum inequality and the fact that $|\widehat{\beta}-\beta_0|+|\widehat{\gamma}-\gamma_0| = O_p(n^{-\frac{1}{2}})$, we have $h^{-\frac{1}{2}}\|\mathbb{G}_n\|_{\mathcal{F}_\delta} = O_p\{h^{-\frac{1}{2}}n^{-\frac{1}{4}}\log(n)\}$. It follows that $(nh)^{\frac{1}{2}}\widehat{d}_a(s) - (nh)^{\frac{1}{2}}\tilde{d}_a(s)$ is asymptotic $o_p(1)$ uniformly in s. This, together with the standard arguments for local linear regression fitting, implies that $(nh)^{\frac{1}{2}}\{\widehat{a}(s) - a(s)\}$ converges to a normal with mean 0 and variance $\sigma^2(s)\dot{g}_0\{a(s)\}^{-2}$, where $m_2 = \int K(s)^2 ds$ and

$$\sigma^2(s) = \frac{m_2}{\tau^2\{\psi(s)\}}E\left[\frac{\{M - \eta_1^{(1)}(s)\}^2}{G(T_0\wedge t_0)}\ \middle|\ \bar{p}_U = s\right]$$

It follows from the delta method that $(nh)^{\frac{1}{2}}\{\widehat{\eta}_1^{(1)}(s) - \eta_1^{(1)}(s)\}$ is asymptotically normal with mean 0 and variance $\sigma^2(s)$. To justify the resampling method, we note that conditional on the data the random variable

$$\frac{(nh)^{\frac{1}{2}}\mathbb{P}_n\left[K_h\{\widehat{\mathcal{E}}(s)\}\widehat{\omega}\left\{\widehat{M} - \widehat{\eta}_1^{(1)}(s)\right\}Z\right]}{\widehat{\tau}\{\psi(s)\}}$$

is asymptotical normally distributed with mean zero and variance

$$\widehat{\sigma}^2(s) = \frac{h\mathbb{P}_n\left[K_h\{\widehat{\mathcal{E}}(s)\}^2\widehat{\omega}^2\left\{\widehat{M} - \widehat{\eta}_1^{(1)}(s)\right\}^2\right]}{\widehat{\tau}^2\{\psi(s)\}}.$$

It follows from the arguments given in Appendix A to show that which converges to $\sigma^2(s)$, as $n \to \infty$.

C Asymptotic null distribution of $\widehat{D}(s)$

From Appendix A, $\sup_s |\widehat{D}(s) - \sum_{l=1}^{L} \tilde{\mathfrak{w}}_l \{\eta_l^{(1)}(s) - \eta_l(s)\xi(s)\}| \to 0$ in probability as $n \to \infty$. Furthermore, as $n \to \infty$, $(nh)^{\frac{1}{2}}[\widehat{D}(s) - \sum_{l=1}^{L} \tilde{\mathfrak{w}}_l \{\eta_l^{(1)}(s) - \eta_l(s)\xi(s)\}]$ can be uniformly approximated by

$$(nh)^{\frac{1}{2}} \sum_{l=1}^{L} \tilde{\mathfrak{w}}_l \left[\{\widehat{\eta}_l^{(1)}(s) - \eta_l^{(1)}(s)\} - \sum_{q=0}^{1} \{\widehat{\eta}_l^{(q)}(s) - \eta_l^{(q)}(s)\}\xi(s) - \sum_{l'=1}^{L} \{\widehat{\eta}_{l'}^{(1)}(s) - \eta_{l'}^{(1)}(s)\}\eta_l(s) \right]$$

$$\approx (nh)^{\frac{1}{2}} \mathbb{P}_n \left[K_h\{\mathcal{E}(s)\}\mathcal{A}(s) \right] \approx (nh)^{\frac{1}{2}} \mathbb{P}_n \left[K_h\{\mathcal{E}(s)\}\mathcal{A}(\bar{p}_U) \right]$$

