Published in final edited form as:

Biometrics. 2015 December ; 71(4): 1139–1149. doi:10.1111/biom.12344.

Assessing Incremental Value of Biomarkers with Multi-phase Nested case-control Studies

Qian M. Zhou,

Department of Statistics and Actuarial Science, Simon Fraser University, Burnaby, BC, Canada, V5A1S6 qmzhou@stat.sfu.ca

Yingye Zheng,

Public Health Sciences Division, Fred Hutchinson Cancer Research Center, Seattle, WA, USA, 98109 yzheng@fhcrc.org

Lori B. Chibnik,

Brigham and Women's Hospital and Harvard Medical School, Boston, MA, USA, 02115 lchibnik@rics.bwh.harvard.edu

Elizabeth W. Karlson, and

Brigham and Women's Hospital and Harvard Medical School, Boston, MA, USA, 02115 EKARLSON@PARTNERS.ORG

Tianxi Cai

Department of Biostatistics, Harvard University, Boston, MA, USA, 02115 tcai@hsph.harvard.edu

Summary

Accurate risk prediction models are needed to identify different risk groups for individualized prevention and treatment strategies. In the Nurses' Health Study, to examine the effects of several biomarkers and genetic markers on the risk of rheumatoid arthritis (RA), a three-phase nested case-control (NCC) design was conducted, in which two sequential NCC subcohorts were formed with one nested within the other, and one set of new markers measured on each of the subcohorts. One objective of the study is to evaluate clinical values of novel biomarkers in improving upon existing risk models because of potential cost associated with assaying biomarkers. In this paper, we develop robust statistical procedures for constructing risk prediction models for RA and estimating the incremental value (IncV) of new markers based on three-phase NCC studies. Our method also takes into account possible time-varying effects of biomarkers in risk modeling, which allows us to more robustly assess the biomarker utility and address the question of whether a marker is better suited for short-term or long-term risk prediction. The proposed procedures are shown to perform well in finite samples via simulation studies.

Keywords

Incremental value; Inverse probability weighting; Nested case-control study; Rheumatoid arthritis; Risk prediction; Time dependent ROC curve analysis

1. Introduction

Accurate risk prediction models are key to forming individualized strategies of disease prevention and treatment management. For rheumatoid arthritis (RA), using risk prediction models, individuals classified as being at high risk may be recommended for intensive preventive or therapeutic treatment in order to improve long-term disease outcomes. In recent years, several biological markers of inflammatory and immune activity, such as soluble tumor necrosis factor II (sTNFR_{II}) and interleukin-6 (IL6), as well as a number of genetic variants, including the HLA-DRB1 shared epitope and PTPN22, have emerged as useful predictors of RA risk (Raychaudhuri et al. 2008; Karlson et al. 2009; Stahl et al. 2010). They have shown great potential in improving existing prediction models, which are built on traditional clinical variables, such as gender, smoking and alcohol intake (Karlson et al. 2010). Two major objectives of the research are (i) to construct 5-year and 15-year RA risk prediction models combining information from clinical, biological and genetic markers; and (ii) to rigorously examine whether including these new markers can lead to sufficiently accurate prediction models for risk stratification, and to quantify the incremental value (IncV) of these new markers in risk prediction. In the following, we first describe a special study design adopted by Nurses' Health Study (NHS) for such evaluation. Then we discuss the practical issues and challenges in analyzing the data.

1.1 Three-phase NCC RA study in NHS

The NHS is a prospective cohort of 121,700 US female registered nurses aged 30 to 55 years in 1976. All the participants in this study completed an initial questionnaire regarding diseases, lifestyle, and health practices, and they were followed by biennial questionnaires to update information. Between 1989 and 1990, 32,826 participants contributed blood for future research. This blood cohort has become a valuable resource to study a wide range of diseases.

To investigate the value of novel biomarkers and genetic markers for RA risk prediction, the most reliable data resource is ideally from a large prospective cohort such as the NHS blood cohort. Due to high measurement cost and limited availability of biospecimens, it is often not feasible to obtain biomarker values for the full cohort. A cost-effective strategy to overcome this is to employ a nested case-control (NCC) study design (Thomas 1977; Prentice and Breslow 1978). Under this design, routine clinical factors are measured on the full cohort, while expensive biomarkers are only measured on the NCC subcohort, which typically includes all cases and a fraction of controls randomly selected from the risk sets of the cases and often matched to the cases on some variables (Gram et al. 2000; Sudbø et al. 2005; Hippisley-Cox and Coupland 2005). Often, as knowledge accumulates and scientific hypotheses evolve over time, the stored specimens might be assayed at multiple stages for different sets of markers.

A three-phase NCC RA study based on the NHS blood cohort was initiated sequentially in 2004 to address two evolving questions: (a) the importance of inflammation markers in predicting RA risks; and (b) the clinical utility of genetic markers in improving RA risk prediction. The full cohort of this RA study includes the participants in the NHS blood cohort who had not developed RA at blood draw. In Phase I, clinical markers, including age, pack-years of smoking and cumulative alcohol intake, were collected at blood draw on the full cohort. In Phase II, for each RA case, 3 controls were randomly selected from the risk set of the case matched on variables including age (± 1 year), menopausal status and postmenopausal hormone use. The Phase II subcohort, denoted by *NCC*, consists of 126 RA cases and 377 selected controls, and their blood samples were assayed for several inflammation biomarkers including IL-6 and sTNFR_{II} to address question (a). In Phase III, 1 control was further randomly selected from the three controls chosen in *NCC*. Samples in this Phase III subcohort, denoted by *NCC*, were genotyped for 39 SNPs identified as associated with RA risks in previous studies to address question (b). After removing those with poor quality genotype data or missing HLA-SE genotype, 122 cases and 123 controls with complete information on these SNPs composed the *NCC*. Figure 1 illustrates the sequential sampling scheme.

In addition, for subjects who developed cancer (except nonmelanoma skin cancer) prior to 2004, their blood samples were excluded from either of these two subcohorts since they were preserved for other studies. By 2004, 143 subjects in the full cohort were confirmed to have definite incident RA. A total of 126 out of the 143 RA cases have blood samples available for the proposed RA study.

1.2 Issues and Challenges in Analyzing the NHS RA data

Recall that the NHS RA study consists of two analytical goals: (i) to construct multivariate 5-year and 15-year RA risk models, and (ii) to evaluate the accuracy of these risk models and the IncVs of novel markers in risk prediction. Both require some deliberation given the specific design the study adopted.

