Published in final edited form as:

Biometrics. 2016 December ; 72(4): 1037–1045. doi:10.1111/biom.12515.

Robust Risk Prediction with Biomarkers under Two-Phase Stratified Cohort Design

Rebecca Payne^{1,†}, Ming Yang^{1,†}, Yingye Zheng², Majken K. Jensen³, and Tianxi Cai^{1,*}

¹Department of Biostatistics, Harvard School of Public Health, Boston, MA

²Fred Hutchinson Cancer Research Center, Seattle, WA

³Department of Nutrition, Harvard School of Public Health, Boston, MA

Summary

Identification of novel biomarkers for risk prediction is important for disease prevention and optimal treatment selection. However, studies aiming to discover which biomarkers are useful for risk prediction often require the use of stored biological samples from large assembled cohorts, and thus the depletion of a finite and precious resource. To make efficient use of such these stored samples, two-phase sampling designs are often adopted as resource-efficient sampling strategies, especially when the outcome of interest is rare. Existing methods for analyzing data from two-phase studies focus primarily on single marker analysis or fitting the Cox regression model to combine information from multiple markers. However, the Cox model may not fit the data well. Under model misspecification, the composite score derived from the Cox model may not perform well in predicting the outcome. Under a general two-phase stratified cohort sampling design, we present a novel approach to combining multiple markers to optimize prediction by fitting a flexible non-parametric transformation model. Using inverse probability weighting to account for the outcome dependent sampling, we propose to estimate the model parameters by maximizing an objective function which can be interpreted as a weighted C-statistic for survival outcomes. Regardless of model adequacy, the proposed procedure yields a sensible composite risk score for prediction. A major obstacle for making inference under two phase studies is due to the correlation induced by the finite population sampling, which prevents standard inference procedures such as the bootstrap from being used for variance estimation. We propose a resampling procedure to derive valid confidence intervals for the model parameters and the C-statistic accuracy measure. We illustrate the new methods with simulation studies and an analysis of a two-phase study of high-density lipoprotein cholesterol (HDL-C) subtypes for predicting the risk of coronary heart disease.

*Correspondence to: tcgai@hsph.harvard.edu.

†Contributed equally to this work

6. Supplementary Materials

The Web Appendices referenced in Section 2.3, Section 2.4, and the R code for the proposed inference procedures are available with this paper at the *Biometrics* website on Wiley Online Library.

Keywords

C-statistic; Finite population sampling; Outcome-dependent sampling; Perturbation resampling; Risk prediction; Robustness; Two-phase stratified cohort sampling

1. Introduction

Novel biomarkers useful for risk prediction can potentially lead to neffective personalized disease prevention and treatment strategies. The discovery of such markers often relies on prospective cohort studies, in which biospecimens are typically collected at baseline and stored for future research. Examples include the Nurses' Health Study (NHS) (Colditz et al., 1997), the Atherosclerosis Risk in Communities (ARIC) study (Folsom et al., 2002), the Health Professional Follow-Up study (HPFS) (Rimm et al., 1991), and the Women's Health Initiative (WHI) (Anderson et al., 2003). These study cohorts along with the stored biospecimens are extremely valuable resources for conducting biomarker studies.

One major challenge for biomarker studies is that it is often impossible to measure biomarkers on the full cohort, since the assessment of biomarkers can be expensive and labor intensive. Furthermore, these banked biologic samples constitute a finite and valuable resource, and thus it is necessary to make efficient use of them even when measurement on the full cohort is feasible. Two-phase sampling designs, including the well-known case cohort (CCH) design (Prentice, 1986) and exposure stratified case cohort (sCCH) design (Borgan et al., 2000), are cost effective strategies for such settings, especially when the outcome of interest is rare. In a two-phase sampling design, a large cohort is first randomly selected from the population of interest in phase I and a (stratified) subset is then chosen from the phase I cohort to form phase II sample. Expensive markers are only measured for subjects selected into the phase II sample. Two-phase designs have been employed widely in existing large cohort studies, for example in the WHI to examine biomarkers which may be risk factors for endometrial cancer (Gunter et al., 2008), colorectal cancer (Gunter et al., 2008), and breast cancer (Gunter et al., 2009), in the MONICA/KORA Ausberg study to examine chemokines predictive of coronary heart disease (Herder et al., 2006), and in the Rotterdam Study to examine biomarkers and the risk of Alzheimer's disease (van Oijen et al., 2006).

Under a general two-phase stratified cohort (sCohort) design (Breslow and Wellner, 2007; Gray, 2009), the phase II sample consists of subjects randomly selected *without replacement* (WOR) within strata defined by case status and possibly other exposure matching variables. The sCohort design thus makes efficient use of biologic and financial resources, although its outcome-dependent sampling scheme introduces significant complexity to analysis because the sampling induces data dependent sampling bias as well as between subjects correlation that is not ignorable at the first order. The sCohort design we consider in this paper can be viewed as a modification and generalization of the CCH or the sCCH design. The CCH and sCCH designs include all cases in Phase II while the sCohort design allows subsampling of the cases. In addition, CCH and sCCH sample 'subcohort' from the entire cohort and subcohort members can overlap with cases, whereas sCohort samples from mutually

exclusive strata. Similar to the sCCH design, the sCohort design can lead to more efficient estimation than the standard CCH design by making use of properly chosen exposure matching variables.

Much of the existing statistical methods for analyzing data from two-phase CCH, sCCH and sCohort studies have focused on efficient inference of relative hazard parameters under the Cox proportional hazards (PH) model (Cox, 1972). Weighted likelihood estimators can be obtained by maximizing the sum of log-likelihood contributions weighted inversely proportional to their probability of selection into the Phase II sample (Chen and Lo, 1999; Kang and Cai, 2009; Gray, 2009; Breslow et al., 2009). The asymptotic theory for IPW estimators are derived in Borgan et al. (2000) and Breslow and Wellner (2007). Under a class of semi-parametric transformation models, Chen and Zucker (2009) proposed an IPW estimator based on an approximate profile likelihood approach.

In this paper, we propose the use of the IPW approach for deriving robust risk prediction rules with sCohort studies under a general class of non-parametric transformation (NPT) models. The NPT model includes many commonly used survival models such as the Cox PH model, the proportional odds model (Murphy et al., 1997), the semi-parametric transformation model (Cheng et al., 1995) and the accelerated failure time model (Wei, 1992) as special cases. The NPT model has been proposed as a robust alternative to the proportional hazards (PH) model and maximum rank correlation (MRC) type estimators have been proposed for estimating the model parameters (Khan and Tamer, 2007; Cai and Cheng, 2008). When the PH assumption holds, the risk scores derived from both the NPT and PH models would achieve the same prediction accuracy asymptotically although the partial likelihood estimator would be more efficient. On the other hand, when the PH assumption fails to hold, the partial likelihood estimate of the risk score may have poor prediction performance (Cai and Cheng, 2008) while the MRC estimator always attains an optimal accuracy with respect to a prediction performance measure, as discussed in Section 2.