where

$$\mathcal{A}(s) = \omega \sum_{l=1}^{L} \sum_{q=0}^{1} \kappa_l^{(q)}(s) \left[M_l^{(q)} - g_0\{a_l^{(q)}(s) + b_l^{(q)}(s)\mathcal{E}(s)\} \right],$$

$$\kappa_l^{(q)}(s) = q\{\tilde{\mathfrak{w}}_l - \eta_l(s)\} - \xi(s)\tilde{\mathfrak{w}}_l,\ a_l^{(q)}(s) = g_0^{-1}\{\eta_l^{(q)}(s)\},\ b_l^{(q)}(s) = \frac{\dot{\eta}_l^{(q)}(s)}{\dot{g}_0\{a_l^{(q)}(s)\}},$$

Under the null hypothesis, $\eta_l^{(1)}(s) = \eta_l(s)\xi(s), l = 1, \cdots, L$ and

$$(nh)^{\frac{1}{2}} \widehat{D}(s) \approx (nh)^{\frac{1}{2}} \mathbb{P}_n \left[K_h\{\mathcal{E}(s)\}\mathcal{A}(\bar{p}_U) \right].$$

Let $r(s) = m_2 E\{\mathcal{A}(\bar{p}_U)^2 | \bar{p}_U = s\}dt$, and

$$\widehat{\mathfrak{D}}(s) = \frac{(nh)^{\frac{1}{2}}}{[r(s)\tau\{\psi(s)\}]^{\frac{1}{2}}} \mathbb{P}_n[\mathcal{A}(\bar{p}_U) K_h\{\mathcal{E}(s)\}].$$

By Lemma 1 of Fan and Zhang (1998), we have

$$\text{pr}\left[r_n \left\{ \sup_{s \in [\rho_l, \rho_r]} \left| \widehat{\mathfrak{D}}(s) \right| - d_n \right\} < x \right] \to \frac{1 + \exp(-2e^{-x})}{2},$$

as $n \to \infty$, where

$$r_n = [2\log\{(\rho_r - \rho_l)/h\}]^{\frac{1}{2}}, \quad \text{and} \quad d_n = r_n + r_n^{-1} \log\left\{ (4\pi m_2)^{-1} \int_{-1}^{1} \dot{K}(t)^2 dt \right\}.$$

This implies that under the null hypothesis,

$$\text{pr}\left[r_n\left\{\sup_{s\in[\rho_l,\rho_r]}\left|\frac{(nh)^{\frac{1}{2}}\widehat{D}(s)}{r(s)\tau\{\psi(s)\}}\right|-d_n\right\}<x\right]\rightarrow\frac{1+\exp(-2e^{-x})}{2},$$

as $n\rightarrow\infty$. To justify the resampling method for approximating the null distribution, we let

$$\widehat{D}^*(s)=\mathbb{P}_n\left\{K_h\{\widehat{\mathcal{E}}(s)\}\sum_{l=1}^{L}\tilde{\mathfrak{w}}_l\left(\widehat{\mathcal{V}}_l^{(1)}-\sum_{q=0}^{1}\widehat{\mathcal{V}}_l^{(q)}\xi(s)-\sum_{l'=1}^{L}\widehat{\mathcal{V}}_{l'}^{(1)}\eta_l(s)\right)Z\right\}$$

and $\widehat{r}(s)=\text{var}\{\widehat{D}^*(s)|\mathcal{O}\}$ and $\mathcal{O}=\{(Y_i,\Delta_i,U_i,V_i),i=1,\cdots,n)\}$. It follows the same arguments as given above that

$$\text{pr}\left[r_n\left\{\sup_{s\in\Omega_h}\left|\frac{(nh)^{\frac{1}{2}}D^*(s)}{\widehat{r}(s)\widehat{\tau}\{\psi(s)\}}\right|-d_n\right\}<x\;\middle|\;\mathcal{O}\right]\rightarrow\frac{1+\exp(-2e^{-x})}{2},$$

as the sample size goes to infinity. Therefore, we can use the conditional distribution of

$$\sup_{s\in\Omega_h}\frac{(nh)^{\frac{1}{2}}\widehat{D}^*(s)}{\widehat{r}(s)\widehat{\tau}\{\psi(s)\}}$$

to approximate the null distribution of

$$\sup_{s\in\Omega_h}\frac{(nh)^{\frac{1}{2}}\widehat{D}(s)}{r(s)\tau\{\psi(s)\}}.$$

