Optimal Cox Regression Subsampling Procedure with Rare Events

Nir Keret and Malka Gorfine
Department of Statistics and Operations Research
Tel Aviv University, Israel

March 16, 2021

Abstract

Massive sized survival datasets are becoming increasingly prevalent with the development of the healthcare industry. Such datasets pose computational challenges unprecedented in traditional survival analysis use-cases. A popular way for coping with massive datasets is downsampling them to a more manageable size, such that the computational resources can be afforded by the researcher. Cox proportional hazards regression has remained one of the most popular statistical models for the analysis of survival data to-date. This work addresses the settings of right censored and possibly left truncated data with rare events, such that the observed failure times constitute only a small portion of the overall sample. We propose Cox regression subsampling-based estimators that approximate their full-data partial-likelihood-based counterparts, by assigning optimal sampling probabilities to censored observations, and including all observed failures in the analysis. Asymptotic properties of the proposed estimators are established under suitable regularity conditions, and simulation studies are carried out to evaluate the finite sample performance of the estimators. Extensions to delayed-entry, time-dependent covariates, time-dependent coefficients and stratified analysis are also discussed and implemented. We further apply our procedure on UK-biobank colorectal cancer data with genetic and environmental risk factors.

keywords: A-optimal; Big data; L-optimal; Left-truncation; Loewner ordering; Optimal sampling; Rare events; Survival analysis.

1 Introduction

The Cox proportional hazards model (Cox, 1972) is one of the most popular methods in use in the context of time-to-event data. In order to estimate the model parameters, the second-order Newton-Raphson (NR) algorithm is typically utilized, and indeed this is the default optimizer in the widely used `coxph` function in the R (R Core Team, 2013) Survival package (Therneau, 2020), as well as in Python's Lifelines package (Davidson-Pilon et al., 2020). In modern applications, massive sized datasets with survival data become increasingly prevalent, with the number of observations go far beyond 10^6 (Mittal et al., 2014). The healthcare industry has been traditionally one of the principal generators of survival data, and the amount of data accumulated in that industry has been growing very rapidly in recent years (Raghupathi and Raghupathi, 2014). Massive datasets might pose a computational barrier to the analysis, due to the large number of observations and covariates. Many a time it is the case that the event of interest constitutes only a very small portion of the overall dataset, so that these failure times are referred to as "rare events". For instance, in the UK-Biobank (UKB) and in the China Kadoorie Biobank, there are about 2800 and 3000 incident cases of colorectal cancer, out of some 500,000 and 509,500 observations, respectively, yielding a failure rate of only about 0.6% (UK-Biobank, 2016; Pang et al., 2018).

In this work, our suggested approach to dealing with the computational challenge of massive data with rare events, is by the means of subsampling. In the context of rare events, our goal is to use a subsample of censored observations in the most efficient way, while retaining all observed failure times in the analyzed dataset. The rationale for this type of subsampling design is that the failure times contribute more information than the censored times, in addition to the potential instability that might result from a subsample with too few (even 0) failure times had the sampling been performed from the entire dataset irrespectively. Settings in which the number of observed failure times is large, such that a subsampling routine is required also for the failure times, calls for a different methodology to be developed, as it does not arise naturally as an extension of this work. We defer such work for future research.

Subsampling methods on the premise of massive data have been developed for other statistical models, for alleviating the computational burden. Dhillon et al. (2013) and Ma et al. (2015) proposed a subsampling algorithm for least squares regression; Wang et al. (2018) derived the optimal sampling probabilities for logistic regression; Ai et al. (2018), Wang and Ma (2020) and Yu et al. (2020) then extended it to generalized linear models, quantile regression and quasi-likelihood estimators, respectively. In the context of survival analysis, Zuo et al. (2020) derived the optimal sampling probabilities for the additive hazards model, under the non-rare-events setting. Practical usage of subsampling methods with survival

data for reducing the computational burden, was observed in Johansson et al. (2015) for Poisson regression, and in Gorfine et al. (2020) for a new semi-competing risks model. In these two works, all failure times were included in the subsample, while a subset of the censored observations was drawn using the inefficient uniform sampling probabilities.

In the models listed above with a derived optimal subsampling procedure, the estimating equations take the simple form of a sum over the observations. Derivation of the optimal sampling probabilities for the Cox regression is substantially more complicated, as the estimating equation is a sum of ratios, such that the observations appear both in the numerator and in the denominator of more than one ratio term. In addition, incorporating all failure times in the analysis, as we do under the rare events settings, produces unpredictable stochastic processes and standard martingale theory does not suffice for deriving all asymptotic properties.

Our suggested approach is related to the case-cohort (CC) design (Prentice, 1986), a well known design in the epidemiological literature. CC designs are put to practice when some covariates are too costly to procure for the entire cohort. As a remedy, only the observations that failed, termed "cases" and a random subset, uniformly sampled from the censored, termed "controls" are measured for their expensive covariates. Another common design aimed at the same goal, is the nested-case-control (NCC) design (Liddell et al., 1977), where all cases are retained, and for each case a small number of controls, typically ranging from 1 to 4, are uniformly sampled from its corresponding riskset. An improvement upon the classic NCC was suggested by Samuelsen (1997), by using each sampled control for more than one failure time. Since these methods are geared at a different use-case, they can be improved upon by a smart subsample selection, and indeed as our simulations show, the classic CC is suboptimal for our needs with regards to efficiency, and both NCC designs (classic and Samuelsen's) are sub-optimal in both computational and efficiency aspects. Some other works for improving the efficiency of classic CC designs (Chen and Lo, 1999; Kulich and Lin, 2004; Kim, Cai, and Lu, 2013; Zheng, Brown, Lok, and Cai, 2017) and of NCC designs (Langholz and Borgan, 1995; Keogh and White, 2013), were published, however these works are established on the grounds of missing covariates, and most importantly, they do not consider computational complexity as a key aspect.

This work presents a novel subsampling approach addressing the above mentioned computational challenge for time-to-event data with rare events. Our method is shown to be both highly efficient and computationally fast, by means of both simulations and theoretical rigor. We prove the consistency and asymptotic normality of the estimated regression coefficients and the estimated cumulative hazard function. Our method naturally accommodates delayed entry (left truncation), stratified analysis, time-dependent covariates and time-dependent coeffi-

cients, as discussed in Section 2.6.

One of the elegant computational devices in the `coxph` function, is the transformation of a dataset into time-dependent form, thus enabling easy accommodation of time-dependent covariates or time-dependent coefficients. Each observation is broken into several pseudo-observations, as elaborated in Therneau et al. (2017) and in Section 2.6.2. One of our novel findings is that consistent and efficient estimators can be achieved by means of sampling from the pseudo-observations, and not necessarily from the original observations. The implication is that it is efficient to use only informative parts of the observation's observed follow-up trajectory.

The performance of our method is demonstrated also for the delayed entry scenario and time-dependent covariates and coefficients both in simulations and in the real data analysis. The colorectal cancer (CRC) incidence cases of the UKB data will be analyzed with common environmental risk factors, and 72 single-nucleotide polymorphisms (SNPs) that have been identified by GWAS to be associated with CRC (Jeon et al., 2018). R code for the data analysis and reported simulations is available at Github site: `https://github.com/nirkeret/subsampling`.

2 Methodology

2.1 Notation and Model Formulation

Let V denote a failure time, C a censoring time and T the observed time $V \wedge C$. Denote $\Delta = I(V < C)$ and let $\mathbf{X}$ be a vector of possibly time-dependent covariates of size $r \times 1$, and for notational convenience $\mathbf{X}$ is used instead of $\mathbf{X}(t)$. Suppose there is a fixed number of independent and identically distributed (iid) observations, denoted n, then the observed data are $\mathcal{D}_n = \{T_i, \Delta_i, \mathbf{X}_i; i = 1, \ldots, n\}$. Out of the n observations, a random number, denoted n_c, are censored and $n_e = n - n_c = \sum_{i=1}^{n} \Delta_i$ have their failure times observed. It is assumed that n_e/n converges to a small positive constant as $n \to \infty$, and that r remains fixed. Let q be the number of censored observations sampled out the full data. The value of q is decided by the researchers, according to their computational limitations, and it is assumed that q is substantially smaller than n, such that $q/n \to 0$ as q and n go to infinity. Out of computational considerations, the sampling will be performed with replacement, and practically, since n_c/q is assumed very large, there should be little difference between sampling with and without replacement. However, when sampling without replacement, after each drawn observation, the remaining sampling probabilities should be updated. This sequential updating procedure makes it harder to derive an optimal policy for the subsampling probabilities from a theoretical point of view, and also greatly increase the runtime from a practical point of view. Denote $\mathcal{C}$ as the set of all censored observations in the

full data ($|\mathcal{C}| = n_c$), $\mathcal{E}$ as the set of all observations whose failure time was observed ($|\mathcal{E}| = n_e$) and $\mathcal{Q}$ as the set containing $\mathcal{E}$, in addition to all censored observations included in the subsample ($|\mathcal{Q}| = n_e + q$).