First the special three-phase NCC design used in the RA study poses substantial difficulties in constructing prediction models. Individuals selected for the marker measurement at various stages can not be regarded as a simple random sample from the original cohort, and accounting for the sampling scheme is necessary for building models while incorporating data collected at various stages. Parameters in a risk model can be estimated by fitting a conditional logistic regression (CLR) model (Goldstein and Langholz 1992) or by maximizing an inverse probability weighted log-partial likelihood function based on the Cox's proportional hazards (PH) models (Samuelsen 1997). The former, conditional on sampled matched sets, can lead to inefficient relative risk estimates due to small numbers of matched sets. For the latter, the sampling probability needs to be modified for the NHS three-phase RA study. An additional complication arises from sample protection. Not all subjects initially sampled into the RA subcohorts have blood samples available. A modified weighting strategy is needed to account for the additional missingness due to sample protection, which further complicates inference, particularly in variance calculations.

The potential time-varying effects of the biomarkers on RA risks pose additional challenges. As indicated in Karlson et al. (2009), sTNFR^{II} appears to be elevated for RA patients years before RA diagnosis, while IL6 appears to be elevated only when time gets closer to RA onset. In fact, risk factors exhibiting time-varying effects occur quite frequently in other studies (Joslyn et al. 2005; Anderson et al. 2008). Existing literature based on the PH or CLR models assumes time invariant marker effects, which is not valid for the RA study. To allow the marker effects to vary over time, Liu et al. (2010) proposed fitting a PH model with time-varying effects for two-phase NCC studies and employed kernel smoothing for estimation. However, in the NHS RA data, the number of the markers of interest is not small and the subcohort sizes are relatively small; the kernel-based estimation procedure may perform poorly due to the curse of dimensionality and limited sample sizes. In this paper, we consider a simple yet robust risk modeling approach that accommodates both three-phase sampling and time-varying effects. We demonstrate that incorporating time-varying marker effects could lead to more accurate risk models.

Finally, evaluating the prediction performance of the estimated multivariate RA risk models with multi-phase data requires new methodology development. No statistical procedures currently exist to make proper inference about commonly used prediction accuracy measures such as the time-dependent receiver operating characteristic (ROC) curve (Gerds et al. 2008) with such data. Time-dependent ROC analysis methods for two-phase NCC studies have been proposed based on inverse probability weighting (IPW) principals (Cai and Zheng 2011, 2012). However, these methods only consider a single marker and are not readily applicable to multivariate settings. Many published NCC data analyses build risk models using CLR methods and then quantify the model prediction performance via simple ROC analyses for binary outcomes by naively treating the data as simple case-control studies (e.g. Gram et al. 2000; Boekholdt et al. 2006; Wolfe and Michaud 2008). These naive ROC analyses not only produced biased estimates of the ROC curves due to ignoring outcome-dependent sampling and censoring, but they also failed to capture potential time-varying predictive performance of the risk models. To evaluate the IncV of the biomarkers or genetic markers, one may compare the prediction performance of two risk models built with and without the new markers. No inference procedures currently exist to make proper inference about such IncV parameters even for two-phase studies. Inference for three-phase NCC designs is even more challenging than two-phase studies since the sequential finite population sampling induces more complex correlation among observations.

In this paper, we develop statistical tools for risk modeling that can address these practical and statistical challenges with three-phase studies. We construct RA risk models with multiple markers collected at different phases of the study while accounting for possible time-varying effects of the markers. We provide nonparametric estimates for various time-dependent accuracy measures to summarize the accuracy of the proposed RA risk models and the IncV of the novel markers with respect to these accuracy measures. We also provide standard error estimates of these estimated quantities. The rest of the paper is organized as follows. Section 2 includes constructing risk models with different sets of markers of interest, and Section 3 demonstrates evaluation of the IncV of the biomarkers and genetic markers. To examine the validity of our method and compare our method with existing

approaches, Section 4 presents a simulation study that mimics the design of the RA study. Concluding remarks are given in Section 5.

2. Constructing Robust Prediction Models for RA

Let T^* denote the time from blood draw to disease onset such as RA diagnosis. Due to censoring, we only observe $T = \min(T^*, C)$ and $\delta = \mathbb{I}(T^* < C)$, where C is the censoring time. Let $\mathbf{Z}$ be the clinical markers collected on the full cohort: age, pack-years of smoking and cumulative alcohol intake. Here, a square root transformation was applied to pack-years of smoking and cumulative alcohol intake as previously suggested in Zhou et al. (2001). Let $\mathbf{B}$ be the two biomarkers, IL-6 and sTNFR2, measured on the Phase II subcohort NCC . Let $\mathbf{G}$ be the genetic risk score (GRS) constructed as a weighted average of the 39 RA SNPs assayed on the Phase III subcohort, NCC , where the weights are the log-odds ratios reported in Stahl et al. (2010).

Let $\mathcal{D} = \{\mathbf{D}_i = (T_i, \delta_i, \mathbf{Z}_{G+B_i}), i=1, \dots, n\}$ denote the underlying complete data of the full cohort, where $\mathbf{Z}_{G+B} = (\mathbf{Z}^\top, \mathbf{B}^\top, \mathbf{G}^\top)^\top$. Recall that the two subcohorts consist of subjects who were selected via the NCC sampling and had biospecimens not protected for competing studies. Let S_i and S_i^* indicate whether the i^{th} subject is initially selected via NCC and NCC sampling, respectively. Let V_i indicate whether the sample of the i^{th} subject is not protected for competing studies. Then $V_i = V_i S_i$ and $V_i^* = V_i S_i^*$ indicate whether the i^{th} subject is included in the final NCC and NCC subcohorts, respectively.

Let $\hat{p}_i = P(S_i = 1 | \mathcal{D})$, $\hat{p}_i^* = P(S_i^* = 1 | \mathcal{D})$, $\hat{p}_i = P(V_i = 1 | \mathcal{D})$ and $\hat{p}_i^* = P(V_i^* = 1 | \mathcal{D})$. It is straightforward to calculate $\hat{p}_i$ following similar arguments as given in Samuelsen (1997). In calculating $\hat{p}_i^*$, we need to take into account the probability of being selected to NCC as a control given that the subject has been selected to NCC as a control. Formulae for $\hat{p}_i$ and $\hat{p}_i^*$ are given in the Web Appendix A. Let $\mathbb{p}_0$ denote the fraction of samples not being protected for competing studies, and it is estimated to be around 90% from the data. Since sample protection is due to competing studies irrelevant to RA, the missingness of biomarkers due to sample protection is assumed to be completely at random. Thus, $\hat{p}_i = \hat{p}_i \mathbb{p}_0$ and $\hat{p}_i^* = \hat{p}_i^* \mathbb{p}_0$. Let $\hat{w}_i = V_i / \hat{p}_i$ and $\hat{w}_i^* = V_i^* / \hat{p}_i^*$ be the IPW weights accounting for the missingness of $\mathbf{B}$ only and of $\mathbf{G}$, respectively.