References

Armitage, P. (1955). Tests for linear trends in proportions and frequencies. *Biometrics* **11**, 375–386.

Bickel, P. J. & Rosenblatt, M. (1973). On some global measures of the deviations of density function estimates (Corr: V3 p1370). *The Annals of Statistics* **1**, 1071–1095.

Buring, J. & Hennekens, C. (1992). The Women's Health Study: summary of the study design. *Journal of Myocardial Ischemia* **4**, 27–9.

Cochran, W. (1954). Some methods for strengthening the common χ 2 tests. *Biometrics* **10**, 417–451.

Cook, N. (2007). Use and Misuse of the Receiver Operating Characteristic Curve in Risk Prediction. *Circulation* **115**, 928.

Cook, N., Buring, J. & Ridker, P. (2006). The effect of including c-reactive protein in cardiovascular risk prediction models for women. *Ann Intern Med* **145**, 21–29.

Fan, J. & Gijbels, I. (1996). Local polynomial modelling and its applications, Vol. 66 of Monographs on Statistics and Applied Probability. *London: Chapman Hall,* .

Fried, L., Borhani, N., Enright, P., Furberg, C., Gardin, J., Kronmal, R., Kuller, L., Manolio, T., Mittelmark, M. & Newman, A. (1991). The Cardiovascular Health Study: design and rationale. *Ann Epidemiol* **1**, 263–76.

Grundy, S., Cleeman, J., Bairey Merz, C., Brewer, H., Clark, L., Hunninghake, D., Pasternak, R., Smith, S. & Stone, N. (2004). Implications of Recent Clinical Trials for the National Cholesterol Education Program Adult Treatment Panel III Guidelines. *Journal of the American College of Cardiology* **44**, 720–732.

Kalbfleisch, J. D. & Prentice, R. L. (2002). *The Statistical Analysis of Failure Time Data*. John Wiley & Sons.

Mosca, L., Appel, L., Benjamin, E. & et al. (American Heart Association) (2004). Evidence-based guidelines for cardiovascular disease prevention in women. *Circulation* **109**, 672–93.

Mosca, L., Banka, C., Benjamin, E. & et al. (for the Expert Panel/Writing Group) (2007). Evidence-based guidelines for cardiovascular disease prevention in women: 2007 update. *Circulation* **115**, 1481–501.

Park, B., Kim, W., Ruppert, D., Jones, M., Signorini, D. & Kohn, R. (1997). Simple transformation techniques for improved non-parametric regression. *Scandinavian journal of statistics* **24**, 145–163.

Pencina, M., D'Agostino Sr, R., D'Agostino Jr, R. & Vasan, R. (2008). Evaluating the added predictive ability of a new marker: From area under the ROC curve to reclassification and beyond. *Stat Med* **27**, 157–72.

Pepe, M., Feng, Z., Huang, Y., Longton, G., Prentice, R., Thompson, I. & Zheng, Y. (2008). Integrating the Predictiveness of a Marker with Its Performance as a Classifier. *American Journal of Epidemiology* **167**, 362.

Pepe, M., Janes, H., Longton, G., Leisenring, W. & Newcomb, P. (2004). Limitations of the Odds Ratio in Gauging the Performance of a Diagnostic, Prognostic, or Screening Marker. *American Journal of Epidemiology* **159**, 882–890.