We assume throughout this work the Cox proportional-hazards model. Extensions of the Cox model will be addressed in Section 2.6. Let $\boldsymbol{\beta}$ be the vector of coefficients corresponding to $\mathbf{X}$, then the instantaneous hazard rate of observation i at time t takes the form

$$\lambda(t|\mathbf{X}_i) = \lambda_0(t)e^{\boldsymbol{\beta}^T\mathbf{X}_i},$$

where $\lambda_0(t)$ is an unspecified positive function, and $\Lambda_0(t)$ is the cumulative baseline hazard function $\int_0^t \lambda_0(u)du$. Denote $\boldsymbol{\beta}^o$,λ_0^o Λ_0^o as the true unknown values of $\boldsymbol{\beta}$, λ_0 and Λ_0, respectively. Let us now introduce the following notation,

$$S^{(0)}(\boldsymbol{\beta}, t) = \sum_{i=1}^{n} e^{\boldsymbol{\beta}^T\mathbf{X}_i} Y_i(t)\,,$$

$$\mathbf{S}^{(1)}(\boldsymbol{\beta}, t) = \sum_{i=1}^{n} e^{\boldsymbol{\beta}^T\mathbf{X}_i} Y_i(t)\mathbf{X}_i\,,$$

and,

$$\mathbf{S}^{(2)}(\boldsymbol{\beta}, t) = \sum_{i=1}^{n} e^{\boldsymbol{\beta}^T\mathbf{X}_i} Y_i(t)\mathbf{X}_i^{\otimes 2}\,,$$

where $\mathbf{X}^{\otimes 2} = \mathbf{X}\mathbf{X}^T$, and $Y_i(t) = I(T_i \geq t)$ is the indicator that observation i is at risk at time t. Denote $\hat{\boldsymbol{\beta}}_{PL}$ the full-sample partial-likelihood estimator for $\boldsymbol{\beta}$ and τ the maximal follow-up time. So, $\hat{\boldsymbol{\beta}}_{PL}$ is the vector that solves the following vectorial equation

$$\frac{\partial l(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}^T} = \sum_{i=1}^{n} \Delta_i \left\{ \mathbf{X}_i - \frac{\mathbf{S}^{(1)}(\boldsymbol{\beta}, T_i)}{S^{(0)}(\boldsymbol{\beta}, T_i)} \right\} = \mathbf{0}\,.$$

Let $\mathbf{p} = (p_1, \ldots, p_{n_c})^T$ be the vector of sampling probabilities for the censored observations, where $\sum_{i=1}^{n_c} p_i = 1$, and let us set

$$w_i = \begin{cases} \frac{1}{p_i q}, & \text{if } \Delta_i = 0, p_i > 0 \\ 1, & \text{if } \Delta_i = 1\,. \end{cases}$$

The subsample-based counterparts of the previously introduced notation, are

$$S_w^{(0)}(\boldsymbol{\beta}, t) = \sum_{i \in \mathcal{Q}} w_i e^{\boldsymbol{\beta}^T\mathbf{X}_i} Y_i(t)\,,$$

$$\mathbf{S}_w^{(1)}(\boldsymbol{\beta}, t) = \sum_{i \in \mathcal{Q}} w_i e^{\boldsymbol{\beta}^T \mathbf{X}_i} Y_i(t) \mathbf{X}_i \,,$$

and,

$$\mathbf{S}_w^{(2)}(\boldsymbol{\beta}, t) = \sum_{i \in \mathcal{Q}} w_i e^{\boldsymbol{\beta}^T \mathbf{X}_i} Y_i(t) \mathbf{X}_i^{\otimes 2} \,.$$

Now, suppose that a subsample of size q was drawn from the censored observations, then the estimator based on $\mathcal{Q}$, denoted $\tilde{\boldsymbol{\beta}}$, is the vector that solves

$$\frac{\partial l^*(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}^T} = \sum_{i \in \mathcal{Q}} \Delta_i \left\{ \mathbf{X}_i - \frac{\mathbf{S}_w^{(1)}(\boldsymbol{\beta}, T_i)}{S_w^{(0)}(\boldsymbol{\beta}, T_i)} \right\} = \mathbf{0} \,.$$

For a given vector of regression coefficients $\boldsymbol{\beta}$, let us define the function

$$\hat{\Lambda}_0(t, \boldsymbol{\beta}) = \sum_{i=1}^{n} \frac{\Delta_i I(T_i \leq t)}{S^{(0)}(\boldsymbol{\beta}, T_i)} \,,$$

and the Breslow estimator (Breslow, 1972) for the cumulative baseline hazard function is thus produced by $\hat{\Lambda}_0(t, \hat{\boldsymbol{\beta}}_{PL})$. In this work we propose Breslow-type estimators, such that different consistent estimators for $\boldsymbol{\beta}^o$ are plugged in, based on our optimal sampling procedure.

Denote R_i as the random variable which counts the number of times observation i was drawn into the subsample $\mathcal{Q}$, and $\mathbf{R} = (R_1, \ldots, R_n)^T$. Conditionally on $\mathcal{D}_n$, $R_i = 1$ if $\Delta_i = 1$, whereas $\mathbf{R}_c | \mathcal{D}_n \sim Multinomial(q, \mathbf{p})$, where $\mathbf{R}_c$ is the $n_c \times 1$ sub-vector of $\mathbf{R}$, corresponding to the censored observations. Using this notation one can replace, for example, $\sum_{i \in \mathcal{Q}} w_i e^{\boldsymbol{\beta}^T \mathbf{X}_i} Y_i(t) \mathbf{X}_i$ with $\sum_{i=1}^{n} w_i e^{\boldsymbol{\beta}^T \mathbf{X}_i} Y_i(t) \mathbf{X}_i R_i$, and both forms will be used interchangeably, depending on convenience. The two forms of presentation illuminate different aspects of the subsampling procedure in the conditional space, namely given $\mathcal{D}_n$. The first presentation is a sum of n_e events and q iid sampled censored observations (similarly to bootstrap sampling), while the second presentation, which spans the full data, emphasizes the dependency induced by sampling a fixed and predetermined number of observations. Additionally, let us introduce the usual counting process notation $N_i(t) = \Delta_i I(T_i \leq t)$, $N_{\cdot}(t) = \sum_{i=1}^{n} N_i(t)$, and finally, $\|.\|_2$ denotes the l_2 Euclidean norm.

2.2 Asymptotic Properties

In this section the asymptotic properties of a general subsampling-based estimator are worked out for any vector of conditionally deterministic probabilities, given the data, that conforms with assumption A.8, as presented below. It will be shown that in the conditional space, i.e. given the observed data, the subsampling-based estimator converges to the full-sample partial-likelihood estimator, while in the

unconditional space, it converges to the true parameter. Based on the results of this section, the optimal sampling probabilities will be derived, motivated by different optimality criteria.

Let us begin with a list of the required assumptions, some of them standard for Cox regression, and some unique to our case, which will be discussed after listing all the assumptions.

A.1 Boundedness of the cumulative hazard function, namely $\int_0^\tau \lambda_0^o(t)dt < \infty$.

A.2 There exists a neighbourhood $\mathcal{B}$ of $\boldsymbol{\beta}^o$ and functions $\mathbf{s}^{(j)}(\boldsymbol{\beta}, t), j = 0, 1, 2$, defined on $\mathcal{B} \times [0, \tau]$ such that for $j = 0, 1, 2$ as $n \to \infty$,

$$\sup_{t\in[0,\tau],\boldsymbol{\beta}\in\mathcal{B}} \frac{1}{n} \left\| \mathbf{S}^{(j)}(\boldsymbol{\beta}, t) - \mathbf{s}^{(j)}(\boldsymbol{\beta}, t) \right\|_2 \xrightarrow{P} 0 \,.$$

A.3 Boundedness of the covariates.

A.4 For all $\boldsymbol{\beta} \in \mathcal{B}$ and $t \in [0, \tau]$,

$$\partial s^{(0)}(\boldsymbol{\beta}, t)/(\partial\boldsymbol{\beta}) = \mathbf{s}^{(1)}(\boldsymbol{\beta}, t) \,,$$

and

$$\partial^2 s^{(0)}(\boldsymbol{\beta}, t)/(\partial\boldsymbol{\beta}^T\partial\boldsymbol{\beta}) = \mathbf{s}^{(2)}(\boldsymbol{\beta}, t) \,.$$

For $j = 0, 1, 2$, $\mathbf{s}^{(j)}(\boldsymbol{\beta}, t)$ is a continuous function of $\boldsymbol{\beta}$, uniformly in $t \in [0, \tau]$. The functions $\mathbf{s}^{(j)}(\boldsymbol{\beta}, t)$, $j = 0, 1, 2$, are bounded, and $s^{(0)}$ is bounded away from 0 on $\mathcal{B} \times [0, \tau]$.

A.5 The matrix

$$\boldsymbol{\Sigma}(\boldsymbol{\beta}^o) = \int_0^\tau \left[\mathbf{s}^{(2)}(\boldsymbol{\beta}^o, t) - \frac{\left\{\mathbf{s}^{(1)}(\boldsymbol{\beta}^o, t)\right\}^{\otimes 2}}{s^{(0)}(\boldsymbol{\beta}^o, t)} \right] \lambda_0(t)dt$$

is positive definite.

A.6 The failure time V and censoring time C are conditionally independent given the covariates $\mathbf{X}$. Additionally, when the data also include delayed entry, the entry and failure times are conditionally quasi-independent (Tsai, 1990) given the covariates $\mathbf{X}$.

A.7 Non-emptyness of the risk set. Namely, $E\{Y_i(\tau)\} > 0$ for all $i = 1, \ldots, n$.

A.1–A.7 are the standard assumptions for the consistency and asymptotical normality of Cox regression in the unconditional space.

A.8 As $n \to \infty$, $p_i n$ is bounded away from 0 for all $i \in \mathcal{C}$.

A.9 The Hessian matrices $\partial^2 l(\boldsymbol{\beta})/(\partial \boldsymbol{\beta}^T \partial \boldsymbol{\beta})$ and $\partial^2 l^*(\boldsymbol{\beta})/(\partial \boldsymbol{\beta}^T \partial \boldsymbol{\beta})$ are non-singular with probability going to 1, as $n \to \infty$ and $q \to \infty$.

A.10 n_e/n converges to a small positive constant as $n \to \infty$, and the number of sampled censored observations is substantially smaller than the full sample, $q = o(n)$.

Assumptions A.8-A.10 are specific to our case. A.8 guarantees that the sampling probabilities do not approach 0 too fast as the sample size increases. For instance, when the sampling probabilities are uniform, $p_i n = n/n_c$, and the condition is indeed satisfied. We do, however, allow for a set of censored observations be assigned a sampling probability of 0, as discussed below. A.9 ensures that as the sample size increases, the subsample-based and full-sample-based information matrices are invertible. A.10 means that on the premise of rare events, we intend to downsample the number of censored observations, to be substantially smaller than n_c.

Generally speaking, for the subsampling estimator to be consistent, each observation in the data should be assigned a sampling probability strictly greater than 0. This condition is a version of the so-called "positivity" assumption in the theory of inverse probability weighting (IPW) estimators. However, it should be recognized, that some observations contribute no information whatsoever to the partial-likelihood-based estimating equations, and no harm will be done if they are assigned a sampling probability of 0. For instance, observations that were censored before the first observed event time, contribute no information at all. Other cases where censored observations contribute no information will be discussed in the relevant context in section 2.6. We allow our methods to assign 0 sampling probabilities exactly to that subset of observations just mentioned, and the theory developed below holds also under this relaxation of the positivity condition. For those methods which allow 0 sampling probabilities, it can be viewed as if this subset of observations were removed in advance. This way, the partial-likelihood remains unaffected, but now for the theoretical results it may be assumed that no observation has a sampling probability of 0.