Here, we propose to make inference under the NPT model in the presence of the additional complication arising from outcome-dependent sampling from the sCohort design. Specifically we propose an IPW MRC estimator that is asymptotically maximizing an objective function that relates to a C-statistic for survival outcomes (Harrell et al., 1996; Uno et al., 2011). As such, regardless of model adequacy, our proposed estimator always yields a sensible risk score that optimizes a prediction accuracy measure. Thus the NPT model can be seen as a natural choice as a working model when the desire is to build a robust risk score for prediction. That is, even if the NPT model is misspecified, the objective function we propose is a natural measure of the discriminatory capacity of the model. The particular strength of this approach is that the estimation procedure would lead to a risk score with a reasonable prediction performance even when the model is wrong.

Another major challenge arises from analyzing sCohort data due to the weak correlation induced by finite population sampling (Samuelsen et al., 2007; Liu et al., 2012). Variance estimation for IPW estimators was traditionally based on explicit estimation that accounts for the correlation among sampling indicators. However, such explicit estimation for the

proposed IPW MRC estimator is not feasible since the limiting variance form involves complex conditional density functions which are difficult to estimate explicitly, especially under two-phase designs. On the other hand, standard resampling methods such as the bootstrap assume independence between observations and hence fail to correctly approximate the variance. For the original CCH design, Wacholder et al. (1989) proposed the use of a super-population model to construct a variance estimator for the log hazard ratio which is only valid under the null of no marker effect. A modified bootstrap scheme was also proposed to approximate the distribution of the Prentice (1986) pseudo-likelihood estimator. Recently, Huang (2014) proposed a novel bootstrap method for several existing pseudo-likelihood based estimators of the proportional hazards model parameters under the CCH design. However, these existing resampling methods and their associated theoretical justifications are not applicable to general IPW estimators under alternative models, especially for studies with the more complex sCohort sampling.

In this paper, we propose a resampling procedure to estimate the distribution of IPW estimators based on data under a sCohort design. Our proposal is based on the equivalence between the IPW estimators constructed with true selection probabilities under sampling WOR and the IPW estimators constructed with estimated sampling weights under sampling *with replacement* (WR) as detailed in Breslow and Wellner (2007). This motivates us to perturb both the sampling indicator and the sampling probabilities to mimic the effect of the correlation among the sampling weights.

The remainder of the paper is structured as follows. In section 2, we discuss the estimation and inference procedures for the NPT model under the sCohort design. Since many of the biomarker studies have substantial batch effects, we also discuss extensions to make inference under a stratified NPT model to account for batch effects. In section 3 we provide simulation results demonstrating that the proposed estimation and inference procedures perform well under a variety of settings in finite samples. In addition, we demonstrate that the NPT based risk score can outperform the partial likelihood based risk score with respect to risk prediction under model misspecification. We apply our proposed methods to a Danish two-phase cohort study of HDL subtypes for predicting the risk of coronary heart disease in section 4. Section 5 concludes with some discussions.

2. Methods

2.1 Data and Sampling Design

Suppose that the full cohort consists of N subjects that are followed for the observation of failure time T . Due to censoring, for T , we only observe a bivariate vector (X, C) , where $X = T \wedge C$, $I(T < C)$ and C is the censoring time. To develop a risk model for predicting T , routine clinical markers $\mathbf{U} = (U_1, \dots, U_q)^T$ are collected for the entire cohort but expensive novel markers $\mathbf{Z} = (Z_1, \dots, Z_p)^T$ are measured only for subjects selected into the Phase II sample. We also allow the sampling to depend on an exposure matching variable $\mathcal{J}$ taking values $\mathcal{J} = 1, \dots, J$. Thus the underlying full cohort data consists of N independent and identically distributed (i.i.d.) random vectors from N subjects:

$$\mathcal{D} = \{\mathbf{D}_i = (X_i, \Delta_i, \mathcal{J}_i, \mathbf{W}_i^\top)^\top, i=1, \dots, N\}, \quad \text{where} \quad \mathbf{W}_i = (\mathbf{U}_i^\top, \mathbf{Z}_i^\top)^\top.$$

Let $V_i = \mathbb{I}_{i \in \text{Phase II sample}}$ indicate whether subject i is selected into the Phase II sample, where V_{1i} and V_{0i} indicate that subject i was selected into the Phase II sample as either a case or non-case respectively. For $\kappa = 0, 1$ and $j = 1, \dots, J$, let $\mathcal{Q}_{\kappa j, i}$ be the matching category indicator taking value 1 if $i = \kappa$ and $\mathcal{J}_i = j$. Within each category indexed by (κ, j) , suppose $n_{\kappa j}$ subjects are sampled WOR from a total of $N_{\kappa j} = \sum_{i=1}^N \mathcal{Q}_{\kappa j, i}$ subjects in this category, where $n_{\kappa j}$ is pre-specified by design.

Conditional on the full-cohort data $\mathcal{D}$, the probability for subject i being sampled into the Phase II sample is

$$P(V_i=1|\mathcal{D}) = \hat{p}_i = \sum_{\kappa=0}^1 \sum_{j=1}^J \mathcal{Q}_{\kappa j, i} \hat{\pi}_{\kappa j}, \quad \text{where} \quad \hat{\pi}_{\kappa j} = n_{\kappa j} / N_{\kappa j}.$$

Define the IPW weight for the i th subject as $\hat{w}_i = V_i / \hat{p}_i$, and let

$\mathcal{F} = \{(X_i, \Delta_i, \mathcal{J}_i, \mathbf{U}_i^\top, V_i \mathbf{Z}_i^\top, V_i)^\top, i=1, \dots, N\}$ denote the entire set of observed data. Note that when $V_i = 0$, $V_i \mathbf{Z}_i = \mathbf{0}$ and the value of $\mathbf{Z}_i$ is not observed.