2.1 Model Specification and Methods

First, we develop an RA absolute risk prediction model with $\mathbf{Z}_{G+B}$. A commonly used approach is to employ a PH model:

$$P(T^* \leq t_0 | \mathbf{Z}_{G+B}) = 1 - \exp \left[-\exp \left\{ \boldsymbol{\gamma}_{+}^{COX}(t_0)^\top \vec{\mathbf{Z}}_{G+B} \right\} \right], \text{ where}$$

$$\boldsymbol{\gamma}_{+}^{COX}(t_0) = \left(\log \Lambda_{G+B}(t_0), \boldsymbol{\beta}_{+}^\top \right)^\top, \boldsymbol{\beta}_{+}^\top \text{ and } \Lambda_{G+B}(\cdot) \text{ are the unknown log HR parameters and baseline cumulative hazard function, respectively. Here and in the sequel, for any vector } \mathbf{a}, \vec{\mathbf{a}} = (1, \mathbf{a}^\top)^\top. \text{ Following Samuelsen (1997), one may estimate } \boldsymbol{\beta}_{G+B} \text{ as } \hat{\boldsymbol{\beta}}_{G+B}, \text{ the maximizer of the IPW log partial likelihood function, and the IPW Breslow's estimator } \hat{\Lambda}_{G+B}(t_0) \text{ can be}$$

used to estimate $\Lambda_{G+B}(t_0)$ as in Cai and Zheng (2012). However, the PH model does not allow marker effects to change over time. To account for possible time-varying effects as discussed in Section 1-2, we consider a time-specific generalized linear model (GLM):

$$P(T^{\dagger} \leq t_0 | \mathbf{Z}_{G+B}) = g \left\{ \boldsymbol{\gamma}_{+}^{GLM}(t_0)^{\top} \vec{\mathbf{Z}}_{G+B} \right\}, \quad (2.1)$$

where $\boldsymbol{\gamma}_{+}^{GLM}(t_0) = \left\{ \alpha(t_0), \boldsymbol{\beta}_{G+B}(t_0)^{\top} \right\}^{\top}$. In contrast to the PH model, the time-specific GLM we adopted here allows for the effects of markers to vary with time via the specification of $\boldsymbol{\beta}_{G+B}(t_0)$.

To estimate $\alpha(t_0)$ and $\boldsymbol{\beta}_{G+B}(t_0)$ with NCC data, we modify the estimator proposed in Uno et al. (2007) and employ a double IPW (DIPW) estimating equation using weights $\hat{w}_i$ and $\hat{\omega}_{t_0i}^C$,

$$\sum_{i=1}^n \hat{w}_i \hat{\omega}_{t_0i}^C \vec{\mathbf{Z}}_{G+B_i} \left[I(T_i \leq t_0) - g \left\{ \boldsymbol{\gamma}_{+}^{GLM}(t_0)^{\top} \vec{\mathbf{Z}}_{G+B_i} \right\} \right] = 0 \quad (2.2)$$

where $\hat{w}_i$ accounts for missingness of $\mathbf{Z}_{G+B_i}$, and

$$\hat{\omega}_{t_0i}^C = \delta_i I(T_i \leq t_0) / \hat{\mathcal{G}}_{\mathbf{Z}_{C_i}}^C(T_i) + I(T_i > t_0) / \hat{\mathcal{G}}_{\mathbf{Z}_{C_i}}^C(t_0)$$

accounts for censoring, $\hat{\mathcal{G}}_{\mathbf{Z}_C}^C(t)$ is a consistent estimator of $\mathcal{G}_{\mathbf{Z}_C}^C(t) = P(C \geq t | \mathbf{Z}_C)$, and $\mathbf{Z}_C$ is a sub-vector of $\mathbf{Z}_{G+B}$ that censoring may depend on. If the censoring is independent of both the event time and markers, $\hat{\mathcal{G}}_{\mathbf{Z}_C}^C(t)$ could be obtained by the Kaplan-Meier (KM) estimator. If the censoring depends on some of the markers, say $\mathbf{Z}$, due to the curse of dimensionality (Robins and Ritov 1997), additional model assumptions are required to make valid inference. For example, a PH model could be fit to estimate $\mathcal{G}_{\mathbf{Z}}^C(t)$.

To predict the t_0 -year absolute risk, one may replace $\boldsymbol{\gamma}_{+}^{GLM}$ in (2.1) with its corresponding estimator, $\hat{\boldsymbol{\gamma}}_{+}^{GLM}$. Here we use $\hat{\mathbf{F}}_{G+B}$ to denote the t_0 -year risk estimate obtained by fitting the time-specific GLM with $\mathbf{Z}_{G+B}$ (model (i)). To evaluate the IncV of the biomarkers and genetic risk score in Section 3, we also fit time-specific GLMs with (ii) $\mathbf{Z}_B = (\mathbf{Z}^{\top}, \mathbf{B}^{\top})^{\top}$, (iii) $\mathbf{Z}_G = (\mathbf{Z}^{\top}, \mathbf{G}^{\top})^{\top}$, and (iv) $\mathbf{Z}$ only separately, and obtain the corresponding estimated risk scores, denoted by $\hat{\mathbf{F}}_B, \hat{\mathbf{F}}_G, \hat{\mathbf{F}}_Z$, respectively. Note that $\hat{w}_i$ is replaced with $\hat{w}_i$ for model (ii), and replaced with 1 for model (iv) since $\mathbf{Z}$ is observed for all subjects.