Theorem 1 establishes the consistency of the subsample-based estimators to their full sample counterparts, conditionally on the observed data, with convergence rates of $q^{-1/2}$.

Theorem 1. *Conditionally on $\mathcal{D}_n$ and given that A.3 and A.7–A.9 hold,*

$$\left\| \tilde{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}_{PL} \right\|_2 = O_{P|\mathcal{D}_n}(q^{-1/2}) \tag{1}$$

and, for each $t \in [0, \tau]$,

$$\hat{\Lambda}_0(t, \tilde{\boldsymbol{\beta}}) - \hat{\Lambda}_0(t, \hat{\boldsymbol{\beta}}_{PL}) = O_{P|\mathcal{D}_n}(q^{-1/2}) \,, \tag{2}$$

where $P|\mathcal{D}_n$ *stands for the conditional probability measure given* $\mathcal{D}_n$.

The proof of Theorem 1 requires the following three lemmas.

Lemma 1. *Based on A.3 and A.8, and for a fixed vector of coefficients* $\boldsymbol{\beta}$*, it holds that*

$$\sup_{t\in[0,\tau],\boldsymbol{\beta}\in\mathcal{B}} \frac{1}{n} \left| \mathbf{S}_w^{(k)}(\boldsymbol{\beta}, t) - \mathbf{S}^{(k)}(\boldsymbol{\beta}, t) \right| = O_{P|\mathcal{D}_n}(q^{-1/2}) \tag{3}$$

for $k = 0, 1, 2$.

For $k = 1, 2$, the lemma is in an element-wise sense in the respective vector/matrix.

Proof of Lemma 1. The proof will be provided for $k = 1$, but similar steps can be repeated for $k = 0, 2$, where in $k = 2$ it is done by treating a general element in the matrix. Let us first rewrite the expression in Eq.(3) in the following manner

$$\frac{1}{n} \left\{ \mathbf{S}_w^{(1)}(\boldsymbol{\beta}, t) - \mathbf{S}^{(1)}(\boldsymbol{\beta}, t) \right\} = \frac{1}{n} \sum_{i=1}^{n} e^{\boldsymbol{\beta}^T \mathbf{X}_i} Y_i(t) \mathbf{X}_i \left(w_i R_i - 1 \right) \,, \tag{4}$$

and observe that its conditional expectation is 0, since $E(R_i|\mathcal{D}_n) = w_i^{-1}$. Examining the conditional variance, it holds that,

$$\begin{aligned}
&Var \left[\frac{1}{n} \left\{ \mathbf{S}_w^{(1)}(\boldsymbol{\beta}, t) - \mathbf{S}^{(1)}(\boldsymbol{\beta}, t) \right\} |\mathcal{D}_n \right] \\
&= \frac{1}{n^2} \left\{ \sum_{i\in\mathcal{C}} \frac{1}{p_i q} e^{2\boldsymbol{\beta}^T \mathbf{X}_i} Y_i(t) \mathbf{X}_i^{\otimes 2} - \sum_{i,j\in\mathcal{C}} \frac{1}{q} e^{\boldsymbol{\beta}^T (\mathbf{X}_i + \mathbf{X}_j)} Y_i(t) Y_j(t) \mathbf{X}_i \mathbf{X}_j^T \right\} \\
&= O_{|\mathcal{D}_n}(q^{-1}) \,,
\end{aligned}$$

where $O_{|\mathcal{D}_n}$ stands for the standard big-O notation (not in the probabilistic sense), in the conditional space. The last equality stems from A.3 and A.8, and based on Chebyshev's inequality, $n^{-1}|\mathbf{S}_w^{(1)}(\boldsymbol{\beta}, t) - \mathbf{S}^{(1)}(\boldsymbol{\beta}, t)| = O_{P|\mathcal{D}_n}(q^{-1/2})$. Finally, the time t affects only the deterministic part in Eq.(4), and since it is bounded, the result holds also for the supremum over t, and the proof of Lemma 1 is complete. □

Lemma 2. *If A.3 and A.7–A.8 are satisfied, then conditionally on $\mathcal{D}_n$ it holds that*

$$\frac{1}{n}\frac{\partial l^*(\hat{\boldsymbol{\beta}}_{PL})}{\partial \boldsymbol{\beta}^T} = O_{P|\mathcal{D}_n}(q^{-1/2})\,,$$

in an element-wise sense.

For clarification, $\partial l^*(\hat{\boldsymbol{\beta}}_{PL})/\partial\boldsymbol{\beta}^T = \partial l^*(\boldsymbol{\beta})/\partial\boldsymbol{\beta}^T|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}_{PL}}$, and this style is adopted throughout this work for notational convenience. Lemma 2 shows that as q and n go to infinity, the subsample-based pseudo-score function tends to 0 at the value $\hat{\boldsymbol{\beta}}_{PL}$, the full-sample partial-likelihood estimator.

Proof of Lemma 2. First, let us write

$$\frac{1}{n}\frac{\partial l^*(\hat{\boldsymbol{\beta}}_{PL})}{\partial \boldsymbol{\beta}^T} = \frac{1}{n}\int_0^\tau \left\{\mathbf{X}_i - \frac{n^{-1}\mathbf{S}_w^{(1)}(\hat{\boldsymbol{\beta}}_{PL},t)}{n^{-1}S_w^{(0)}(\hat{\boldsymbol{\beta}}_{PL},t)}\right\} dN_{\cdot}(t)\,. \tag{5}$$

In the conditional space, $N_{\cdot}(t)$ is deterministic for all t, and so is $\hat{\boldsymbol{\beta}}_{PL}$. For the integrand of Eq.(5), a first order Taylor expansion about $\left(n^{-1}S^{(0)}(\hat{\boldsymbol{\beta}}_{PL},t), n^{-1}\mathbf{S}^{(1)T}(\hat{\boldsymbol{\beta}}_{PL},t)\right)^T$ yields

$$\begin{aligned}
&\frac{1}{n}\int_0^\tau \left[\mathbf{X}_i - \frac{n^{-1}\mathbf{S}^{(1)}(\hat{\boldsymbol{\beta}}_{PL},t)}{n^{-1}S^{(0)}(\hat{\boldsymbol{\beta}}_{PL},t)} + \frac{1}{\eta_t}n^{-1}\left\{\mathbf{S}_w^{(1)}(\hat{\boldsymbol{\beta}}_{PL},t) - \mathbf{S}^{(1)}(\hat{\boldsymbol{\beta}}_{PL},t)\right\}\right.\\
&\quad \left. -\frac{\boldsymbol{\xi}_t}{\eta_t^2}n^{-1}\left\{S_w^{(0)}(\hat{\boldsymbol{\beta}}_{PL},t) - S^{(0)}(\hat{\boldsymbol{\beta}}_{PL},t)\right\}\right] dN_{\cdot}(t)\\
&= \frac{1}{n}\frac{\partial l(\hat{\boldsymbol{\beta}}_{PL})}{\partial \boldsymbol{\beta}^T} + \frac{1}{n}\int_0^\tau \left[\frac{1}{\eta_t}n^{-1}\left\{\mathbf{S}_w^{(1)}(\hat{\boldsymbol{\beta}}_{PL},t) - \mathbf{S}^{(1)}(\hat{\boldsymbol{\beta}}_{PL},t)\right\}\right.\\
&\quad \left. -\frac{\boldsymbol{\xi}_t}{\eta_t^2}n^{-1}\left\{S_w^{(0)}(\hat{\boldsymbol{\beta}}_{PL},t) - S^{(0)}(\hat{\boldsymbol{\beta}}_{PL},t)\right\}\right] dN_{\cdot}(t)\,,
\end{aligned}$$

where $(\eta_t, \boldsymbol{\xi}_t^T)^T$ is on the line segment between $\left(n^{-1}S_w^{(0)}(\hat{\boldsymbol{\beta}}_{PL},t), n^{-1}\mathbf{S}_w^{(1)T}(\hat{\boldsymbol{\beta}}_{PL},t)\right)^T$ and $\left(n^{-1}S^{(0)}(\hat{\boldsymbol{\beta}}_{PL},t), n^{-1}\mathbf{S}^{(1)T}(\hat{\boldsymbol{\beta}}_{PL},t)\right)^T$. Thanks to continuity, η_t and $\boldsymbol{\xi}_t$ converge to $n^{-1}\mathbf{S}^{(k)}(\hat{\boldsymbol{\beta}}_{PL},t)$, $k=0,1$, respectively, and therefore η_t^{-1} and $\boldsymbol{\xi}_t\eta_t^{-2}$ are conditionally bounded in probability due to the continuous mapping theorem, A.3 and A.7. Therefore, based on Lemma 1, we have proven Lemma 2. □

For Lemma 3 denote

$$\boldsymbol{\mathcal{I}}(\boldsymbol{\beta}) \;=\; \frac{1}{n}\frac{\partial^2 l(\boldsymbol{\beta})}{\partial\boldsymbol{\beta}^T\partial\boldsymbol{\beta}} = -\frac{1}{n}\int_0^\tau \left\{\frac{\mathbf{S}^{(2)}(\boldsymbol{\beta},t)}{S^{(0)}(\boldsymbol{\beta},t)} - \left(\frac{\mathbf{S}^{(1)}(\boldsymbol{\beta},t)}{S^{(0)}(\boldsymbol{\beta},t)}\right)^{\otimes 2}\right\} dN_{\cdot}(t)$$

and,

$$\tilde{\boldsymbol{\mathcal{I}}}(\boldsymbol{\beta}) \;=\; \frac{1}{n}\frac{\partial^2 l^*(\boldsymbol{\beta})}{\partial\boldsymbol{\beta}^T\partial\boldsymbol{\beta}} = -\frac{1}{n}\int_0^\tau \left\{ \frac{\mathbf{S}_w^{(2)}(\boldsymbol{\beta},t)}{S_w^{(0)}(\boldsymbol{\beta},t)} - \left(\frac{\mathbf{S}_w^{(1)}(\boldsymbol{\beta},t)}{S_w^{(0)}(\boldsymbol{\beta},t)}\right)^{\otimes 2} \right\} dN_{\cdot}(t)\,.$$

Additionally, denote,

$$S_k^{(1)}(\boldsymbol{\beta},t) = \sum_{i=1}^n e^{\boldsymbol{\beta}^T\mathbf{X}_i} Y_i(t) X_{ik}\,,$$

$$S_{wk}^{(1)}(\boldsymbol{\beta},t) = \sum_{i=1}^n w_i e^{\boldsymbol{\beta}^T\mathbf{X}_i} Y_i(t) X_{ik} R_i\,,$$

where X_{ik} stands for the k'th element in the vector $\mathbf{X}_i$, $k = 1,\ldots,r$.