Ridker, P., Buring, J., Rifai, N. & Cook, N. (2007). Development and validation of improved algorithms for the assessment of global cardiovascular risk in women: the reynolds risk score. *JAMA* **297**, 611–9.

Tian, L., Cai, T., Goetghebeur, E. & Wei, L. (2007). Model evaluation based on the sampling distribution of estimated absolute prediction error. *Biometrika* **94**, 297.

Uno, H., Cai, T., Tian, L. & Wei, L. J. (2007). Evaluating Prediction Rules for t-Year Survivors With Censored Regression Models. *J. Am. Stat. Assoc.* **102**, 527–537.

Wand, M. & Jones, M. (1995). *Kernel Smoothing*. Chapman & Hall/CRC.

Wand, M., Marron, J. & Ruppert, D. (1991). Transformation in density estimation (with comments). *J. Am. Stat. Assoc.* **86**, 343–361.

Wang, T., Gona, P., Larson, M., Tofler, G., Levy, D., Newton-Cheh, C., Jacques, P., Rifai, N., Selhub, J., Robins, S. et al. (2006). Multiple Biomarkers for the Pre-

diction of First Major Cardiovascular Events and Death. *New England Journal of Medicine* **355**, 2631.

Wilson, P., D'Agostino, R., Levy, D., Belanger, A., Silbershatz, H. & Kannel, W. (1998). Prediction of coronary heart disease using risk factor categories. *Circulation* **97**, 1837–47.

Figure 1: Evaluating CRP incremental values for female participants from Cardiovascular Heath Study by treating non-CHD death as non-event; (a) Density function estimate of the conventional risk estimate; (b) Reclassification rate estimates; (c)-(e) $\widehat{\eta}_l^{(1)}(\cdot), l = 1, 2, 3$, and corresponding expected values under the null; (f) Point- and simultaneous-p-value curves