2.2 Risk Modeling with NHS RA Data

To develop RA risk models, we fit both the time-specific GLM and PH models to the NHS RA data, using the IPW framework described in Section 2.1. For time-specific GLMs, we set different prediction time $t_0 \in \mathcal{T}_0 = \{5, 6, 8, 10, 12, 14, 15\}$, allowing us to examine how the effects of the markers changes over t_0 . We let $g(x) = 1 - \exp\{-\exp(x)\}$ in the GLM such

that (2.1) reduces to the PH model when the marker effects are constant over time. In addition, it is not likely that the censoring time directly depends on the markers measured at the baseline. Thus, to calculate the weights $\hat{\omega}_{t_0 i}^C$ in (2.2), the survival function of the censoring time $\mathcal{G}_{Z_c}^C(t)$ is estimated by the KM estimator under the assumption of independence censoring. We also performed a sensitivity analysis by allowing C to depend on $\mathbf{Z}$ through a Cox's PH model and obtained similar accuracy estimates, as shown in Figure 4 of the Web Appendix E. The IncV and regression parameters estimates are also similar (results not shown).

In Figure 2, we present the point and interval estimates of the regression parameters for the full risk models with $\mathbf{Z}_{G+B}$. In addition, we constructed risk models with (ii) $\mathbf{Z}_B$, (iii) $\mathbf{Z}_G$, and (iv) $\mathbf{Z}$ only, and these three models are called biomarker model, genetic model and clinical model, respectively. The regression parameter estimates for these four risk models are shown in Table 1 of Web Appendix E for $t_0 = 5$ and 15. The results show that pack-years of smoking is more strongly associated with increased longer term RA risks although the association was not statistically significant except for the clinical model. Consistent with previous findings (Källberg et al. 2009; Liao et al. 2009), alcohol consumption appears to be inversely associated with RA risk. The estimated effect for alcohol under the time-specific GLM with $\mathbf{Z}_{G+B}$ is -0.24 (95% CI: $[-0.49, 0.01]$, p-value 0.061) for the 5-year model and -0.16 (95% CI: $[-0.31, -0.01]$, p-value 0.043) for the 15-year model. Age is not a significant risk factor under most of the models. In general, higher levels of sTNFR2 and IL6 are associated with higher risk of RA, but the effect of sTNFR2 appears to be strong for both short term and long term risks while the effect of IL6 appears to be much stronger for short-term risks than long-term risks. For example, the estimated effects under the 5-year and 15-year GLM with $\mathbf{Z}_{G+B}$ were 0.82 and 0.89 for sTNFR2; and 0.89 and 0.27 for IL6, respectively. The GRS appears to be highly predictive of RA risks across all $t_0 \in \mathcal{T}_0$ with effect estimated as 0.80 at 5-year and 0.56 at 15-year under the time-specific GLM with $\mathbf{Z}_{G+B}$. The time-varying patterns of the biomarker effects are consistent with results reported in the literature and indicate that the PH model may not be a good choice in this application.

3. Evaluation of the Risk Models and IncV of Novel Markers

To evaluate and compare the prediction performance of the four RA risk models constructed in Section 2, one may quantify the predictiveness of the corresponding risk scores based on commonly used accuracy measures, including the time-dependent ROC curve, the positive predictive values (PPV) and the negative predictive values (NPV) (See Heagerty et al. 2000; Pepe 2004). Let $\mathbb{F}$ denote the generic t_0 -year risk estimate obtained from fitting one of the models under consideration. The false positive rate (FPR) and the true positive rate (TPR) functions for predicting t_0 -year RA risk associated with $\mathbb{F}$ are defined as

$$FPR_{t_0}(c) = P(\mathbb{F} > c \mid T^\dagger > t_0) \text{ and } TPR_{t_0}(c) = P(\mathbb{F} > c \mid T^\dagger \leq t_0).$$

The corresponding ROC curve is then $ROC_{t_0}(u) = TPR_{t_0}\{FPR_{t_0}^{-1}(u)\}$ with area under the ROC curve (AUC)

being $AUC_{t_0} = \int_0^1 ROC_{t_0}(u) du$. To evaluate the IncV of new markers, say $\mathbf{B}$, on top of existing markers, say $\mathbf{Z}$, we compare the accuracy measures of two risk models, m_1 with $\mathbf{B} +$

$\mathbf{Z}$ and m_2 with $\mathbf{Z}$ only. The IncV of $\mathbf{B}$ with respect to an arbitrary accuracy measure $\mathbb{D}$ can be quantified by $\mathbb{I}^{m_1/m_2} = \mathbb{D}^{\mathbb{F}_{m_1}} - \mathbb{D}^{\mathbb{F}_{m_2}}$, where $\mathbb{F}_m$ denotes the risk score for model m and $\mathbb{D}^{\mathbb{F}_m}$ denotes the accuracy of the risk score $\mathbb{F}_m$ for $m \in \{m_1, m_2\}$. For example, the IncV of $\mathbf{B}$ and $\mathbf{G}$ on top of $\mathbf{Z}$ in AUC is defined by $\mathbb{I}_{AUC}^{+ / Z} = AUC^{\mathbb{F}_{G+B}} - AUC^{\mathbb{F}_Z}$.

3.1 Estimation and Inference Procedure for Prediction Accuracy and IncV

To estimate the accuracy and IncV parameters, we propose DIPW estimators by extending those given in Uno et al. (2007) and Cai and Zheng (2011) to account for the additional complexity due to three-phase sampling and multiple markers. For example, the IncV $\mathbb{I}_{AUC}^{m_1/m_2}$ can be estimated by $\hat{\mathbb{I}}_{AUC}^{m_1/m_2} = \widehat{AUC}^{\hat{\mathbb{F}}_{m_1}}(c) - \widehat{AUC}^{\hat{\mathbb{F}}_{m_2}}(c)$, where

$$\widehat{AUC}^{\hat{\mathbb{F}}_m}(c) = \frac{\sum_{i=1}^n \sum_{j=1}^n \hat{w}_i^m \hat{\omega}_{t_0 i}^C \hat{w}_j^m \hat{\omega}_{t_0 j}^C I \left\{ \hat{\mathbb{F}}_{mi} > \hat{\mathbb{F}}_{mj}, T_i \leq t_0, T_j > t_0 \right\}}{\sum_{i=1}^n \sum_{j=1}^n \hat{w}_i^m \hat{\omega}_{t_0 i}^C \hat{w}_j^m \hat{\omega}_{t_0 j}^C I(T_i \leq t_0, T_j > t_0)}$$

for $m \in \{m_1, m_2\}$, and $\hat{w}_i^m = \hat{w}_i$ for models include $\mathbf{G}$; $\hat{w}_i^m = \hat{w}_i$ for models involve $\mathbf{B}$ but not $\mathbf{G}$; $\hat{w}_i^m = 1$ for models that only involve $\mathbf{Z}$. However, the above estimates may be too optimistic due to over-fitting. To correct for the over-fitting, one may employ standard procedures such as cross-validation or 0.632 Bootstrap (Efron 1983; Jiang and Simon 2007).