2.2 Model and Estimation

We consider the following NPT failure time model:

$$h_0(T_i) = \beta_0^\top \mathbf{W}_i + \varepsilon_i, \quad (1)$$

where $h_0(\cdot)$ is an unspecified smooth increasing function and ε_i are i.i.d. random error with an unknown distribution. Obviously, β_0 is only identifiable up to a scalar and thus to ensure estimability, we consider estimation over a compact subset of the parameter space $\Omega_1 = \{\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)^\top : \|\boldsymbol{\beta}\|_1 = 1\}$, where $\|\boldsymbol{\beta}\|_1 = \sum_j |\beta_j|$ is the L_1 norm of $\boldsymbol{\beta}$.

With full cohort survival data, Khan and Tamer (2007) proposed an MRC estimator as the maximizer of

$$Q_n(\boldsymbol{\beta}) = \frac{1}{n(n-1)} \sum_{i \neq j} I(\beta^\top \mathbf{W}_i < \beta^\top \mathbf{W}_j, X_i < X_j) \Delta_i, \quad (2)$$

which converges in probability uniformly in $\boldsymbol{\beta}$ to

$$q(\beta) = P(\beta^\top \mathbf{W}_1 < \beta^\top \mathbf{W}_2 | X_1 < X_2, \Delta_1 = 1).$$

In the absence of censoring, $q(\beta)$ corresponds to the Kendall's rank correlation between T and $\beta^\top \mathbf{W}$. In the presence of censoring, $q(\beta) = P(\beta^\top \mathbf{W}_1 < \beta^\top \mathbf{W}_2 | T_1 < T_2, T_1 \leq C_1 \wedge C_2)$ corresponds to the Harrell's C-statistic and hence still possesses the interpretation as a concordance measure (Harrell et al., 1996; Uno et al., 2011). Under independent censoring when $C \perp \{T, \mathbf{W}\}$, $q(\beta)$ can be viewed as a weighted area under an incident/dynamic ROC curve (Heagerty and Zheng, 2005) with

$$q(\beta) = \int \text{AUC}_\beta(t) w(t) dt,$$

where $\text{AUC}_\beta(t) = P(\beta^\top \mathbf{W}_1 < \beta^\top \mathbf{W}_2 | T_1 = t, T_2 > t)$ is the incident/dynamic time-specific AUC, and $w(t) = f(t)S(t)G(t)^2 / \int f(u)S(u)G(u)^2 du$ is the weight function that involves the marginal survival and censoring distributions. Here $f(t)$ is the density function of T , $S(t) = P(T > t)$ and $G(t) = P(C > t)$.

However, under the sCohort design, $\mathbf{Z}$ is only observed for subjects in the Phase II sample. To account for non-random missingness due to sampling, we propose an IPW MRC estimator of β_0 , $\hat{\beta} = \arg\max_{\beta \in \Omega_1} \hat{Q}(\beta)$, where

$$\hat{Q}(\beta) = \frac{\sum_{ij} \hat{w}_i \hat{w}_j I(\beta^\top \mathbf{W}_i < \beta^\top \mathbf{W}_j) I_{ij}}{\sum_{ij} \hat{w}_i \hat{w}_j I_{ij}}. \quad (3)$$

where $I_{ij} = I(X_i < X_j)$. In Web Appendix A we show that $\hat{Q}(\beta)$ converges in probability uniformly to $q(\beta)$, the same limiting objective function corresponding to the full cohort data, which implies the consistency of $\hat{\beta}$ for β_0 . Consequently, the empirical C-statistic estimate $\hat{Q}(\hat{\beta})$ converges in probability to $q(\beta_0)$.

2.3 Inference via Resampling

To construct confidence intervals for β_0 we show in Web Appendix A that $\sqrt{n}(\hat{\beta} - \beta_0)$ converges in distribution to a multivariate normal with mean $\mathbf{0}$ and some covariance matrix Σ . Establishing the normality of $\hat{\beta}$ is complicated first by the fact the IPW weights $\hat{w}_i$ induce correlation in what would otherwise be a U-statistic, so that standard theory for i.i.d. U-processes cannot be directly applied. A second complication is the fact that the objective function $\hat{Q}_n(\beta)$ is not smooth in β ; to address this we approximate it with a smooth function $\hat{Q}^\dagger(\beta)$, defined in Web Appendix A, which is differentiable in β and has the same limiting function $q(\beta)$ as $\hat{Q}(\beta)$. Subsequently, we show that $n^{1/2}(\hat{\beta}^\dagger - \hat{\beta}) = o_p(1)$ and $n^{1/2}(\hat{\beta}^\dagger - \beta_0)$ converges in distribution to $N(0, \Sigma)$, where $\hat{\beta}^\dagger = \arg\max_{\beta} \hat{Q}^\dagger(\beta)$. This implies the weak convergence of $\sqrt{n}(\hat{\beta} - \beta_0) \rightarrow N(0, \Sigma)$.

Explicitly estimating Σ would require the estimation of conditional density functions, which is difficult both statistically and computationally. As such, it would be preferable to employ a simple resampling technique such as the bootstrap to approximate Σ . However, the standard bootstrap ignores negative between-subject correlations induced by sampling and consequently provides consistently inflated estimates of the variance, which may substantially reduce power in what is already a low-resource setting. Here, we propose a novel resampling method, which is able to recover the effect of correlation among the V_i 's by both perturbing the sampling indicators and estimating the sampling probabilities. Specifically, let $B_1, \dots, B_N$ be i.i.d. non-negative random variables with both mean and variance equal to one, and generated independently of the observed data $\mathcal{A}$. Then we perturb the sampling weights as

$$\hat{w}_i^* = \frac{V_i B_i}{\hat{p}_i^*}, \quad \hat{p}_i^* = \sum_{\kappa_j} \mathcal{I}_{\kappa_j, i} \hat{\pi}_{\kappa_j}^*, \quad \text{and} \quad \hat{\pi}_{\kappa_j}^* = \frac{\sum_{i=1}^N B_i V_i \mathcal{I}_{\kappa_j, i}}{\sum_{i=1}^N B_i \mathcal{I}_{\kappa_j, i}}.$$