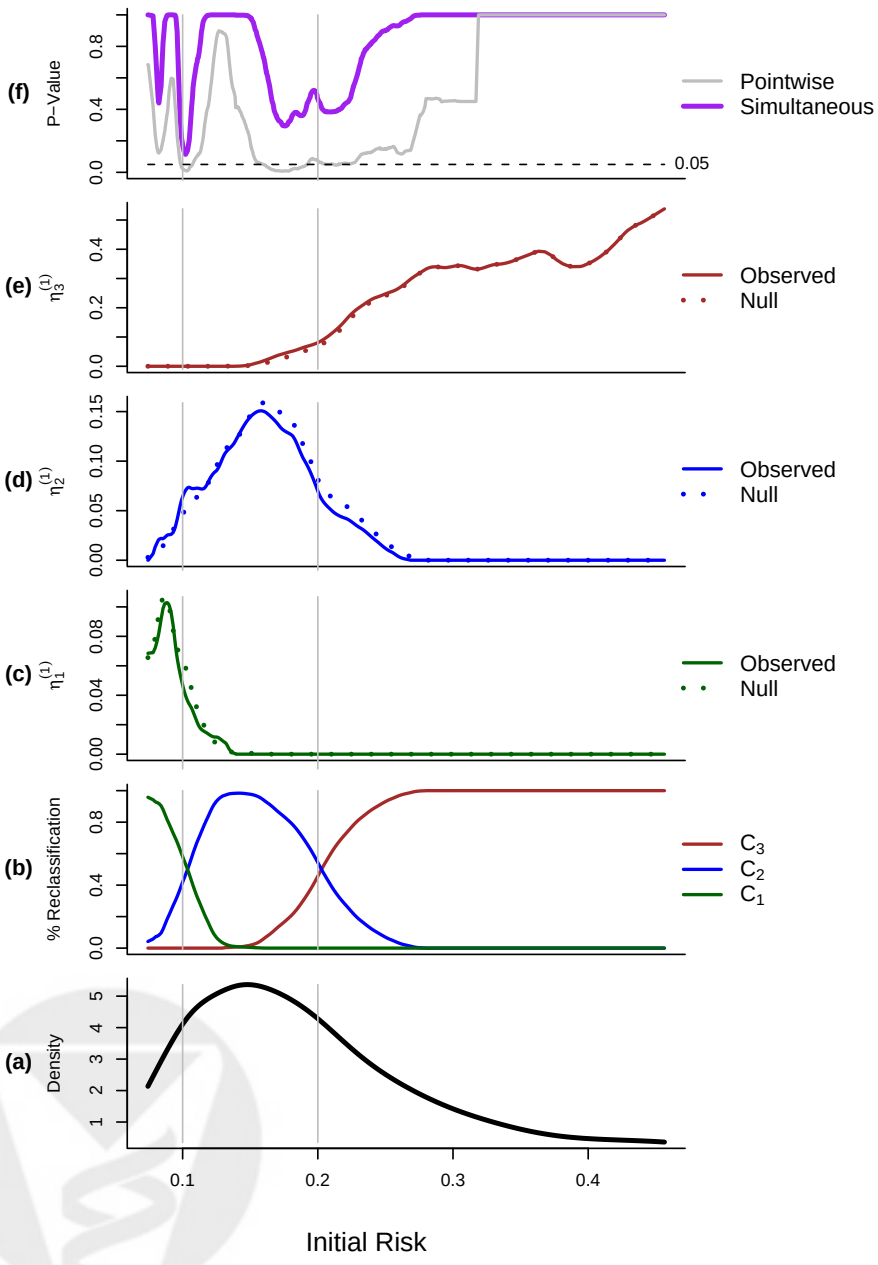

Figure 2: Evaluating CRP incremental values for female participants from Cardiovascular Heath Study by treating non-CHD death as censoring event; (a) Density function estimate of the conventional risk estimate; (b) Reclassification rate estimates; (c)-(e) $\widehat{\eta}_l^{(1)}(\cdot), l = 1, 2, 3$, and corresponding expected values under the null; (f) Point- and simultaneous-p-value curves