To construct confidence intervals (CIs), we provide, in Web Appendix B, variance formulae for our accuracy and IncV estimators. These variance calculations account for additional variations from various sources including the estimated model parameters and the estimated censoring probabilities. Since the PH model may perform well for some applications, we also provide the formulae for the asymptotic variances (aVARs) of the estimators for the relative risk parameters, accuracy measures and IncV parameters obtained from fitting PH models in Web Appendix C. These variance forms are useful in practice as the standard resampling procedures such as the bootstrap fail to approximate the variance of IPW estimators constructed with NCC data (Gray 2009; Cai and Zheng 2013). The finite population sampling of the NCC studies induces complex between-subject correlations for the sampling variables, which are not ignorable at the first order. The aVAR of a generic DIPW estimator typically consists of several components that account for the variability

from sampling variables $\{S_i, S_i\}$, V_i and $\hat{\mathcal{G}}^C(t)$ as well as the correlation between $\{S_i, S_i\}$ and $\{S_j, S_j\}$ when $i \neq j$ due to finite population sampling. For example,

consider a DIPW statistic constructed with NCC data, $\hat{\mathcal{W}}_{\mathbf{R}_{t_0}}^C = n^{-\frac{1}{2}} \sum_{i=1}^n \hat{w}_i \hat{\omega}_{t_0 i}^C \mathbf{R}_{t_0}(\mathbf{D}_i)$,

where $E\{\mathbf{R}_{t_0}(\mathbf{D}_i)\} = 0$. Its aVAR, $\Sigma_{\mathbf{R}_{t_0}}^C$, can be decomposed into two components as shown in (B.2) of Web Appendix B.1. The first component accounts for the variability from S_i , V_i and $\hat{\mathcal{G}}^C(t)$, ignoring the negative correlation between S_i and S_j . The second term in (B.2) corrects for this correlation, ignoring which always results in over-estimation of the variance. In addition, the variance calculation of the IncV estimators accounts for the

correlation among the sampling indicators from multiple phases, which has not been previously studied in the literature.

3.2 Evaluating RA Risk Models and IncV of Novel Markers with NHS RA Data

To evaluate RA risk models constructed in Section 2.2 and assess the IncV of adding the biomarkers and/or genetic score, we consider the following accuracy measures: AUC, $TPR_{t_0} \{FPR_{t_0}^{-1}(0.1)\}$, $PPV_{t_0} \{FPR_{t_0}^{-1}(0.1)\}$ and $NPV_{t_0} \{FPR_{t_0}^{-1}(0.1)\}$. The point estimates for the accuracy measures were obtained via an average over 200 repeated 0.632 bootstrap estimates. The point and interval estimates of the accuracy parameters are shown in Figure 1 of Web Appendix E for both the time-specific GLM and PH models with $\mathbf{Z}_{G+B}$ across all $t_0 \in \mathcal{T}_0$ and in Table 2 of Web Appendix E for all the 5-year and 15-year models.

First, the clinical models have very low discriminatory capacity. The AUCs of the time-specific GLMs with $\mathbf{Z}$ are 0.53 (95% CI: [0.42, 0.64]) for predicting 5-year RA risk and 0.58 (95% CI: [0.53, 0.63]) for 15-year risk. The AUCs of the time-specific GLM with $\mathbf{Z}_{G+B}$ increase to 0.70 (95% CI: [0.62, 0.79]) for 5-year risk and 0.62 (95% CI: [0.55, 0.69]) for 15-year risk.

Second, when compared to the PH modeling, the accuracy measures of the time-specific GLM models tend to be higher for predicting short-term risks but similar for longer-term risks. For example, for predicting 5-year RA risk with $\mathbf{Z}_B$, the $TPR_{t_0} \{FPR_{t_0}^{-1}(0.1)\}$ under the time-specific GLM is 0.256, which is almost twice the TPR under the PH model, which is 0.135; and the $PPV_{t_0} \{FPR_{t_0}^{-1}(0.1)\}$ under the time-specific GLM is 0.0032 in contrast to the PPV, 0.0017, under the PH model. However, for predicting the 15-year risk, the accuracy measures of the PH and time-specific GLM based biomarker models are much closer. These results along with the coefficients given in Table 1 of Web Appendix E indicate that the fitted PH model in this study is mainly summarizing the long term effects of the markers, while the time-specific GLM enables us to capture both long-term and short-term effects.

The estimates of the IncV by adding $\mathbf{B}$ and $\mathbf{G}$ are shown in Figure 3 for comparing the models with $\mathbf{Z}_{G+B}$ to those with $\mathbf{Z}$ alone across all $t_0 \in \mathcal{T}_0$. The results for all the comparisons between different models at $t_0 = 5$ and 15 are shown in Table 3 of Web Appendix E. First, comparing with the PH models, the time-specific GLM appear to capture the IncV of biomarkers for the short term risks more effectively due to its ability to model their time-varying effects on the RA risk. For example, for predicting the 5-year RA risk, the gain in AUC from adding $\mathbf{B}$ on top of $\mathbf{Z}$ is 0.12 (95% CI: [0.01, 0.23], p-value: 0.03) by using the time-specific GLM for risk modeling. On the other hand, the gain in AUC is only 0.06 (95% CI: [-0.06, 0.18], p-value: 0.31) with the PH modeling. For predicting long-term risks, the time-specific GLM and PH models produce similar IncV estimates. Secondly, the IncV of novel markers also appear to vary over t_0 . For example, the biomarkers appear to be more helpful for improving the AUC for shorter term risks compared to for longer term risks. Using the time-specific GLM, the improvement in AUC by adding $\mathbf{B}$ on top of $\mathbf{Z}$ is

0.12 (95% CI: [0.01, 0.23], p-value: 0.03) for 5-year RA risk and -0.01 (95% CI: $[-0.07, 0.05]$, p-value: 0.75) for 15-year risk, respectively.

We also performed the naive ROC analysis, as discussed in Section 1.2, in which the HR was estimated using the CLR method and accuracy measures were estimated by treating the data as standard case-control studies. This naive approach produced estimates with substantial biases. For example, for the full model with all the markers $\mathbf{Z}_{G+B}$, the estimates for AUC, $\text{TPR}\{\text{FPR}^{-1}(0.1)\}$, $\text{PPV}\{\text{FPR}^{-1}(0.1)\}$ and $\text{NPV}\{\text{FPR}^{-1}(0.1)\}$ are 0.534, 0.125, 0.505 and 0.577. In contrast to an AUC of 0.70 from our estimate for predicting 5-year risk, this naive AUC estimate would lead to misleading conclusions.