As described in section 6 of Breslow and Wellner (2007), using estimated sampling fractions under a Bernoulli (with replacement) sampling scheme constructs an IPW estimator asymptotically equivalent to the one obtained using the true sampling fractions under a finite population WOR sampling scheme. Thus, to approximate the distribution of our IPW estimator obtained under WOR sampling with true weights $\hat{w}_i$, we carry out our resampling procedure independent perturbations, which resembles Bernoulli sampling, but estimate the true sampling fractions $\hat{\pi}_{\kappa_j}$ in $\hat{w}_i$ as $\hat{\pi}_{\kappa_j}^*$ to create the perturbed weights $\hat{w}_i^*$. The variations in $\hat{\pi}_{\kappa_j}^*$ that mimics the unconditional induces a correlation among $\{\hat{w}_i^*\}$ conditional on $\mathcal{A}$ correlation structure among V_i . With these perturbed weights, we obtain a stochastically perturbed counterpart of $\hat{\beta}$ defined as $\hat{\beta}^* = \operatorname{argmax}_{\beta \in \Omega_1} \hat{Q}^*(\beta)$,

$$\hat{Q}^*(\beta) = \frac{\sum_{ij} \hat{w}_i^* \hat{w}_j^* I(\beta^\top \mathbf{W}_i < \beta^\top \mathbf{W}_j) I_{ij}}{\sum_{ij} \hat{w}_i^* \hat{w}_j^* I_{ij}} \quad (4)$$

In Web Appendix B we provide justifications for the resampling method for a weighted generic statistic. Using these justifications along with similar arguments given in the Web Appendix A, we may show that $\sqrt{n}(\hat{\beta}^* - \hat{\beta})$ conditional on the observed data $\mathcal{A}$ converges in distribution to $N(\mathbf{0}, \Sigma)$, the same limiting distribution as $\sqrt{n}(\hat{\beta} - \beta_0)$. Similar arguments can be used to show that the conditional distribution of $\sqrt{n}\{\hat{Q}^*(\hat{\beta}^*) - \hat{Q}(\hat{\beta})\}$ given $\mathcal{A}$ can be used to approximate the unconditional distribution of $\sqrt{n}\{\hat{Q}(\hat{\beta}) - q(\beta_0)\}$.

2.4 Stratified NPT Model

Batch effects are frequently encountered in many biomarker studies when biological samples are processed in multiple batches in varying experimental conditions and/or by different laboratory technicians. The presence of batch effects introduces non-biological experimental

variation to the data. We assume that there are K finite number of batches and let $S_j \in \{1, \dots, K\}$ index the batch that the i th sample belongs to. To account for the heterogeneity between the batches, we consider the following stratified NPT model:

$$h_{S_i}(T_i) = \beta_0^\top \mathbf{W}_i + \varepsilon_{S_i}, \quad (5)$$

where ε_{S_i} represents the error term that corresponds to observations from batch S_i . Under the stratified NPT model, it is reasonable to make comparisons within each batch and then combine evidence across the batches. Therefore, we propose to estimate β_0 as $\hat{\beta}_S = \arg \max_{\beta \in \Omega_1} \hat{Q}(\beta)$, the maximizer of the stratified C-statistic

$$\hat{Q}(\beta) = \frac{\sum_{i,j} \hat{w}_i \hat{w}_j I(\beta^\top \mathbf{W}_i < \beta^\top \mathbf{W}_j) I_{ij} I(S_i = S_j)}{\sum_{i,j} \hat{w}_i \hat{w}_j I_{ij} I(S_i = S_j)} = \sum_{k=1}^K \hat{\omega}_k \hat{Q}_k(\beta),$$

where

$$\hat{\omega}_k = \frac{\sum_{S_i=S_j=k} \hat{w}_i \hat{w}_j I_{ij}}{\sum_{k=1}^K \sum_{S_i=S_j=k} \hat{w}_i \hat{w}_j I_{ij}} \quad \text{and} \quad \hat{Q}_k(\beta) = \frac{\sum_{S_i=S_j=k} \hat{w}_i \hat{w}_j I(\beta^\top \mathbf{W}_i < \beta^\top \mathbf{W}_j) I_{ij}}{\sum_{S_i=S_j=k} \hat{w}_i \hat{w}_j I_{ij}}.$$

To approximate the distribution of $\hat{\beta}_S$, we perturb the stratified C-statistic as

$$\hat{Q}^*(\beta) = \frac{\sum_{i,j} \hat{w}_i^* \hat{w}_j^* I(\beta^\top \mathbf{W}_i < \beta^\top \mathbf{W}_j) I_{ij} I(S_i = S_j)}{\sum_{i,j} \hat{w}_i^* \hat{w}_j^* I_{ij} I(S_i = S_j)} = \sum_{k=1}^K \hat{\omega}_k^* \hat{Q}_k^*(\beta),$$

where

$$\hat{\omega}_k^* = \frac{\sum_{S_i=S_j=k} \hat{w}_i^* \hat{w}_j^* I_{ij}}{\sum_{k=1}^K \sum_{S_i=S_j=k} \hat{w}_i^* \hat{w}_j^* I_{ij}} \quad \text{and} \quad \hat{Q}_k^*(\beta) = \frac{\sum_{S_i=S_j=k} \hat{w}_i^* \hat{w}_j^* I(\beta^\top \mathbf{W}_i < \beta^\top \mathbf{W}_j) I_{ij}}{\sum_{S_i=S_j=k} \hat{w}_i^* \hat{w}_j^* I_{ij}}.$$

3. Simulation

In this section we present results from extensive simulation studies to assess the performance of the proposed estimation procedure under differing sample sizes and models. For simplicity, we considered the setting without additional clinical variables. For the sCohort sampling design, we considered an additional 4-level matching exposure factor $\mathcal{J}$ used for sampling. Within each matching group j defined by $\mathcal{J}$, we sampled all cases with $n_{1j} = N_{1j}$ and subsampled $n_{0j} = mn_{1j}$ controls, for $m = 1$ or 3 . Throughout, the censoring time was generated as the minimum of a $\text{Gamma}(0.5, 0.5) + 0.1$ variable and a $\text{Uniform}(0.5, 1)$.

We considered cohort sizes of $N = 5000$ and $N = 10000$. For each configuration, we simulated 1,000 datasets to summarize the performance of the estimators. For each simulated dataset, 500 perturbations were used to calculate the standard errors and confidence intervals.