Lemma 3. *If A.3 and A.7–A.8 are satisfied, then conditionally on $\mathcal{D}_n$, for a vector of fixed coefficients $\boldsymbol{\beta}$,*

$$\tilde{\boldsymbol{\mathcal{I}}}(\boldsymbol{\beta}) - \boldsymbol{\mathcal{I}}(\boldsymbol{\beta}) = O_{P|\mathcal{D}_n}(q^{-1/2})\,, \tag{6}$$

in the sense that it holds for each element in the matrix.

Lemma 3 shows that the subsample-based observed information matrix for $\boldsymbol{\beta}$ converges to the corresponding observed information matrix based on the full sample.

Proof of Lemma 3.

$$\begin{aligned}\tilde{\boldsymbol{\mathcal{I}}}(\boldsymbol{\beta}) - \boldsymbol{\mathcal{I}}(\boldsymbol{\beta}) \;=\; & \frac{1}{n}\int_0^\tau \left\{ \frac{\mathbf{S}^{(2)}(\boldsymbol{\beta},t)}{S^{(0)}(\boldsymbol{\beta},t)} - \frac{\mathbf{S}_w^{(2)}(\boldsymbol{\beta},t)}{S_w^{(0)}(\boldsymbol{\beta},t)} \right\} dN_{\cdot}(t) \\ & -\frac{1}{n}\int_0^\tau \left[\left\{ \frac{\mathbf{S}^{(1)}(\boldsymbol{\beta},t)}{S^{(0)}(\boldsymbol{\beta},t)} \right\}^{\otimes 2} - \left\{ \frac{\mathbf{S}_w^{(1)}(\boldsymbol{\beta},t)}{S_w^{(0)}(\boldsymbol{\beta},t)} \right\}^{\otimes 2} \right] dN_{\cdot}(t)\,.\end{aligned}$$

The first addend can be shown to be $O_{P|\mathcal{D}_n}(q^{-1/2})$ in an identical manner to Lemma 1, by examining a general element in the matrix. We now turn to showing that the second addend is also $O_{P|\mathcal{D}_n}(q^{-1/2})$. Let us examine a general element in the matrix, in the m'th row and the l'th column and use a Taylor expansion for

$$n^{-1}S_{wm}^{(1)}(\boldsymbol{\beta},t)n^{-1}S_{wl}^{(1)}(\boldsymbol{\beta},t)/\{n^{-1}S_w^{(0)}(\boldsymbol{\beta},t)\}\,.$$

That is,

$$\begin{aligned}
&\frac{1}{n}\int_0^\tau \left[\left\{\frac{n^{-1}\mathbf{S}^{(1)}(\boldsymbol{\beta},t)}{n^{-1}S^{(0)}(\boldsymbol{\beta},t)}\right\}_{ml}^{\otimes 2} - \left\{\frac{n^{-1}\mathbf{S}_w^{(1)}(\boldsymbol{\beta},t)}{n^{-1}S_w^{(0)}(\boldsymbol{\beta},t)}\right\}_{ml}^{\otimes 2}\right] dN_\cdot(t) \\
&= \frac{1}{n}\int_0^\tau \left[\frac{\xi_t}{\omega_t^2}\left\{n^{-1}S_{wm}^{(1)}(\boldsymbol{\beta},t) - n^{-1}S_m^{(1)}(\boldsymbol{\beta},t)\right\}\right. \\
&\quad + \frac{\eta_t}{\omega_t^2}\left\{n^{-1}S_{wl}^{(1)}(\boldsymbol{\beta},t) - n^{-1}S_l^{(1)}(\boldsymbol{\beta},t)\right\} \\
&\quad \left. - \frac{\xi_t\eta_t}{\omega_t^3}\left\{n^{-1}S_w^{(0)}(\boldsymbol{\beta},t) - n^{-1}S^{(0)}(\boldsymbol{\beta},t)\right\}\right] dN_\cdot(t)\,,
\end{aligned}$$

where $(\eta_t,\, \xi_t,\, \omega_t)^T$ is on the line segment connecting $\left(n^{-1}S_{wm}^{(1)}(\boldsymbol{\beta},t), n^{-1}S_{wl}^{(1)}(\boldsymbol{\beta},t),\right.$ $\left. n^{-1}S_w^{(0)}(\boldsymbol{\beta},t)\right)^T$ with $\left(n^{-1}S_m^{(1)}(\boldsymbol{\beta},t), n^{-1}S_l^{(1)}(\boldsymbol{\beta},t), n^{-1}S^{(0)}(\boldsymbol{\beta},t)\right)^T$. Based on the continuous mapping theorem, Lemma 1, A.3 and A.7, Eq.(6) follows. □

We are now in position to prove Eq.(1) of Theorem 1.

Proof of Theorem 1(1). First, let us introduce some additional notation as follows,

$$S_{wkl}^{(1)}(\boldsymbol{\beta},t) = \sum_{i=1}^n w_i e^{\boldsymbol{\beta}^T\mathbf{X}_i} Y_i(t) X_{ik} X_{il} R_i\,,$$

and

$$S_{wklm}^{(1)}(\boldsymbol{\beta},t) = \sum_{i=1}^n w_i e^{\boldsymbol{\beta}^T\mathbf{X}_i} Y_i(t) X_{ik} X_{il} X_{im} R_i\,.$$

Denote $l_k^{*'}(\boldsymbol{\beta})$ as the derivative of $l^*(\boldsymbol{\beta})$ with respect to β_k, and $\tilde{\boldsymbol{\mathcal{I}}}_k$ as the k'th row of the matrix $\tilde{\boldsymbol{\mathcal{I}}}$. A first order Taylor expansion for $n^{-1}l_k^{*'}(\tilde{\boldsymbol{\beta}})$ about $\hat{\boldsymbol{\beta}}_{PL}$ is

$$0 = \frac{1}{n}l_k^{*'}(\tilde{\boldsymbol{\beta}}) = \frac{1}{n}l_k^{*'}(\hat{\boldsymbol{\beta}}_{PL}) + \tilde{\boldsymbol{\mathcal{I}}}_k^T(\hat{\boldsymbol{\beta}}_{PL})(\tilde{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}_{PL}) + \frac{1}{n}\mathrm{Res}_k \tag{7}$$

and

$$\frac{1}{n}\mathrm{Res}_k = (\tilde{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}_{PL})^T \frac{\partial\tilde{\boldsymbol{\mathcal{I}}}_k(\boldsymbol{\xi})}{\partial\boldsymbol{\beta}^T}(\tilde{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}_{PL})\,,$$

where $\boldsymbol{\xi}$ is on the line segment between $\tilde{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\beta}}_{PL}$, $k = 1,\ldots,r$. Examining a general element of $\partial\tilde{\boldsymbol{\mathcal{I}}}_k(\boldsymbol{\beta})/\partial\boldsymbol{\beta}^T$, we get

$$\begin{aligned}
\frac{1}{n}\frac{\partial^2 l_k^{*'}(\boldsymbol{\beta})}{\partial\boldsymbol{\beta}_l\partial\boldsymbol{\beta}_k} &= \frac{1}{n}\int_0^\tau \frac{1}{S_w^{(0)3}(\boldsymbol{\beta},t)}\left[S_{wklm}^{(1)}(\boldsymbol{\beta},t)S_w^{(0)2}(\boldsymbol{\beta},t) + \left\{S_{wkl}^{(1)}(\boldsymbol{\beta},t)S_{wm}^{(1)}(\boldsymbol{\beta},t)\right.\right. \\
&\quad \left. + S_{wkm}^{(1)}(\boldsymbol{\beta},t)S_{wl}^{(1)}(\boldsymbol{\beta},t) + S_{wlm}^{(1)}(\boldsymbol{\beta},t)S_{wk}^{(1)}(\boldsymbol{\beta},t)\right\} S_w^{(0)}(\boldsymbol{\beta},t) \\
&\quad \left. - 2S_{wk}^{(1)}(\boldsymbol{\beta},t)S_{wl}^{(1)}(\boldsymbol{\beta},t)S_{wm}^{(1)}(\boldsymbol{\beta},t)\right] dN_\cdot(t)\,.
\end{aligned}$$

Now it will be shown that $\partial\tilde{\boldsymbol{\mathcal{I}}}_k(\boldsymbol{\xi})/\partial\boldsymbol{\beta}^T$ is bounded in conditional probability. Let us note that based on Lemma 1, the continuous mapping theorem and A.7, the subsample-based pseudo-score function $n^{-1}\partial l^*(\boldsymbol{\beta})/\partial\boldsymbol{\beta}^T$ converges in conditional probability to the full sample score function $n^{-1}\partial l(\boldsymbol{\beta})\partial\boldsymbol{\beta}^T$. Since $\mathcal{B}$ is compact and the function $l(\boldsymbol{\beta})$ is continuous and convex, $\hat{\boldsymbol{\beta}}_{PL}$ is its unique global maximizer, then based on van der Vaart (1998, Theorem 5.9, page 46) we know that $\tilde{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}_{PL} = o_{P|\mathcal{D}_n}(1)$. This result establishes that $\tilde{\boldsymbol{\beta}}$ is indeed consistent for $\hat{\boldsymbol{\beta}}_{PL}$ in the conditional space, but does not teach us about its rate of convergence. Due to continuity, and since $\boldsymbol{\xi}$ is on the line segment between $\tilde{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\beta}}_{PL}$, it follows that $\boldsymbol{\xi} - \hat{\boldsymbol{\beta}}_{PL} = o_{P|\mathcal{D}_n}(1)$, hence $\boldsymbol{\xi} = O_{P|\mathcal{D}_n}(1)$. Based on this observation, A.3 and A.8, one can verify that the functions $n^{-1}\mathbf{S}_w^{(k)}(\boldsymbol{\xi}, t)$, $k = 0, 1, 2$, are all bounded in conditional probability, being the mean of q conditionally independent and bounded addends, plus the sum of n_e bounded constants divided by n. Namely, let us write

$$n^{-1}\mathbf{S}_w^{(0)}(\boldsymbol{\xi}, t) = \frac{1}{q}\sum_{i=1}^{q}\frac{e^{\boldsymbol{\xi}^T X_i^*}Y_i^*(t)}{np_i^*} + \frac{1}{n}\sum_{i\in\mathcal{E}} e^{\boldsymbol{\xi}^T X_i}Y_i(t)\,,$$

where the "*" sign denotes sampling with replacement, and similarly the same presentation holds for $n^{-1}\mathbf{S}_w^{(k)}(\boldsymbol{\xi}, t)$, $k = 1, 2$, $n^{-1}S_{wkl}^{(1)}(\boldsymbol{\xi}, t)$ and $n^{-1}S_{wklm}^{(1)}(\boldsymbol{\xi}, t)$. Therefore, and also based on A.7, we find that $\partial\tilde{\boldsymbol{\mathcal{I}}}_k(\boldsymbol{\xi})/\partial\boldsymbol{\beta}^T = O_{P|\mathcal{D}_n}(1)$, and as a result, for all $k = 1, \ldots, r$,

$$\frac{1}{n}\text{Res}_k = O_{P|\mathcal{D}_n}\left(\|\tilde{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}_{PL}\|_2^2\right)\,. \tag{8}$$