Although the inclusion of the genetic and biomarker information appears to improve risk prediction substantially over the clinical model with respect to AUC, their utility in identifying high-risk subpopulations may be better reflected by assessing the PPV at a given threshold value for defining a high-risk group. To this end, we choose a cut-off value for classifying subjects as high-risk by achieving an FPR of 0.1. Under the time-specific GLM using all the markers $\mathbf{Z}_{G+B}$ with $\text{FPR} = 0.1$, the probability of developing RA within 5 years is 0.35% for the high-risk group and 0.11% for the low-risk group. Although there is a more than three-fold increase in absolute risk for the high-risk group compared to the low-risk group, the absolute risk for the high-risk group remains low due to the low overall incidence rate. A lower FPR level might be helpful to refine the definition of high risk group. However, the limited sample size in our study does not allow a stable estimation of PPV for such tail probabilities. This highlights the potential utility of including additional risk factors such as family history and focusing on the identification of high risk groups within certain subpopulations.

4. Simulation

In this section, we report results from simulation studies aiming to (i) examine the validity of the proposed inference procedures, and (ii) compare the time-specific GLMs with the PH models with regard to prediction accuracy in the presence of time-varying marker effects.

The simulation studies mimic the NHS RA study design. We consider three markers $\mathbf{Z}_{G+B} = (Z, B, G)^T$: a conventional marker $Z \sim \mathcal{N}(0, 1)$, a biological marker $B = Z + e_B$ with $e_B \sim \mathcal{N}(0, 1)$, and a genetic score $G \sim \text{Uniform}[-4, 0]$. The event time $T^\dagger$ given $\mathbf{Z}_{G+B}$ is

generated from $\log(T^\dagger - 2) = 2.5 + \frac{1}{1.5}(\epsilon_T - 0.5Z - 0.5B) / (2 - G)$, where ϵ_T is generated from a standard extreme value distribution. This leads to

$$P(T^\dagger \leq t | Z, B, G) = 1 - \exp[-\exp\{\alpha(t) + 0.5Z + 0.5B + \beta_G(t)G\}], \quad (4.3)$$

where $\alpha(t) = 3\{\log(t-2) - 2.5\}$ and $\beta_G(t) = -1.5\{\log(t-2) - 2.5\}$. This implies different magnitudes of the effects of the genetic score on the risk for different time periods. We assume that Z is available for the Phase I full cohort, and B is available for only the Phase II NCC subcohort, and G is available only for the Phase III NCC subcohort. To construct these two NCC subcohorts, in the first stage, we select three controls, for each case, from

either (i) the risk set without matching, or (ii) the risk set with matching on variables $\mathbf{M} = (M_1, M_2)$. The matching variables $\mathbf{M}$ are generated as follows: (i) $M_1 = I(\tilde{M}_1 > 0.5)$ with $\tilde{M}_1 = \Phi(Z + e_1)$ and $e_1 \sim N(0, 1)$; (ii) M_2 is the closest integer to $\tilde{M}_2 = 5\Phi(B + e_2)$, where $e_2 \sim N(0, 1)$. Exact matching and matching up to ± 1 are required for M_1 and M_2 , respectively. In the second stage, for each case, one control is randomly selected from the three controls chosen from the first stage. The sample size of the full cohort is 10,000 and all the results are based on 1000 independent replicates.

Here, we consider independent censoring with $C = \min\{C_1 + 9.5, C_2\}$, where $C_1 \sim \text{Gamma}(2, 2)$ and $C_2 \sim \text{Uniform}[5, 15]$. This leads to about 90% censoring rate. To predict $T^\dagger \leq t_0$ for a given t_0 , we fit two full prediction models with all the markers $\mathbf{Z}_{G+B}$: (i) $\mathbb{F}_+^{GLM}$: a time-specific GLM $P(T^\dagger \leq t_0) = g\{\alpha(t_0) + \beta_Z(t_0)Z + \beta_B(t_0)B + \beta_G(t_0)G\}$, where $g(x) = 1 - \exp\{-\exp(x)\}$; (ii) $\mathbb{F}_+^{COX}$: a PH model $\lambda_{\mathbf{Z}_{G+B}}(t) = \lambda_0(t) \exp(\beta_Z Z + \beta_B B + \beta_G G)$. To investigate possibly different time-varying marker effects, we consider prediction time intervals $t_0 = 5, 6, 7, 8, 9, 10$. Figure 4 shows the average point estimates of β_Z , β_B and β_G obtained from the time-specific GLM models and the Cox PH model as well as their empirical 95% confidence intervals. The plots clearly demonstrate that, compared with the PH models, the time-specific GLM models are able to more accurately characterize the time-varying effects of marker G . The PH model is more likely to capture the effects of the markers on the long-term risk. In addition, we also estimate the log HRs by a CLR model, and the estimates are very close to those obtained from the IPW PH model.

Tables 4 and 5 in Web Appendix E report the performance of the DIPW estimators of AUC, TPR, NPV and PPV at the cutoff value such that FPR = 0.1 of the full prediction models $\mathbb{F}_+^{GLM}$ and $\mathbb{F}_+^{COX}$ for the risk prediction with both the short-term $t_0 = 6$ and the long-term $t_0 = 10$. To assess the IncV of G on top of $\mathbf{Z}_B = (Z, B)^\top$, we also fit a time-specific GLM $\mathbb{F}_+^{GLM}$ and a Cox's PH model $\mathbb{F}_+^{COX}$ with the markers $\mathbf{Z}_B$. Tables 6 and 7 in Web Appendix E report the DIPW estimators of the IncV of adding G on top of (Z, B) for both the short-term and long-term risk prediction. Across all the settings we considered, the results suggest that our procedures provide valid point and interval estimators for the accuracy or IncV measures. Specifically, the adjusted variance estimators provide good approximations to the true empirical variances. On the other hand, the naive standard error estimators without accounting for correlation among sampling indicators could substantially over-estimate the variance, especially in the presence of additional matching. For example, as shown in Table 5 in Web Appendix E, the over-estimation of the variance could be as high as 40% for NPV using the naive estimator. The coverage probabilities using the proposed variance estimators are close to the nominal level. Note that only the model $\mathbb{F}_+^{GLM}$ correctly specifies the underlying true model and the rest of the fitted models are misspecified. Our results also suggest that the proposed inference procedures based on the DIPW estimators are valid regardless of the specification of the underlying regression model.