We first considered the case when the NPT model holds and generated survival time T_i from a Weibull distribution with scale parameter equal to 1 and shape parameter a_i :

$$\log a_i = (5.5 + Z_{i,1} + 0.5Z_{i,2} - Z_{i,3} + \mu_{S_i} b_i) / 2,$$

where b_i is generated from the standard normal distribution, and $S_i \in \{1, 2, 3\}$ indexes the batch. We considered setting (I) with no batch effects by letting $\mu_{S_i} = 0$ and setting (II) with $\mu_{S_i} = S_i - 1$. The event rates for setting (I) and setting (II) are 4.8% and 5.6%, respectively. We simulated $Z_{i,1}$, $Z_{i,2}$ and $Z_{i,3}$ from independent standard normal. The exposure matching variable used for two-phase stratified sampling was generated as

$\mathcal{J}_i = I(\mathcal{J}_i^\dagger > 0.25) + I(\mathcal{J}_i^\dagger > 0.5) + I(\mathcal{J}_i^\dagger > 0.75) + 1$, where $\mathcal{J}_i^\dagger = \Phi(Z_{i,1} + Z_{i,2})$, and $\Phi(\cdot)$ is the standard normal cumulative distribution function. We generate $\mathcal{J}$ to mimic the scenario that an inexpensive exposure variable, correlated with the expensive novel markers, is available to assist in sampling to improve the efficiency. In general, the higher correlation $\mathcal{J}$ has with $\mathbf{Z}$, the more efficiency gain one may expect from the stratified sampling (Borgan and Olsen, 1999; Borgan et al., 2000; Breslow et al., 2009). For comparison, we also obtained estimates of the model parameters and C-statistic from fitting corresponding Cox models and the estimated regression coefficients are normalized to have unit L_1 norm.

Tables 1 and 2 summarize the simulation results for the Cox and NPT models under settings (I) and (II), respectively. For setting (II), stratified Cox and stratified NPT models were fit to account for the batch effect. The proposed NPT estimators of both β_0 and the C-statistic $q(\beta_0)$ have negligible bias and the bias decreases as the number of available samples increases. The resampling procedure works well in approximating the distribution of the proposed estimators. The average of the estimated standard errors are close to the empirical standard errors. The empirical coverage probabilities of the 95% confidence intervals (CI) are close to the nominal level. As shown in Tables 1 and 2, the rescaled NPT estimators and the partial likelihood estimators are both consistent for the underlying standardized coefficient since the NPT model holds and the Cox model is only misspecified in the link (Li and Duan, 1989). Interestingly, the NPT estimators are more efficient than the partial likelihood estimators since the PH assumption fails. The standard error estimates are unstable for the Cox model in these rare event settings although they do get better when the event rate increases (results not shown). The estimated C-statistics and their standard errors are similar for the two approaches.

To demonstrate the robustness of the proposed IPW MRC estimator under model misspecification, we examined the prediction performance of the estimated risk score and compared to the risk score derived from the Cox model. Specifically, we generated the survival time from the following model with interactive effects:

$$\log T_i = (4 - 0.5Z_{i,1} + 0.5Z_{i,2} - 0.5Z_{i,1}Z_{i,2} + \varepsilon_i) / 4,$$

where ε_i was generated from the standard extreme value distribution, $Z_{i,1}$ and $Z_{i,2}$ were generated from independent normal mixtures, $I_i \sim N(-2, 1) + (1 - I_i) \sim N(2, 1)$, and $I_i \sim \text{Bernoulli}$ with success probability 0.2. For simplicity, no batch effects or matching variable was considered for the misspecified case. The event rate of the population is approximately 5%. For each of the four settings $N = (5000, 10000)$ and $m = (1, 3)$, we fit both the Cox and NPT models with $Z_{i,1}$ and $Z_{i,2}$ as linear predictors omitting the interaction terms. To evaluate the predictive ability of the Cox and NPT models, we calculated the C-statistics using a large independent validation data of size 1,000,000 such that the variability in the observed C-statistic estimates is due to the variability in estimating β from the two procedures. As shown in Table 3, the proposed estimates of the risk score derived under the NPT model showed superior predictive capacity with average C-statistics around 0.73 while the PH model based risk score has an average of 0.68. The PH model based estimator also exhibits substantially higher variability with resulting C-statistic estimates lower than the null value of 0.5 for some realizations.

4. Application

We applied our proposed methods to the Danish Diet, Cancer and Health (DCH) study to assess the effect of high-density lipoprotein (HDL) subtypes and other traditional clinical markers on the risk of coronary heart disease (CHD). Initiated in 1993, the DCH study enrolled a total of 57,053 Danish born residents (27,178 men and 29,875 women) between ages of 50 and 64 without prior diagnosis of cancer in the Danish Cancer Registry (Tjønneland et al., 2007; Jensen et al., 2008). Blood was drawn from the participants at baseline and stored for future research. The primary goal of the DCH study is to study the association between diet and the risk of cancer and other chronic diseases including type 2 diabetes and cardiovascular diseases. However, with biospecimens linked to detailed outcome data on the study participants, the DCH study provides unique opportunities to evaluate novel molecular markers for risk prediction. For example, Jensen et al. (2009) studied the effect of a genetic variant on the risk of coronary heart disease (CHD) using the DCH study. Recently, Jensen et al. (2012) used the NHS and HPFS to evaluate novel lipid biomarkers representing subtypes of HDL cholesterol (HDL-C), HDL-C with apoC-III (HDL-C₊) and without apoC-III (HDL-C₋), on the risk of CHD. Interestingly, it was shown that the two HDL-C subtypes have differential association with the CHD risk.

In this paper, we assess the effect of HDL-C subtypes on the CHD risk using the DCH study. The HDL-C subtypes were measured on a Phase II sample consisting of 3,863 participants and sampled by matching on gender. All CHD cases were included and a subset of controls were sampled from male and female categories separately. The number of cases and controls in each of the strata defined by case-control status and gender is shown in Table 4, which we use as the basis for calculating the sampling weights. Among the 57,053 DCH participants, 1,506 subjects had prevalent CHD, prevalent cancer, or missingness and are therefore excluded from both the analysis and weight calculation. Among the Phase II sample, 3,633

(94%) participants had complete records on the two HDL-C subtypes and other demographical and clinical information (age, gender, education, smoking, alcohol, hypertension, and BMI). The missingness within the Phase II sample was treated as missing completely at random. Since the biomarkers were measured in 44 different batches, stratified NPT models were fit to account for the batch effects.

Table 5 shows the estimated model parameters for the stratified Cox and stratified NPT models, respectively. Consistent with the findings in Jensen et al. (2012) based on the NHS and HPFS, HDL-C₋ is inversely associated with CHD risk ($P < 0.001$ from both Cox and NPT models). However, no association between HDL-C₊ and CHD risk was found (Cox model: $P = 0.812$; NPT model: $P = 0.887$) in the DCH cohort. Testing for the difference between the effects of HDL-C₋ and HDL-C₊, we obtain a p-value of 0.002 from the Cox model and a p-value of less than 0.001 from the NPT model, suggesting that the two subtypes of HDL-C may have different anti-atherogenic potentials. The estimated C-statistics for the resulting stratified Cox and stratified NPT models were $\hat{Q} = 0.742$ (95% CI: [0.725, 0.759]) and $\hat{Q} = 0.743$ (95% CI: [0.725, 0.760]) respectively, suggesting that these two models have very similar accuracy in predicting CHD risks.