From Eq.'s (7)–(8) and A.9 we obtain

$$\tilde{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}_{PL} = -\tilde{\boldsymbol{\mathcal{I}}}^{-1}(\hat{\boldsymbol{\beta}}_{PL})\left\{\frac{1}{n}\frac{\partial l^*(\hat{\boldsymbol{\beta}}_{PL})}{\partial\boldsymbol{\beta}^T} + O_{P|\mathcal{D}_n}\left(\|\tilde{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}_{PL}\|_2^2\right)\right\}\,. \tag{9}$$

Since matrix inversion is a continuous operation, then due to Lemma 3 and the continuous mapping theorem, $\tilde{\boldsymbol{\mathcal{I}}}^{-1}(\hat{\boldsymbol{\beta}}_{PL}) - \boldsymbol{\mathcal{I}}^{-1}(\hat{\boldsymbol{\beta}}_{PL}) = o_{P|\mathcal{D}_n}(1)$, which yields $\tilde{\boldsymbol{\mathcal{I}}}^{-1}(\hat{\boldsymbol{\beta}}_{PL}) = O_{P|\mathcal{D}_n}(1)$. Therefore, combining Lemmas 1–2 and Eq.(9), it holds that

$$\tilde{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}_{PL} = O_{P|\mathcal{D}_n}(q^{-1/2}) + o_{P|\mathcal{D}_n}\left(\|\tilde{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}_{PL}\|_2\right)\,,$$

hence,

$$\tilde{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}_{PL} = O_{P|\mathcal{D}_n}(q^{-1/2})\,, \tag{10}$$

which completes the proof of Eq.(1) in Theorem 1. □

Proof of Theorem 1(2). Let us write

$$\hat{\Lambda}_0(t,\tilde{\boldsymbol{\beta}}) - \hat{\Lambda}_0(t,\hat{\boldsymbol{\beta}}_{PL}) = \frac{1}{n}\int_0^t \left\{ \frac{1}{n^{-1}S^{(0)}(\tilde{\boldsymbol{\beta}},u)} - \frac{1}{n^{-1}S^{(0)}(\hat{\boldsymbol{\beta}}_{PL},u)} \right\} dN_{\cdot}(u)\,.$$

Using a Taylor expansion about $\hat{\boldsymbol{\beta}}_{PL}$ yields

$$\frac{1}{n}(\tilde{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}_{PL}) \int_0^t \frac{S^{(1)}(\boldsymbol{\xi},u)}{n^{-1}S^{(0)2}(\boldsymbol{\xi},u)} dN_{\cdot}(u)\,, \tag{11}$$

where $\boldsymbol{\xi}$ is on the line segment between $\tilde{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\beta}}_{PL}$. Based on A.3, A.7 and Eq.(10), Eq.(11) is a product of a term which is $O_{P|\mathcal{D}_n}(q^{-1/2})$ and a term bounded in conditional probability, thus proving Eq.(2) and completing Theorem 1. □

We now turn to stating and proving Theorem 2, which concerns the asymptotic properties of the subsample-based estimators in the unconditional space.

Theorem 2. *Given that A.1–A.10 hold, then as* $q \to \infty$ *and* $n \to \infty$,

$$\sqrt{q}\mathbb{V}_{\tilde{\boldsymbol{\beta}}}(\mathbf{p},\boldsymbol{\beta}^o)^{-1/2}(\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta}^o) \xrightarrow{D} N(0,I) \tag{12}$$

and, for all $t \in [0,\tau]$,

$$\sqrt{q}\mathbb{V}_{\hat{\Lambda}_0(t,\tilde{\boldsymbol{\beta}})}(\mathbf{p},\boldsymbol{\beta}^o,t)^{-1/2}\{\hat{\Lambda}_0(t,\tilde{\boldsymbol{\beta}}) - \Lambda_0^o(t)\} \xrightarrow{D} N(0,1) \tag{13}$$

where

$$\mathbb{V}_{\tilde{\boldsymbol{\beta}}}(\mathbf{p},\boldsymbol{\beta}) = \frac{q}{n}\boldsymbol{\mathcal{I}}^{-1}(\boldsymbol{\beta}) + \boldsymbol{\mathcal{I}}^{-1}(\boldsymbol{\beta})\boldsymbol{\varphi}(\mathbf{p},\boldsymbol{\beta})\boldsymbol{\mathcal{I}}^{-1}(\boldsymbol{\beta})\,,$$

$$\boldsymbol{\varphi}(\mathbf{p},\boldsymbol{\beta}) = \frac{1}{n^2}\left\{ \sum_{i\in\mathcal{C}} \frac{\mathbf{a}_i(\boldsymbol{\beta})\mathbf{a}_i(\boldsymbol{\beta})^T}{p_i} - \sum_{i,j\in\mathcal{C}} \mathbf{a}_i(\boldsymbol{\beta})\mathbf{a}_j(\boldsymbol{\beta})^T \right\}\,,$$

$$\mathbf{a}_i(\boldsymbol{\beta}) = \int_0^\tau \left\{ \mathbf{X}_i - \frac{\mathbf{S}^{(1)}(\boldsymbol{\beta},t)}{S^{(0)}(\boldsymbol{\beta},t)} \right\} \frac{Y_i(t)e^{\boldsymbol{\beta}^T\mathbf{X}_i}}{S^{(0)}(\boldsymbol{\beta},t)} dN_{\cdot}(t)\,,$$

$$\mathbb{V}_{\hat{\Lambda}_0(t,\tilde{\boldsymbol{\beta}})}(\mathbf{p},\boldsymbol{\beta},t) = \frac{q}{n}\int_0^t \frac{dN_{\cdot}(u)}{n^{-1}S^{(0)2}(\boldsymbol{\beta},u)} + \mathbf{H}^T(\boldsymbol{\beta},t)\mathbb{V}_{\tilde{\boldsymbol{\beta}}}(\mathbf{p},\boldsymbol{\beta})\mathbf{H}(\boldsymbol{\beta},t)\,,$$

and,

$$\mathbf{H}(\boldsymbol{\beta},t) = \int_0^t \frac{\mathbf{S}^{(1)}(\boldsymbol{\beta},u)}{S^{(0)2}(\boldsymbol{\beta},u)} dN_{\cdot}(u)\,.$$

First, let us show that $\tilde{\boldsymbol{\beta}}$ is consistent to $\boldsymbol{\beta}^o$.

Lemma 4. *Given that A.1–A.9 hold, then as* $q \to \infty$ *and* $n \to \infty$,

$$\lim_{n,q\to\infty} \Pr(\|\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta}^o\|_2 > \epsilon) = 0 \,, \forall \epsilon > 0 \,. \tag{14}$$

Proof of Lemma 4. In Theorem 1 it was established that for all $\epsilon > 0$,

$$\lim_{n,q\to\infty} \Pr\left(\|\tilde{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}_{PL}\|_2 \geq \epsilon | \mathcal{D}_n\right) = 0 \,.$$

In the unconditional probability space, $\Pr\left(\|\tilde{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}_{PL}\|_2 \geq \epsilon | \mathcal{D}_n\right)$ is itself a random variable, let us denote it as $\pi_{n,q}$. It then follows that

$$\Pr\left(\lim_{n,q\to\infty} \pi_{n,q} = 0\right) = 1 \,,$$

meaning that $\pi_{n,q} \xrightarrow[n,q\to\infty]{a.s} 0$. Now we have that for all $\epsilon > 0$,

$$\lim_{n,q\to\infty} \Pr\left(\|\tilde{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}_{PL}\|_2 > \epsilon\right) = \lim_{n,q\to\infty} E(\pi_{n,q}) = E(\lim_{n,q\to\infty} \pi_{n,q}) = 0 \,, \tag{15}$$

where the interchange of expectation and limit is allowed due to the dominated convergence theorem, since $\pi_{n,q}$ is trivially bounded by 1. We proceed and write

$$\begin{aligned}
\Pr\left(\|\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta}^o\|_2 \geq \epsilon\right) &= \Pr\left(\|\tilde{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}_{PL} + \hat{\boldsymbol{\beta}}_{PL} - \boldsymbol{\beta}^o\|_2 \geq \epsilon\right) \\
&\leq \Pr\left(\|\tilde{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}_{PL}\|_2 + \|\hat{\boldsymbol{\beta}}_{PL} - \boldsymbol{\beta}^o\|_2 \geq \epsilon\right) \\
&\leq \Pr\left(\{\|\tilde{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}_{PL}\|_2 \geq \epsilon/2\} \cup \{\|\hat{\boldsymbol{\beta}}_{PL} - \boldsymbol{\beta}^o\|_2 \geq \epsilon/2\}\right) \\
&\leq \Pr\left(\|\tilde{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}_{PL}\|_2 \geq \epsilon/2\right) + \Pr\left(\|\hat{\boldsymbol{\beta}}_{PL} - \boldsymbol{\beta}^o\|_2 \geq \epsilon/2\right) \,.
\end{aligned}$$

Taking limits on both sides,

$$\begin{aligned}
&\lim_{n,q\to\infty} \Pr\left(\|\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta}^o\|_2 \geq \epsilon\right) \\
&\quad \leq \lim_{n,q\to\infty} \Pr\left(\|\tilde{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}_{PL}\|_2 \geq \epsilon/2\right) + \lim_{n,q\to\infty} \Pr\left(\|\hat{\boldsymbol{\beta}}_{PL} - \boldsymbol{\beta}^o\|_2 \geq \epsilon/2\right) = 0 \,,
\end{aligned}$$

where the first addend on the right is 0 due to Eq.(15), and the second addend is 0 based on known results for Cox regression and A.1–A.7. By that we have shown Eq.(14). □