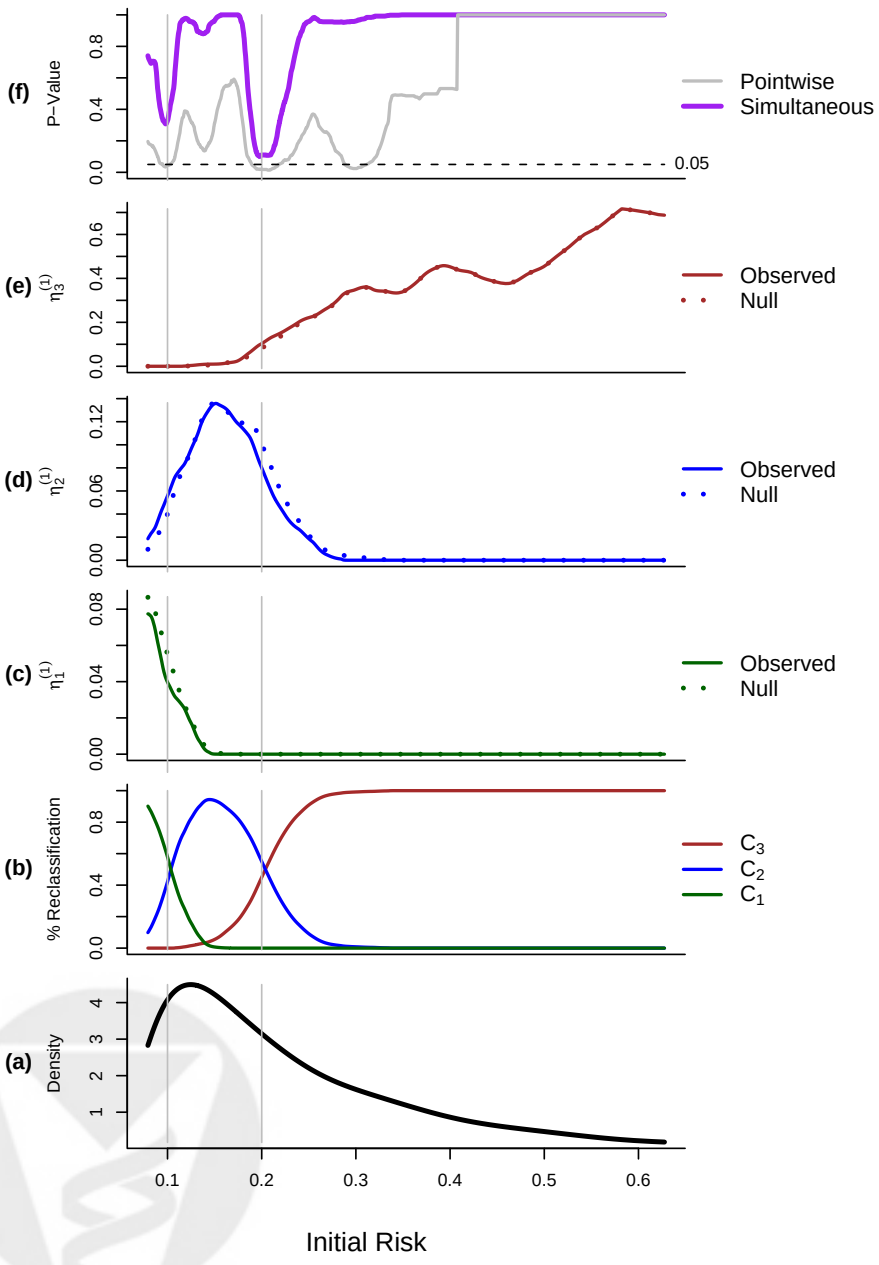

Table 1: Estimates for $\widehat{\eta}_l, \widehat{\eta}_l^{(0)}$ and $\widehat{\eta}_l^{(1)}$ for female participants from the Cardiovascular Health Study with various conventional risk estimates (shown in the parenthesis are corresponding standard error estimates).

(a) Treating non-CHD death as non-event

s		$\widehat{\eta}_l(s)$	$\widehat{\eta}_l^{(1)}(s)$	$\widehat{\eta}_l^{(0)}(s)$
0.11	l=1	0.379(0.025)	0.027(0.009)	0.349(0.025)
	l=2	0.621(0.025)	0.072(0.014)	0.543(0.025)
	l=3	0.000(0.000)	0.000(0.000)	0.000(0.000)
0.18	l=1	0.000(0.000)	0.000(0.000)	0.000(0.000)
	l=2	0.798(0.017)	0.127(0.014)	0.664(0.020)
	l=3	0.202(0.017)	0.050(0.009)	0.150(0.015)
0.21	l=1	0.000(0.000)	0.000(0.000)	0.000(0.000)
	l=2	0.407(0.021)	0.052(0.010)	0.349(0.020)
	l=3	0.593(0.021)	0.108(0.014)	0.476(0.022)

(b) Treating non-CHD death as censoring event

s		$\widehat{\eta}_l(s)$	$\widehat{\eta}_l^{(1)}(s)$	$\widehat{\eta}_l^{(0)}(s)$
0.10	l=1	0.587(0.023)	0.040(0.015)	0.540(0.024)
	l=2	0.413(0.023)	0.055(0.014)	0.352(0.022)
	l=3	0.000(0.001)	0.000(0.000)	0.000(0.001)
0.21	l=1	0.000(0.000)	0.000(0.000)	0.000(0.000)
	l=2	0.429(0.021)	0.061(0.015)	0.356(0.019)
	l=3	0.571(0.021)	0.130(0.017)	0.426(0.020)
0.30	l=1	0.000(0.000)	0.000(0.000)	0.000(0.000)
	l=2	0.009(0.004)	0.000(0.001)	0.009(0.004)
	l=3	0.991(0.004)	0.346(0.026)	0.642(0.026)