5. Discussion

In this paper, we proposed IPW procedures to derive and evaluate robust risk scores for predicting survival outcomes under a flexible NPT model with data from two-phase sCohort studies. We demonstrate through numerical studies that the risk score has similar prediction performance as those derived from the Cox model when the PH assumption holds and outperforms the Cox model based score when the PH assumption fails to hold. One main challenge arising from analyzing sCohort data is due to the correlation induced by finite population sampling. We propose a novel resampling procedure that can mimic such correlation structures.

Our resampling method differs from the nested case-control setting (Cai and Zheng, 2011, 2013) due to the difference in the calculation of the perturbed sampling probability. The perturbation also differs from the Huang (2014) method as our procedure can accommodate the two-phase sCohort design while the bootstrap method given in Huang (2014) is only applicable to the original CCH sampling design. Following similar principles, we expect that our proposed resampling technique can be easily modified to accommodate the CCH and sCCH designs. Details on the modification and its justification warrant further research.

Like the estimator of Khan and Tamer (2007), our proposed estimator $\hat{\beta}$ depends on the study-specific censoring distribution under model misspecification. Assuming that $C_i \perp (T_i, \mathbf{W}_i)$, we could maximize a sampling probability reweighted counterpart of the objective function proposed in Cai and Cheng (2008). The limit of this estimator will be censoring free even under model misspecification. However, when censoring does depend on $\mathbf{W}$ and the NPT model is correctly specified, this estimator may not be consistent while the estimator proposed in this paper is consistent. Under covariate independent censoring, one may also estimate β consistently up to a scalar by fitting a binary regression model $P(C_i = 1 | \mathbf{W}_i) = g(\alpha - \beta^T \mathbf{W}_i)$ under specific assumptions for the covariate distribution given in Li and

Duan (1989). However, we expect that the estimator based on the binary regression would be less efficient than the proposed estimator based on rank regression which makes use of the observed survival time information.

The C-statistic considered in this paper is a measure of predictive performance of a statistical model and hence can potentially be used as a criterion for model selection. However, like R^2 for ordinary linear regression, caution is needed when using such estimates for selecting nested models due to overfitting. To overcome the overfitting bias, an adjusted C-statistic with a penalty term of model complexity needs to be developed to guide the model selection. One approach is to use cross-validation to estimate $(\alpha\beta_0)$ such that training data used for estimating β_0 are independent of validation data used for estimating Q (Uno et al., 2007; Jiang and Simon, 2007). Other criteria such as Brier score can also be used for model evaluation. However, unlike semiparametric transformation models, the NPT model does not attempt to estimate the error distribution or the monotone transformation function, and the $\hat{\beta}$ from the NPT model does not directly lead to an estimate of the absolute risk function. In order to produce a risk estimate under the NPT model, one may estimate $R_{ik}(t) = P(T_i \leq t | \hat{\beta}^T \mathbf{W}_i = u)$ as $\hat{R}_{ik}(t)$ via non-parametric kernel smoothing and then predict the risk of developing an event by time t_0 for a future subject with $\mathbf{W}$ as $\hat{R}_{\hat{\beta}}^T \mathbf{W}(t_0)$. Subsequently, one may evaluate the prediction performance of the NPT model with respect to both discrimination and calibration based on $\hat{R}_{\hat{\beta}}^T \mathbf{W}(t_0)$. We expect the NPT model to outperform the Cox model with respect to prediction performance under a variety of misspecified models. Further research is warranted to investigate the bias and variance trade-off under different choices of survival models.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Payne and Yang contributed equally to this work. This research was partially supported by NIH grants R01 NS089638, R01 GM085047, P50 MH106999, and R01 GM079330. The Diet, Cancer and Health Study is supported by the Danish Research Council and the Danish Cancer Society. HDL apoC-III measurements were funded by An Elite Research Award from the Danish Research Council for Independent Research.

References

- Anderson GL, Manson J, Wallace R, Lund B, Hall D, Davis S, Shumaker S, Wang CY, Stein E, Prentice RL. Implementation of the Women's Health Initiative study design. *Annals of Epidemiology*. 2003; 13:S5–17. [PubMed: 14575938]
- Borgan O, Langholz B, Samuelsen SO, Goldstein L, Pogoda J. Exposure stratified case-cohort designs. *Lifetime Data Analysis*. 2000; 6:39–58. [PubMed: 10763560]
- Borgan O, Olsen EF. The efficiency of simple and counter-matched nested case-control sampling. *Scandinavian Journal of Statistics*. 1999; 26:493–509.
- Breslow NE, Lumley T, Ballantyne CM, Chambless LE, Kulich M. Improved horvitz–thompson estimation of model parameters from two-phase stratified samples: Applications in epidemiology. *Statistics in Biosciences*. 2009; 1:32–49. [PubMed: 20174455]
- Breslow NE, Wellner JA. Weighted likelihood for semiparametric models and two-phase stratified samples, with application to Cox regression. *Scandinavian Journal of Statistics*. 2007; 34:86–102.