Proof of Theorem 2. Similarly to Eq.(9), and based on Lemma 4, let us use a Taylor expansion for the subsample-based pseudo-score function evaluated at $\tilde{\boldsymbol{\beta}}$, about $\boldsymbol{\beta}^o$, and get

$$\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta}^o = -\tilde{\boldsymbol{\mathcal{I}}}^{-1}(\boldsymbol{\beta}^o) \left\{ \frac{1}{n} \frac{\partial l^*(\boldsymbol{\beta}^o)}{\partial \boldsymbol{\beta}^T} + o_P\left(\|\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta}^o\|_2 \right) \right\} . \tag{16}$$

Similarly to Samuelsen (1997), we decompose the subsample-based pseudo-score function into two separate components,

$$\frac{1}{n} \frac{\partial l^*(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}^T} = \frac{1}{n} \frac{\partial l(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}^T} + \frac{1}{n} \sum_{i \in \mathcal{C}} (1 - w_i R_i) \breve{\mathbf{a}}_i(\boldsymbol{\beta}) , \tag{17}$$

where

$$\breve{\mathbf{a}}_i(\boldsymbol{\beta}) = \int_0^\tau \left\{ \mathbf{X}_i - \frac{\mathbf{S}^{(1)}(\boldsymbol{\beta}, t)}{S^{(0)}(\boldsymbol{\beta}, t)} \right\} \frac{Y_i(t) e^{\boldsymbol{\beta}^T \mathbf{X}_i}}{S_w^{(0)}(\boldsymbol{\beta}, t)} dN_\cdot(t) .$$

Verifying Eq. (17),

$$\begin{aligned}
\sum_{i \in \mathcal{C}} (1 - w_i R_i) \breve{\mathbf{a}}_i(\boldsymbol{\beta}) &= \sum_{i=1}^n (1 - w_i R_i) \breve{\mathbf{a}}_i(\boldsymbol{\beta}) \\
&= \sum_{i=1}^n \Delta_i \sum_{j=1}^n \left\{ \mathbf{X}_j - \frac{\mathbf{S}^{(1)}(\boldsymbol{\beta}, T_i)}{S^{(0)}(\boldsymbol{\beta}, T_i)} \right\} \frac{Y_j(T_i) e^{\boldsymbol{\beta}^T \mathbf{X}_j}}{S_w^{(0)}(\boldsymbol{\beta}, T_i)} \\
&\quad - \sum_{i=1}^n \Delta_i \sum_{j=1}^n w_j R_j \left\{ \mathbf{X}_j - \frac{\mathbf{S}^{(1)}(\boldsymbol{\beta}, T_i)}{S^{(0)}(\boldsymbol{\beta}, T_i)} \right\} \frac{Y_j(T_i) e^{\boldsymbol{\beta}^T \mathbf{X}_j}}{S_w^{(0)}(\boldsymbol{\beta}, T_i)} \\
&= \sum_{i=1}^n \Delta_i \left\{ \frac{\mathbf{S}^{(1)}(\boldsymbol{\beta}, T_i)}{S_w^{(0)}(\boldsymbol{\beta}, T_i)} - \frac{\mathbf{S}^{(1)}(\boldsymbol{\beta}, T_i)}{S^{(0)}(\boldsymbol{\beta}, T_i)} \frac{S^{(0)}(\boldsymbol{\beta}, T_i)}{S_w^{(0)}(\boldsymbol{\beta}, T_i)} \right\} \\
&\quad - \sum_{i=1}^n \Delta_i \left\{ \frac{\mathbf{S}_w^{(1)}(\boldsymbol{\beta}, T_i)}{S_w^{(0)}(\boldsymbol{\beta}, T_i)} - \frac{\mathbf{S}^{(1)}(\boldsymbol{\beta}, T_i)}{S^{(0)}(\boldsymbol{\beta}, T_i)} \frac{S_w^{(0)}(\boldsymbol{\beta}, T_i)}{S_w^{(0)}(\boldsymbol{\beta}, T_i)} \right\} \\
&= \sum_{i=1}^n \Delta_i \left\{ \frac{\mathbf{S}^{(1)}(\boldsymbol{\beta}, T_i)}{S^{(0)}(\boldsymbol{\beta}, T_i)} - \frac{\mathbf{S}_w^{(1)}(\boldsymbol{\beta}, T_i)}{S_w^{(0)}(\boldsymbol{\beta}, T_i)} \right\} \\
&= \frac{\partial l^*(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}^T} - \frac{\partial l(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}^T} .
\end{aligned}$$

So, based on Eq's.(16)–(17) we can write

$$\sqrt{q}(\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta}^o) = -\tilde{\boldsymbol{\mathcal{I}}}^{-1}(\boldsymbol{\beta}^o) \frac{\sqrt{q}}{n} \frac{\partial l(\boldsymbol{\beta}^o)}{\partial \boldsymbol{\beta}^T} + \frac{\sqrt{q}}{n} \sum_{i \in \mathcal{C}} (1 - w_i R_i) \breve{\mathbf{a}}_i(\boldsymbol{\beta}^o) + o_P(\sqrt{q} \|\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta}^o\|_2) \tag{18}$$

Let us examine the first addend. From standard Cox regression results, assuming A.1–A.7 hold, $\{-\boldsymbol{\mathcal{I}}(\boldsymbol{\beta}^o)\}^{-1/2} n^{-1/2} \partial l(\boldsymbol{\beta}^o)/\partial \boldsymbol{\beta}^T \xrightarrow{D} N(0, I)$. Based on Lemma 3 and the dominated convergence theorem, one can show that $\tilde{I}(\boldsymbol{\beta}^o)$ is consistent to $\boldsymbol{\Sigma}(\boldsymbol{\beta}^o)$, see A.5, similarly to how Lemma 4 was proved, and therefore

$$\left\{-\tilde{\boldsymbol{\mathcal{I}}}(\boldsymbol{\beta}^o)\right\}^{-1/2} \frac{1}{\sqrt{n}} \frac{\partial l(\boldsymbol{\beta}^o)}{\partial \boldsymbol{\beta}^T} \xrightarrow{D} N(0, I)\,. \tag{19}$$

When this term is further multiplied by a factor of $q^{1/2} n^{-1/2} \left\{-\tilde{\boldsymbol{\mathcal{I}}}(\boldsymbol{\beta}^o)\right\}^{-1/2}$, as in Eq.(18), then due to Slutsky's theorem, the first addend vanishes to 0. Our goal now is to show that the second addend is asymptotically normal. It should be emphasized that since all observed failure times are included with a weight valued 1, the weights are not predictable, and standard martingale theory does not apply here, so instead we pursue a different direction. First, one can write

$$\frac{\sqrt{q}}{n} \sum_{i \in \mathcal{C}} (1 - w_i R_i) \breve{\mathbf{a}}_i(\boldsymbol{\beta}) = \frac{\sqrt{q}}{n} \int_0^\tau \frac{\mathbf{S}^{(1)}(\boldsymbol{\beta}, t) S_w^{(0)}(\boldsymbol{\beta}, t) - \mathbf{S}_w^{(1)}(\boldsymbol{\beta}, t) S^{(0)}(\boldsymbol{\beta}, t)}{S^{(0)}(\boldsymbol{\beta}, t) S_w^{(0)}(\boldsymbol{\beta}, t)} dN_\cdot(t)\,, \tag{20}$$

and

$$\frac{\sqrt{q}}{n} \sum_{i \in \mathcal{C}} (1 - w_i R_i) \mathbf{a}_i(\boldsymbol{\beta}) = \frac{\sqrt{q}}{n} \int_0^\tau \frac{\mathbf{S}^{(1)}(\boldsymbol{\beta}, t) S_w^{(0)}(\boldsymbol{\beta}, t) - \mathbf{S}_w^{(1)}(\boldsymbol{\beta}, t) S^{(0)}(\boldsymbol{\beta}, t)}{\{S^{(0)}(\boldsymbol{\beta}, t)\}^2} dN_\cdot(t)\,. \tag{21}$$

By means of the functional delta method, conditionally on $\mathcal{D}_n$, it can be shown that Eq.(20) is asymptotically normally distributed, with a zero-mean and covariance matrix identical to that of Eq.(21). For the sake of simpler presentation and derivation, we show the asymptotic normality of Eq.(21) instead and identify its covariance matrix, and it will follow that Eq.(20) and Eq.(21) are asymptotically equivalent. From the conditional space, we proceed to the unconditional space and establish the corresponding results there. Importantly, $\mathbf{a}_i(\boldsymbol{\beta}^o)$ is constant upon conditioning on $\mathcal{D}_n$, so

$$\frac{\sqrt{q}}{n} \sum_{i \in \mathcal{C}} w_i R_i \mathbf{a}_i(\boldsymbol{\beta})$$

can be alternatively expressed as a sum of q iid observations in the conditional space, marked with a "*" sign, as follows,

$$\frac{\sqrt{q}}{n} \sum_{i=1}^{q} w_i^* \mathbf{a}_i^*(\boldsymbol{\beta}^o) = \frac{1}{\sqrt{q}} \sum_{i=1}^{q} \frac{\mathbf{a}_i^*(\boldsymbol{\beta}^o)}{n p_i^*} \equiv \frac{1}{\sqrt{q}} \sum_{i=1}^{q} \boldsymbol{\gamma}_i(\mathbf{p}, \boldsymbol{\beta}^o)\,.$$

Since the distribution of $\boldsymbol{\gamma}_i(\mathbf{p},\boldsymbol{\beta}^o)$ changes as n and q increase, the Lindeberg-Feller (van der Vaart, 1998, proposition 2.27) condition should be established, as it covers the settings of triangular arrays. First, denote $\boldsymbol{\varphi}(\mathbf{p},\boldsymbol{\beta}) \equiv Var(\boldsymbol{\gamma}(\mathbf{p},\boldsymbol{\beta})|\mathcal{D}_n)$, and let us show that $\boldsymbol{\varphi}(\mathbf{p},\boldsymbol{\beta}^o)$ is finite.