Figure 5 summarizes the estimated AUC_{t_0} for the models constructed with $\mathbf{Z}_{G+B}$ and the IncV of marker G . Compared with the Cox PH model, the time-specific GLM is better at

capturing different magnitudes of improvement by adding G for the long-term and short-term prediction. This demonstrates the importance of incorporating time-varying effects in risk modeling when the markers differentially affect the long-term and short-term risks.

Furthermore, our results illustrate that our proposed methods are able to accurately exhibit time-varying contribution from the new markers in risk prediction. For example, the IncV of G on top of (Z, B) in AUC via the time-specific GLM is 0.13 for $t_0 = 6$, which is almost 2.6 times of the IncV 0.049 for $t_0 = 10$. In contrast, naively taking the event indicators as binary outcomes and performing ROC analyses as those considered in the clinical literature will produce misleading results on the IncV. To see this, we also obtained such naive estimates with our simulated data, for which the CLR was used to estimate the risk model parameters as in the literature. The average naive estimate of the IncV in AUC is only 0.042, which is substantially lower than the AUC gain estimated from our proposed DIPW methods that account for both time-varying marker effects and time-varying prediction performance, especially for short-term risk prediction. The naive method would have incorrectly suggested that the gain by adding the genetic score is minimal. This comparison illustrates the appealing advantages of our method.

5. Remarks

In large cohort studies, multi-phase NCC designs are often considered in practice, where multiple sets of biomarkers are measured at different phases on sequentially selected subsamples of the study population. Quantifying the clinical utility of biomarkers ascertained in such data, especially their IncV over routine factors, is challenging. As illustrated in our example of the NHS RA study, the complex sampling schemes need to be accounted with fidelity in both the estimation and inference procedures in order to reach valid conclusion. In this paper, we proposed robust approaches to evaluate the IncV of new markers in risk prediction under a three-phase NCC design. Our methods also take into account the time-varying effects of the markers on disease risks. This is particularly helpful in situations where we do not have enough knowledge about how the risk factors are associated with the disease. Incorporating such information would help differentiate usefulness of these markers in improving short-term and long-term risk prediction. In addition, via a DIPW approach, our method is robust to model misspecification in the sense that the proposed inference for the accuracy parameters are valid regardless of the adequacy of the fitted models. We provided the aVARs of our proposed estimators, accounting for various sources of variations as well as correlations among the subcohort sampling. When the correlation is strong, the adjustment for the correlation in the formulae would be essential in preventing over-estimation of the variance, as confirmed in our simulation results.

For illustration, we focused on the accuracy measures in the ROC paradigm and predictive values. Our DIPW-based estimation procedures can be easily extended to other accuracy measures, for example, the integrated discrimination improvement (IDI) index and the net reclassification improvement (NRI) index (Pencina et al. 2008).

Our DIPW accuracy estimators require estimation of the censoring survival function $\mathcal{G}_{Z_c}^C(\cdot)$. Under independent censoring, $\mathcal{G}_{Z_c}^C(\cdot)$ can be estimated by the KM estimator. When the censoring depends on some covariates Z_c , flexible semiparametric models such as the PH model can be used to estimate $\mathcal{G}_{Z_c}^C(\cdot)$. We find through additional simulation studies, shown in Web Appendix D, that the KM-based estimates are reasonably robust to slight violation of the independent censoring assumption.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGEMENTS

The authors wish to thank the participants, investigators and study staff of the Nurses' Health Studies, headquartered at the Channing Laboratory, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA. This research is supported by the Natural Sciences and Engineering Research Council of Canada, P01 CA053996, R01 GM079330, R01 GM085047, U01 CA086368, and U54 HG007963.

References

- Anderson WF, Rosenberg PS, Menashe I, Mitani A, Pfeiffer RM. Age-related crossover in breast cancer incidence rates between black and white ethnic groups. *Journal of the National Cancer Institute*. 2008; 100:1804–1814. [PubMed: 19066264]
- Boekholdt SM, Hack CE, Sandhu MS, Luben R, Bingham SA, Wareham NJ, Peters RJ, Jukema JW, Day NE, Kastelein JJ, Khaw K-T. C-reactive protein levels and coronary artery disease incidence and mortality in apparently healthy men and women: the epic-norfolk prospective population study 1993-2003. *Atherosclerosis*. 2006; 187:415–422. [PubMed: 16257408]
- 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. Evaluating prognostic accuracy of biomarkers under nested case-control studies. *Biostatistics*. 2012; 13:89–100. [PubMed: 21856652]
- 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]
- Efron B. Estimating the error rate of a prediction rule: improvement on cross-validation. *Journal of the American Statistical Association*. 1983; 78:316–331.
- Gerds TA, Cai T, Schumacher M. The performance of risk prediction models. *Biometrical Journal*. 2008; 50:457–479. [PubMed: 18663757]
- Goldstein L, Langholz B. Asymptotic theory for nested case-control sampling in the Cox regression model. *The Annals of Statistics*. 1992; 20:1903–1928.
- Gram J, Bladbjerg E-M, Møller L, Sjøll A, Jespersen J. Tissue-type plasminogen activator and c-reactive protein in acute coronary heart disease. A nested case-control study. *Journal of Internal Medicine*. 2000; 247:205–212. [PubMed: 10692083]
- Gray RJ. Weighted analyses for cohort sampling designs. *Lifetime data analysis*. 2009; 15:24–40. [PubMed: 18712477]
- Heagerty P, Lumley T, Pepe M. Time-dependent ROC curves for censored survival data and a diagnostic marker. *Biometrics*. 2000; 56:337–344. [PubMed: 10877287]
- Hippisley-Cox J, Coupland C. Risk of myocardial infarction in patients taking cyclo-oxygenase-2 inhibitors or conventional non-steroidal anti-inflammatory drugs: population based nested case-control analysis. *Bmj*. 2005; 330:1366. [PubMed: 15947398]