- Cai T, Cheng S. Robust combination of multiple diagnostic tests for classifying censored event times. *Biostatistics*. 2008; 9:216–233. [PubMed: 18056687]
- Cai T, Zheng Y. Nonparametric evaluation of biomarker accuracy under nested case-control studies. *Journal of the American Statistical Association*. 2011; 106:569–580. [PubMed: 22844169]
- Cai T, Zheng Y. Resampling procedures for making inference under nested case-control studies. *Journal of the American Statistical Association*. 2013; 108:1532–1544. [PubMed: 24436503]
- Chen K, Lo SH. Case-cohort and case-control analysis with cox's model. *Biometrika*. 1999; 86:755–764.
- Chen YH, Zucker DM. Case-cohort analysis with semiparametric transformation models. *Journal of Statistical Planning and Inference*. 2009; 139:3706–3717.
- Cheng S, Wei L, Ying Z. Analysis of transformation models with censored data. *Biometrika*. 1995; 82:835–845.
- Colditz GA, Manson JE, Hankinson SE. The Nurses' Health Study: 20-year contribution to the understanding of health among women. *Journal of Women's Health*. 1997; 6:49–62.
- Cox DR. Regression models and life-tables. *Journal of the Royal Statistical Society. Series B (Methodological)*. 1972; 34:187–220.
- Folsom A, Aleksic N, Catellier D, Juneja H, Wu K. C-reactive protein and incident coronary heart disease in the Atherosclerosis Risk In Communities (ARIC) study. *American Heart Journal*. 2002; 144:233–238. [PubMed: 12177639]
- Gray RJ. Weighted analyses for cohort sampling designs. *Lifetime Data Analysis*. 2009; 15:24–40. [PubMed: 18712477]
- Gunter MJ, Hoover DR, Yu H, Wassertheil-Smoller S, Manson JE, Li J, Harris TG, Rohan TE, Xue X, Ho GYF, Einstein MH, Kaplan RC, Burk RD, Wylie-Rosett J, Pollak MN, Anderson G, Howard BV, Strickler HD. A prospective evaluation of Insulin and Insulin-like growth factor-I, endogenous estradiol, and risk of colorectal cancer in postmenopausal women. *Cancer Epidemiology, Biomarkers and Prevention*. 2008; 17:921–929.
- Gunter MJ, Hoover DR, Yu H, Wassertheil-Smoller S, Rohan TE, Manson JE, Li J, Ho GY, Xue X, Anderson GL, Kaplan RC, Harris TG, Howard BV, Wylie-Rosett J, Burk RD, Strickler HD. Insulin, Insulin-like growth factor-I, and risk of breast cancer in postmenopausal women. *Journal of the National Cancer Institute*. 2009; 101:48–60. [PubMed: 19116382]
- Harrell FE, Lee KL, Mark DB. Multivariate prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Statistics in Medicine*. 1996; 15:361–387. [PubMed: 8668867]
- Heagerty PJ, Zheng Y. Survival model predictive accuracy and ROC curves. *Biometrics*. 2005; 61:92–105. [PubMed: 15737082]
- Herder C, Baumert J, Thorand B, Martin S, Löwel H, Kolb H, Koenig W. Chemokines and incident coronary heart disease: results from the MON-ICA/KORA Augsburg case-cohort study, 1984–2002. *Arteriosclerosis, Thrombosis, and Vascular Biology*. 2006; 26:147–152.
- Huang Y. Bootstrap for the case cohort design. *Biometrika*. 2014; 101:465–476. [PubMed: 25170175]
- Jensen MK, Chiuve SE, Rimm EB, Dethlefsen C, Tjønneland A, Joensen AM, Overvad K. Obesity, behavioral lifestyle factors, and risk of acute coronary events. *Circulation*. 2008; 117:3062–3069. [PubMed: 18541738]
- Jensen MK, Rimm EB, Furtado JD, Sacks FM. Apolipoprotein C-III as a potential modulator of the association between HDL-cholesterol and incident coronary heart disease. *Journal of the American Heart Association*. 2012; 1:jah3-e000232. [PubMed: 23130121]
- Jensen MK, Rimm EB, Mukamal KJ, Edmondson AC, Radar DJ, Vogel U, Tjønneland A, Sorensen TIA, Schmidt EB, Overvad K. The T1111 variant in the endothelial lipase gene and risk of coronary heart disease in three independent populations. *European Heart Journal*. 2009; 30:1584–1589. [PubMed: 19411665]
- Jiang W, Simon R. A comparison of bootstrap methods and an adjusted bootstrap approach for estimating the prediction error in microarray classification. *Statistics in Medicine*. 2007; 26:5320–5334. [PubMed: 17624926]
- Kang S, Cai J. Marginal hazards model for case-cohort studies with multiple disease outcomes. *Biometrika*. 2009; 96:887–901. [PubMed: 23946547]

- Khan S, Tamer E. Partial rank estimation of duration models with general forms of censoring. *Journal of Econometrics*. 2007; 136:251–280.
- Li KC, Duan N. Regression analysis under link violation. *The Annals of Statistics*. 1989; 17:1009–1052.
- Liu D, Cai T, Zheng Y. Evaluating the predictive value of biomarkers with stratified case-cohort design. *Biometrics*. 2012; 68:1219–1227. [PubMed: 23173848]
- Murphy S, Rossini A, van der Vaart AW. Maximum likelihood estimation in the proportional odds model. *Journal of the American Statistical Association*. 1997; 92:968–976.
- Prentice RL. A case-cohort design for epidemiologic cohort studies and disease prevention trials. *Biometrika*. 1986; 73:1–11.
- Rimm E, Giovannucci E, Willett W, Colditz G, Ascherio A, Rosner B, Stampfer M. Prospective study of alcohol consumption and risk of coronary disease in men. *The Lancet*. 1991; 338:464–468.
- Samuelson SO, ÅNestad H, Skrdal A. Stratified case-cohort analysis of general cohort sampling designs. *Scandinavian Journal of Statistics*. 2007; 34:103–119.
- Tjønneland A, Olsen A, Boll K, Stripp C, Christensen J, Engholm G, Overvad K. Study design, exposure variables, and socioeconomic determinants of participation in Diet, Cancer and Health: A population-based prospective cohort study of 57,053 men and women in Denmark. *Scandinavian Journal of Public Health*. 2007; 35:432–441. [PubMed: 17786808]
- Uno H, Cai T, Pencina MJ, D'Agostino RB, Wei LJ. On the C-statistics for evaluating overall adequacy of risk prediction procedures with censored survival data. *Statistics in Medicine*. 2011; 30:1105–1117. [PubMed: 21484848]
- Uno H, Cai T, Tian L, Wei LJ. Evaluating prediction rules for t-year survivors with censored regression models. *Journal of the American Statistical Association*. 2007; 102:527–537.
- van Oijen M, Hofman A, Soares H, Koudstaal PJ, Breteler MMB. Plasma $A\beta_{1-40}$ and $A\beta_{1-42}$ and the risk of dementia: a prospective case-cohort study. *The Lancet Neurology*. 2006; 5:655–660. [PubMed: 16857570]
- Wacholder S, Gail M, Pee D, Brookmeyer R. Alternative variance and efficiency calculations for the case-cohort design. *Biometrika*. 1989; 76:117–123.
- Wei L. The accelerated failure time model: a useful alternative to the cox regression model in survival analysis. *Statistics in Medicine*. 1992; 11:1871–1879. [PubMed: 1480879]

Table 1

Finite sample properties of the estimators for setting (I) in the absence of batch effects (ESE: empirical standard errors; ASE: average of standard errors; CovP: coverage probability).