$$\begin{aligned}\boldsymbol{\varphi}(\mathbf{p},\boldsymbol{\beta}^o) &= E\left(\boldsymbol{\gamma}(\mathbf{p},\boldsymbol{\beta}^o)\boldsymbol{\gamma}^T(\mathbf{p},\boldsymbol{\beta}^o)|\mathcal{D}_n\right) - E\left(\boldsymbol{\gamma}(\mathbf{p},\boldsymbol{\beta}^o)|\mathcal{D}_n\right)E\left(\boldsymbol{\gamma}(\mathbf{p},\boldsymbol{\beta}^o)|\mathcal{D}_n\right)^T \\ &= \frac{1}{n^2}\left\{\sum_{i\in\mathcal{C}}\frac{\mathbf{a}_i(\boldsymbol{\beta}^o)\mathbf{a}_i(\boldsymbol{\beta}^o)^T}{p_i} - \sum_{i,j\in\mathcal{C}}\mathbf{a}_i(\boldsymbol{\beta}^o)\mathbf{a}_j(\boldsymbol{\beta}^o)^T\right\} = O_{|\mathcal{D}_n}(1)\,,\end{aligned}$$

where the last equality is due to A.2–A.3, A.7–A.8. It is interesting to note, that the sum of $\mathbf{a}_i(\hat{\boldsymbol{\beta}}_{PL})$ over all observations is equal to $\mathbf{0}$, but in our case the sum is not $\mathbf{0}$ because only the censored observations appear in the summation, and because the $\mathbf{a}_i$'s are evaluated at the true parameter value $\boldsymbol{\beta}^{\mathbf{o}}$. Now, for every $\epsilon > 0$, and for some $\delta > 0$,

$$\begin{aligned}&\sum_{j=1}^{q} E\left\{\|q^{-1/2}\boldsymbol{\gamma}_i(\mathbf{p},\boldsymbol{\beta}^o)\|_2^2 I\left(\|q^{-1/2}\boldsymbol{\gamma}_i(\mathbf{p},\boldsymbol{\beta}^o)\|_2 > \epsilon\right)|\mathcal{D}_n\right\} \\ &\quad\le \frac{1}{q^{1+\delta/2}\epsilon^{\delta}}\sum_{i=1}^{q} E\left\{\left\|\frac{\mathbf{a}_i^*(\boldsymbol{\beta}^o)}{np_i^*}\right\|_2^{2+\delta}\Bigg|\mathcal{D}_n\right\} \\ &\quad= \frac{1}{q^{\delta/2}\epsilon^{\delta}n^{2+\delta}}\sum_{i\in\mathcal{C}}\frac{\|\mathbf{a}_i(\boldsymbol{\beta}^o)\|_2^{2+\delta}}{p_i^{1+\delta}} = o_{|\mathcal{D}_n}(1)\,,\end{aligned}$$

where the first inequality is due to van der Vaart (1998, pg.21). By that, since $E(1-w_iR_i|\mathcal{D}_n)=0$, it holds that $n^{-1}\sqrt{q}\boldsymbol{\varphi}(\mathbf{p},\boldsymbol{\beta}^o)^{-1/2}\sum_{i\in\mathcal{C}}(1-w_iR_i)\check{\mathbf{a}}_i(\boldsymbol{\beta}^o)$ converges conditionally on $\mathcal{D}_n$ to a standard multivariate normal distribution. Put differently, for all $\mathbf{u}\in\mathbb{R}^r$,

$$P\left\{n^{-1}\sqrt{q}\boldsymbol{\varphi}(\mathbf{p},\boldsymbol{\beta}^o)^{-1/2}\sum_{i\in\mathcal{C}}(1-w_iR_i)\check{\mathbf{a}}_i(\boldsymbol{\beta}^o)\le\mathbf{u}\Big|\mathcal{D}_n\right\}\xrightarrow{P}\Phi(\mathbf{u})\,,\tag{22}$$

where Φ is the standard multivariate normal cumulative distribution function. Since the conditional probability is a bounded random variable in the unconditional space, which converges to a constant, then due to Eq.(22) and the dominated convergence theorem, it follows that

$$P\left\{n^{-1}\sqrt{q}\boldsymbol{\varphi}(\mathbf{p},\boldsymbol{\beta}^o)^{-1/2}\sum_{i\in\mathcal{C}}(1-w_iR_i)\check{\mathbf{a}}_i(\boldsymbol{\beta}^o)\le\mathbf{u}\right\}\xrightarrow{P}\Phi(\mathbf{u})\,.\tag{23}$$

It should also be observed, that since the first addend in Eq.(18) goes to 0 as $q\to\infty$ and $n\to\infty$, based on A.10, the two addends are asymptotically independent.

Combining Eq.'s (18),(19),(23), we arrive at Eq.(12). Finally, the first expression within $\mathbb{V}_{\tilde{\boldsymbol{\beta}}}(\mathbf{p}, \boldsymbol{\beta}^o)$, $qn^{-1}\boldsymbol{\mathcal{I}}^{-1}(\boldsymbol{\beta}^o)$ goes to 0 as $q \to \infty$ and $n \to \infty$, however we prefer to keep it as it provides a more accurate expression for finite samples. The two variance components correspond to two orthogonal sources of variance, namely the variance generated by the data itself, and the variance generated by the subsampling procedure. □

Based on A.1–A.10 and Eq.(12), the same steps as Fleming and Harrington (2011, pg. 300), can be taken in order to prove Eq.(13), with some suitable adjustments for q and n. As with $\mathbb{V}_{\tilde{\boldsymbol{\beta}}}(\mathbf{p}, \boldsymbol{\beta}^o)$, the first addend in $\mathbb{V}_{\hat{\Lambda}_0(t,\tilde{\boldsymbol{\beta}})}(\mathbf{p}, \boldsymbol{\beta}^o, t)$ vanishes as $q \to \infty$ and $n \to \infty$, but we prefer to keep it for the sake of better approximation for finite samples.

2.3 Optimal Sampling Probabilities

There is more than one way to define "optimal", and one should choose a criterion to base optimality upon. In the field of optimal design of experiments, some well-known criteria are the A, D and E optimal designs (Pukelsheim, 2006), standing for "Average", "Determinant" and "Eigen", respectively, which correspond to the minimization of the trace/determinant/maximal eigenvalue of the estimated regression coefficients covariance matrix. We choose in this paper to work with the A-optimal design criterion due to its analytical convenience and because it is equivalent to minimizing the asymptotical MSE of the estimated coefficients. Namely, the sampling probability vector which minimizes $Tr(\mathbb{V}_{\tilde{\boldsymbol{\beta}}}(\mathbf{p}, \boldsymbol{\beta}^o))$ is derived, where Tr is the trace operator.

Theorem 3. *The A-optimal sampling probabilities vector* $\mathbf{p}^A$ *is of the form*

$$p_m^A = \frac{\|\boldsymbol{\mathcal{I}}^{-1}(\boldsymbol{\beta}^o)a_m(\boldsymbol{\beta}^o)\|_2}{\sum_{i\in\mathcal{C}} \|\boldsymbol{\mathcal{I}}^{-1}(\boldsymbol{\beta}^o)a_i(\boldsymbol{\beta}^o)\|_2} \text{ for all } m \in \mathcal{C}\,. \tag{24}$$

Proof of Theorem 3.

$$Tr\left(\mathbb{V}_{\tilde{\boldsymbol{\beta}}}(\mathbf{p}, \boldsymbol{\beta}^o)\right) = Tr\left\{\boldsymbol{\mathcal{I}}^{-1}(\boldsymbol{\beta}^o)\boldsymbol{\varphi}(\mathbf{p}, \boldsymbol{\beta}^o)\boldsymbol{\mathcal{I}}^{-1}(\boldsymbol{\beta}^o)\right\} + d\,,$$

where d is a constant not involving the vector $\mathbf{p}$. Proceeding with the first term,

$$\begin{aligned} Tr\left\{\boldsymbol{\mathcal{I}}^{-1}(\boldsymbol{\beta}^o)\boldsymbol{\varphi}(\mathbf{p}, \boldsymbol{\beta}^o)\boldsymbol{\mathcal{I}}^{-1}(\boldsymbol{\beta}^o)\right\} &= Tr\left[\frac{1}{n^2}\boldsymbol{\mathcal{I}}^{-1}(\boldsymbol{\beta}^o)\left\{\sum_{i\in\mathcal{C}}\frac{1}{p_i}\mathbf{a}_i(\boldsymbol{\beta}^o)\mathbf{a}_i^T(\boldsymbol{\beta}^o)\right.\right. \\ &\quad \left.\left. - \sum_{i,j\in\mathcal{C}}\mathbf{a}_i(\boldsymbol{\beta}^o)\mathbf{a}_j^T(\boldsymbol{\beta}^o)\right\}\boldsymbol{\mathcal{I}}^{-1}(\boldsymbol{\beta}^o)\right]. \end{aligned}$$

By omitting the part which does not involve $\mathbf{p}$,

$$Tr\left\{\frac{1}{n^2}\sum_{i\in\mathcal{C}}\frac{1}{p_i}\boldsymbol{\mathcal{I}}^{-1}(\boldsymbol{\beta}^o)\mathbf{a}_i(\boldsymbol{\beta}^o)\mathbf{a}_i^T(\boldsymbol{\beta}^o)\boldsymbol{\mathcal{I}}^{-1}(\boldsymbol{\beta}^o)\right\} = \frac{1}{n^2}\sum_{i\in\mathcal{C}}\frac{1}{p_i}Tr\left\{\mathbf{a}_i^T(\boldsymbol{\beta}^o)\boldsymbol{\mathcal{I}}^{-2}(\boldsymbol{\beta}^o)\mathbf{a}_i(\boldsymbol{\beta}^o)\right\}$$
$$= \frac{1}{n^2}\sum_{i\in\mathcal{C}}\frac{1}{p_i}\|\boldsymbol{\mathcal{I}}^{-1}(\boldsymbol{\beta}^o)\mathbf{a}_i(\boldsymbol{\beta}^o)\|_2^2\,.$$

Removing the factor of n^{-2} does not alter the optimization solution, so let us now define the Lagrangian function, with multiplier γ, as

$$g(\mathbf{p}) = \sum_{i\in\mathcal{C}}\frac{1}{p_i}\|\boldsymbol{\mathcal{I}}^{-1}(\boldsymbol{\beta}^o)\mathbf{a}_i(\boldsymbol{\beta}^o)\|_2^2 + \gamma(1-\sum_{i\in\mathcal{C}}p_i)\,,$$