- 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]
- Joslyn SA, Foote ML, Nasser K, Coughlin SS, Howe HL. Racial and ethnic disparities in breast cancer rates by age: Naaccr breast cancer project. *Breast Cancer Research and Treatment*. 2005; 92:97–105. [PubMed: 15986118]
- Källberg H, Jacobsen S, Bengtsson C, Pedersen M, Padyukov L, Garred P, Frisch M, Karlson E, Klareskog L, Alfredsson L. Alcohol consumption is associated with decreased risk of rheumatoid arthritis: results from two Scandinavian case–control studies. *Annals of the Rheumatic Diseases*. 2009; 68:222–227. [PubMed: 18535114]
- Karlson E, Chibnik L, Kraft P, Cui J, Keenan B, Ding B, Raychaudhuri S, Klareskog L, Alfredsson L, Plenge R. Cumulative association of 22 genetic variants with seropositive rheumatoid arthritis risk. *Annals of the Rheumatic Diseases*. 2010; 69:1077–1085. [PubMed: 20233754]
- Karlson E, Chibnik L, Tworoger S, Lee I, et al. Biomarkers of inflammation and development of rheumatoid arthritis in women from two prospective cohort studies. *Arthritis & Rheumatism*. 2009; 60:641–652. [PubMed: 19248103]
- Liao K, Alfredsson L, Karlson E. Environmental influences on risk for rheumatoid arthritis. *Current Opinion in Rheumatology*. 2009; 21:279–283. [PubMed: 19318947]
- Liu M, Lu W, Shore RE, Zeleniuch-Jacquotte A. Cox regression model with time-varying coefficients in nested case–control studies. *Biostatistics*. 2010; 11:693–706. [PubMed: 20525697]
- Pencina M, D’Agostino R Sr, D’Agostino R Jr, Vasan R. Evaluating the added predictive ability of a new marker: from area under the ROC curve to reclassification and beyond. *Statistics in Medicine*. 2008; 27:157–172. [PubMed: 17569110]
- Pepe, M. *The Statistical Evaluation of Medical Tests for Classification and Prediction*. Oxford University Press; USA: 2004.
- Prentice R, Breslow N. Retrospective studies and failure time models. *Biometrika*. 1978; 65:153–158.
- Raychaudhuri S, Remmers E, Lee A, Hackett R, Guiducci C, Burt N, Gianniny L, Korman B, Padyukov L, Kurreeman F, et al. Common variants at CD40 and other loci confer risk of rheumatoid arthritis. *Nature Genetics*. 2008; 40:1216–1223. [PubMed: 18794853]
- Robins J, Ritov Y. Toward a curse of dimensionality appropriate (CODA) asymptotic theory for semi-parametric models. *Statistics in Medicine*. 1997; 16:285–319. [PubMed: 9004398]
- Samuelsen S. A pseudolikelihood approach to analysis of nested case-control studies. *Biometrika*. 1997; 84:379–394.
- Stahl E, Raychaudhuri S, Remmers E, Xie G, Eyre S, Thomson B, Li Y, Kurreeman F, Zhernakova A, Hinks A, et al. Genome-wide association study meta-analysis identifies seven new rheumatoid arthritis risk loci. *Nature Genetics*. 2010; 42:508–514. [PubMed: 20453842]
- Sudbø J, Lee J, Lippman S, Mork J, Sagen S, Flatner N, Ristimäki A, Sudbø A, Mao L, Zhou X, et al. Non-steroidal anti-inflammatory drugs and the risk of oral cancer: a nested case-control study. *The Lancet*. 2005; 366:1359–1366.
- Thomas D, Liddell FDK, McDonald JC, Thomas DC. Addendum to: Methods of cohort analysis: appraisal by application to asbestos mining. *Journal of the Royal Statistical Society. Series A (General)*. 1977; 140:469–491.
- Uno H, Cai T, Tian L, Wei L. Evaluating prediction rules for t-year survivors with censored regression models. *Journal of the American Statistical Association*. 2007; 102:527–537.
- Wolfe F, Michaud K. The risk of myocardial infarction and pharmacologic and nonpharmacologic myocardial infarction predictors in rheumatoid arthritis: a cohort and nested case–control analysis. *Arthritis & Rheumatism*. 2008; 58:2612–2621. [PubMed: 18759273]
- Zhou W, Thurston SW, Liu G, Xu LL, Miller DP, Wain JC, Lynch TJ, Su L, Christiani DC. The interaction between microsomal epoxide hydrolase polymorphisms and cumulative cigarette smoking in different histological subtypes of lung cancer. *Cancer Epidemiology Biomarkers & Prevention*. 2001; 10:461–466.

Figure 1.
Sequential NCC sampling design in NHS.

Figure 2.

Analysis of NHS RA Data: estimated regression coefficients (circles) and their 95% confidence intervals (vertical bars) of (a) age (b) pack-years of smoking (c) cumulative alcohol intake (d) sTNFRII (e) IL-6 (f) GRS under the t_0 -year GLM with $\mathbf{Z}_{\mathbf{G}+\mathbf{B}}$ ($t_0 = 5, 6, 8, 10, 12, 14, 15$). Shown also are the estimates from the Cox PH model with $\mathbf{Z}_{\mathbf{G}+\mathbf{B}}$.

Figure 3.

Analysis of the NHS RA data: cross-validated point estimates (dots) along with 95% confidence intervals (vertical bars) of the IncV by adding $\{\mathbf{B}, \mathbf{G}\}$ on top of $\mathbf{Z}$ with respect to AUC_{t_0} , TPR_{t_0} , PPV_{t_0} and NPV_{t_0} at the cut-off values chosen to achieve $FPR_{t_0} = 0.1$ for $t_0 = 5, 6, 8, 10, 12, 14, 15$. The solid and dashed lines represent the IncV estimates corresponding to risk modeling with time-specific GLMs and PH models, respectively.

Figure 4.

Simulation Results: average point estimates (circles) of the regression coefficients for marker Z , B and G in the full time-specific GLM along with their empirical 95% confidence intervals (solid vertical bars) at $t_0 = 5, 6, 7, 8, 9, 10$. Shown also are the regression coefficient estimates along with their 95% confidence intervals from the PH model (diamond with dashed vertical bars) and from the CLR (triangle with dotted vertical bars). The dotdashed line represents the true values of the regression coefficients.

(a) AUC of the models with Z_{G+B} (b) IncV of G on top of Z_B **Figure 5.**

Simulation Results: average DIPW estimates (circles: time-specific GLM, diamonds: PH model) of the time-dependent AUC for the model with Z_{G+B} and the IncV in AUC of G on top of Z_B along with their empirical 95% confidence intervals (solid bars: time-specific GLM, dashed bars: PH model) for $t_0 = 5, 6, 7, 8, 9, 10$. Shown also are the naive estimates obtained from using the CLR to estimate the log hazard ratios but treats the data as standard case-control study to perform ROC analyses.