N	m	Truth	Cox				NPT			
			Bias	ESE	ASE	CovP	Bias	ESE	ASE	CovP
5000	1	$\beta_{0,1} = 0.400$	-0.005	0.103	0.077	0.851	0.002	0.051	0.050	0.951
		$\beta_{0,2} = 0.200$	-0.023	0.114	0.081	0.831	-0.006	0.054	0.051	0.929
		$\beta_{0,3} = -0.400$	-0.019	0.106	0.077	0.837	-0.004	0.053	0.049	0.911
	3	$Q = 0.860$	-0.006	0.024	0.024	0.947	0.002	0.021	0.020	0.939
		$\beta_{0,1} = 0.400$	-0.002	0.071	0.059	0.899	-0.001	0.036	0.037	0.946
		$\beta_{0,2} = 0.200$	-0.010	0.075	0.063	0.913	-0.001	0.037	0.038	0.947
		$\beta_{0,3} = -0.400$	-0.012	0.074	0.060	0.880	-0.002	0.037	0.037	0.945
	10000	$Q = 0.860$	-0.002	0.017	0.017	0.954	0.002	0.016	0.015	0.945
		$\beta_{0,1} = 0.400$	-0.000	0.078	0.061	0.881	0.001	0.036	0.036	0.945
		$\beta_{0,2} = 0.200$	-0.016	0.079	0.065	0.886	-0.005	0.037	0.036	0.926
5000	1	$\beta_{0,3} = -0.400$	-0.015	0.079	0.062	0.874	-0.005	0.036	0.035	0.933
		$Q = 0.860$	-0.004	0.016	0.017	0.963	0.001	0.015	0.015	0.949
		$\beta_{0,1} = 0.400$	-0.002	0.050	0.044	0.914	0.000	0.025	0.026	0.959
	3	$\beta_{0,2} = 0.200$	-0.007	0.053	0.047	0.916	-0.003	0.025	0.026	0.954
		$\beta_{0,3} = -0.400$	-0.009	0.052	0.046	0.920	-0.002	0.025	0.026	0.968
		$Q = 0.860$	-0.001	0.011	0.011	0.947	0.001	0.011	0.011	0.942

Finite sample properties of the estimators for setting (II) in the presence of batch effects (ESE: empirical standard errors; ASE: average of standard errors; CovP: coverage probability).

Table 2

N	m	Truth	Cox				NPT			
			Bias	ESE	ASE	CovP	Bias	ESE	ASE	CovP
5000	1	$\beta_{0,1} = 0.400$	-0.008	0.087	0.066	0.865	0.003	0.055	0.057	0.942
		$\beta_{0,2} = 0.200$	-0.019	0.093	0.069	0.856	-0.006	0.059	0.056	0.930
		$\beta_{0,3} = -0.400$	-0.024	0.090	0.066	0.839	-0.003	0.055	0.054	0.946
	3	$Q = 0.820$	-0.002	0.023	0.023	0.955	0.004	0.022	0.021	0.943
		$\beta_{0,1} = 0.400$	-0.002	0.061	0.053	0.907	-0.000	0.041	0.043	0.964
		$\beta_{0,2} = 0.200$	-0.012	0.064	0.056	0.917	-0.004	0.042	0.043	0.947
		$\beta_{0,3} = -0.400$	-0.014	0.061	0.054	0.911	-0.004	0.041	0.042	0.943
	10000	$Q = 0.820$	-0.001	0.017	0.017	0.961	0.003	0.017	0.016	0.949
		$\beta_{0,1} = 0.400$	-0.002	0.066	0.051	0.884	-0.000	0.038	0.040	0.962
		$\beta_{0,2} = 0.200$	-0.015	0.066	0.054	0.887	-0.003	0.038	0.039	0.952
10000	1	$\beta_{0,3} = -0.400$	-0.017	0.068	0.052	0.852	-0.003	0.038	0.039	0.946
		$Q = 0.820$	-0.001	0.016	0.016	0.944	0.002	0.016	0.015	0.932
		$\beta_{0,1} = 0.400$	-0.001	0.045	0.039	0.909	-0.001	0.028	0.029	0.953
	3	$\beta_{0,2} = 0.200$	-0.007	0.046	0.041	0.928	-0.002	0.030	0.030	0.944
		$\beta_{0,3} = -0.400$	-0.008	0.045	0.040	0.919	-0.003	0.028	0.029	0.946
		$Q = 0.820$	-0.000	0.012	0.012	0.950	0.001	0.012	0.012	0.941

Table 3
C-statistic for Cox and NPT under model misspecification based on 1000 simulated datasets.

N	m	Cox		NPT	
		Mean (%)	SD (%)	Mean (%)	SD (%)
5000	1	68.0	2.11	72.9	0.21
	3	67.9	1.78	73.0	0.18
10000	1	68.0	1.58	73.0	0.18
	3	68.0	1.31	73.0	0.17

IPW weight calculation for the CHD data from the DCH study. The two-phase stratified cohort sampling was applied to select participants into the Phase II sample for data analysis.

Table 4

Status	Gender	Full Cohort	Phase II Sample	Weight	Complete Data
Case	Male	1,526	1,526	1.0	1,417
	Female	578	578	1.0	526
Control	Male	24,658	917	26.9	890
	Female	28,785	842	34.2	800
		55,547	3,863		3,633

Estimated regression coefficients for the stratified Cox and stratified NPT models. For comparison, the coefficients from the stratified Cox model are also normalized to have unit L_1 norm. A positive coefficient corresponds to a beneficial effect and prolonged survival time.

Table 5

	Cox			NPT		
	Coef.	S.E.	P-value	Coef.	S.E.	P-value
HDL-C ₊ (log ₂ mg/dL)	0.005	0.021	8.12e-01	-0.002	0.017	8.87e-01
HDL-C ₋ (log ₂ mg/dL)	0.124	0.030	3.26e-05	0.144	0.030	1.20e-06
Age at baseline (yrs)	-0.035	0.004	2.27e-16	-0.033	0.004	4.98e-15
Sex (female vs. male)	0.290	0.032	1.30e-19	0.311	0.029	3.29e-27
Education (<8 vs. 8 yrs)	-0.082	0.031	8.36e-03	-0.053	0.028	5.32e-02
Current smoker (yes vs. no)	-0.235	0.028	2.83e-17	-0.240	0.025	2.52e-21
Alcohol (g/d)	0.001	0.001	4.44e-01	0.001	0.001	4.02e-01
Hypertension (yes vs. no)	-0.215	0.031	7.06e-12	-0.201	0.030	1.86e-11
BMI (kg/m ²)	-0.014	0.004	1.21e-03	-0.015	0.004	1.24e-04