Differentiating with respect to $p_m, m\in\mathcal{C}$ and setting the derivative to 0 we get

$$\frac{\partial g(\mathbf{p})}{\partial p_m} = -\frac{\|\boldsymbol{\mathcal{I}}^{-1}(\boldsymbol{\beta}^o)\mathbf{a}_m(\boldsymbol{\beta}^o)\|_2^2}{p_m^2} - \gamma \equiv 0\,,$$

hence

$$p_m = \frac{\|\boldsymbol{\mathcal{I}}^{-1}(\boldsymbol{\beta}^o)\mathbf{a}_m(\boldsymbol{\beta}^o)\|_2}{\sqrt{-\gamma}}\,.$$

Since all probabilities sum up to 1, it follows that

$$\sqrt{-\gamma} = \sum_{i\in\mathcal{C}}\|\boldsymbol{\mathcal{I}}^{-1}(\boldsymbol{\beta}^o)\mathbf{a}_i(\boldsymbol{\beta}^o)\|_2\,,$$

which yields Eq.(24). □

These expressions are not readily usable, as they require the unknown vector $\boldsymbol{\beta}^o$. Interestingly, had we plugged the partial-likelihood estimator $\hat{\boldsymbol{\beta}}_{PL}$ in place of $\boldsymbol{\beta}^o$, it would turn out that these optimal probabilities are in fact proportional to the norm of the commonly used approximation for each observation's "Dfbeta" statistic. The Dfbeta is defined as $\hat{\boldsymbol{\beta}}_{PL} - \hat{\boldsymbol{\beta}}_{PL(i)}$, where $\hat{\boldsymbol{\beta}}_{PL(i)}$ is the vector of estimated coefficients obtained by removing observation i from the sample (Klein and Moeschberger, 2006, pg. 391). The Dfbeta's are commonly used to detect influential observations, hence providing insight into which observations are assigned a higher sampling probability with this method. However, $\hat{\boldsymbol{\beta}}_{PL}$ is also unknown, as it entails running Cox regression on the full sample, which is what we are trying to circumvent, and in practice a two-step estimator, in the spirit of Wang et al. (2018) will be used, as discussed in section 2.4.

We now turn to discuss a different, yet related, optimality criterion, motivated by the Loewner ordering (Pukelsheim, 2006), which enjoys a computational advantage. The Loewner ordering states that for two positive definite matrices $\mathbf{M_1}$ and

$\mathbf{M_2}, \mathbf{M_1} \leq \mathbf{M_2}$ if and only if $\mathbf{M_2} - \mathbf{M_1}$ is a nonnegative definite matrix. The part in $\mathbb{V}_{\tilde{\boldsymbol{\beta}}}(\mathbf{p}, \boldsymbol{\beta}^o)$ which involves $\mathbf{p}$ is composed of 3 matrices $\boldsymbol{\mathcal{I}}^{-1}(\boldsymbol{\beta}^o)\boldsymbol{\varphi}(\mathbf{p}, \boldsymbol{\beta}^o)\boldsymbol{\mathcal{I}}^{-1}(\boldsymbol{\beta}^o)$ where $\boldsymbol{\varphi}$ is the only one among them which actually contains $\mathbf{p}$. It follows that for two potential sampling probability vectors $\mathbf{p}^{(1)}, \mathbf{p}^{(2)}$, $\mathbb{V}_{\tilde{\boldsymbol{\beta}}}(\mathbf{p}^{(1)}, \boldsymbol{\beta}^o) \leq \mathbb{V}_{\tilde{\boldsymbol{\beta}}}(\mathbf{p}^{(2)}, \boldsymbol{\beta}^o)$ if and only if $\boldsymbol{\varphi}(\mathbf{p}^{(1)}, \boldsymbol{\beta}^o) \leq \boldsymbol{\varphi}(\mathbf{p}^{(2)}, \boldsymbol{\beta}^o)$. This idea suggests that one can try to minimize $Tr\left(\boldsymbol{\varphi}(\mathbf{p}, \boldsymbol{\beta}^o)\right)$ instead, which is equivalent to the L-optimal criterion (Pukelsheim, 2006), standing for "Linear" and aims at the minimization of the trace of the covariance matrix of a linearly transformed vector of estimated regression coefficients. In our case, minimizing the trace of $\boldsymbol{\varphi}(\mathbf{p}, \boldsymbol{\beta}^o)$ is equivalent to minimizing the asymptotical mean squared error (MSE) of $\boldsymbol{\mathcal{I}}(\boldsymbol{\beta}^o)\tilde{\boldsymbol{\beta}}$.

Theorem 4. *The L-optimal sampling probabilities vector* $\mathbf{p}^L$ *is of the form*

$$p_m^L = \frac{\|\mathbf{a}_m(\boldsymbol{\beta}^o)\|_2}{\sum_{i \in \mathcal{C}} \|\mathbf{a}_i(\boldsymbol{\beta}^o)\|_2} \text{ for all } m \in \mathcal{C}\,. \tag{25}$$

The proof of Theorem 4 is straightforward and similar to the steps of Theorem 3, and thus is omitted. Again, using $\hat{\boldsymbol{\beta}}_{PL}$ instead of $\boldsymbol{\beta}^o$, one finds that each observation's sampling probability is in fact proportional to the norm of its vector of score residuals (Therneau et al., 1990; Klein and Moeschberger, 2006, pg 385 Eq. 11.6.1). The score residual is another useful quantity for the identification of influential observations, and is analogous to leverage scores in linear regression. Therefore, the "L" in L-optimality can be loaded in our case with a second interpretation, namely "Leverage-optimality". Roughly speaking, large values for $\|\mathbf{a}_i(\boldsymbol{\beta}^o)\|_2$ usually involve some of the following situations:

- Extreme covariate values.
- The observation is at risk until a relatively late time.
- The linear predictor $\boldsymbol{\beta}^{oT}\mathbf{X}_i$ is relatively large.

Since $\boldsymbol{\beta}^o$ is unknown, we employ the two-step estimator to be discussed in the next section.

The A-optimal probabilities take into account the variances and correlation between the estimated coefficients, through the information matrix $\boldsymbol{\mathcal{I}}$, in contrast to the L-optimal criterion which treats each covariate on its own and is agnostic to their variances and correlation. We should thus expect that the A-optimality criterion outperform L-optimality under unequal variances and when covariates are strongly correlated. In the numerical study in Section 3 we show that the efficiency-loss differences between L-opt and A-opt tend to be relatively small.

Finally, observations that were censored prior to the first observed failure time, have a sampling probability of 0, according to both optimal criteria. It is efficient

since these observations do not contribute any information to the partial-likelihood estimating equations.

2.4 Two-Step Procedure

The results obtained in Eqs.(24),(25) are impractical as they contain the unknown $\boldsymbol{\beta}^o$. Instead, a two-step procedure is implemented, which can be described as follows.

> **Step 1:** Sample q observations uniformly from $\mathcal{C}$, join them with the observed failure times to form $\mathcal{Q}_{pilot}$. Run a weighted Cox regression on $\mathcal{Q}_{pilot}$ to obtain a crude estimator $\tilde{\boldsymbol{\beta}}_U$, and use it to derive approximated optimal sampling probabilities using Eq.(24) or Eq.(25).
>
> **Step 2:** Sample another q observations from $\mathcal{C}$ using the probabilities computed at Step 1. Join these observations with the observed failure times to form $\mathcal{Q}$, and rerun weighted Cox regression on $\mathcal{Q}$ to obtain the two-step estimator $\tilde{\boldsymbol{\beta}}_{TS}$.

Theorem 5 is the analog of Theorem 2, and establishes asymptotic properties of $\tilde{\boldsymbol{\beta}}_{TS}$ and $\hat{\Lambda}_0(t, \tilde{\boldsymbol{\beta}}_{TS})$.

Theorem 5. *Under A.1–A.10, the following holds,*

$$\sqrt{q}\mathbb{V}_{\tilde{\boldsymbol{\beta}}}(\mathbf{p}^{opt}, \boldsymbol{\beta}^o)^{-1/2}(\tilde{\boldsymbol{\beta}}_{TS} - \boldsymbol{\beta}^o) \xrightarrow{D} N(0, I)\,, \tag{26}$$

$$\sqrt{q}\mathbb{V}_{\hat{\Lambda}_0(t,\tilde{\boldsymbol{\beta}})}(\mathbf{p}^{opt}, \boldsymbol{\beta}^o, t)^{-1/2}\{\hat{\Lambda}_0(t, \tilde{\boldsymbol{\beta}}_{TS}) - \Lambda_0^o(t)\} \xrightarrow{D} N(0, 1) \tag{27}$$

where $\mathbf{p}^{opt}$ is either $\mathbf{p}^L$ or $\mathbf{p}^A$, depending on the chosen optimality criterion.

Proof of Theorem 5. There are two key points essential for proving Eq.'s (26)–(27). The first, is that in the conditional space, if $\tilde{\boldsymbol{\beta}}_U$ is also conditioned upon, the sampling probabilities become deterministic, and we return to the settings of Theorem 2. The second, is that the consistency and normality results which were derived for $\tilde{\boldsymbol{\beta}}$, apply for any vector of deterministic sampling probabilities that satisfies A.8. Now, for all $\mathbf{u} \in \mathbb{R}^r$,

$$\begin{aligned}
&\Pr\left\{\mathbb{V}_{\tilde{\boldsymbol{\beta}}}^{-1/2}(\mathbf{p}^{opt}, \boldsymbol{\beta}^o)\left(\sqrt{q}(\tilde{\boldsymbol{\beta}}_{TS} - \boldsymbol{\beta}^o)\right) \le \mathbf{u}\right\} \\
&= E\left[P\left\{\mathbb{V}_{\tilde{\boldsymbol{\beta}}}^{-1/2}(\mathbf{p}^{opt}, \boldsymbol{\beta}^o)\left(\sqrt{q}(\tilde{\boldsymbol{\beta}}_{TS} - \boldsymbol{\beta}^o)\right) \le \mathbf{u} \middle| \mathcal{D}_n, \tilde{\boldsymbol{\beta}}_U\right\}\right]\,.
\end{aligned}$$

Based on the points mentioned above and Theorem 2, it follows that

$$\Pr\left\{\mathbb{V}_{\tilde{\boldsymbol{\beta}}}^{-1/2}(\mathbf{p}^{opt}, \boldsymbol{\beta}^o)\left(\sqrt{q}(\tilde{\boldsymbol{\beta}}_{TS} - \boldsymbol{\beta}^o)\right) \le \mathbf{u} \middle| \mathcal{D}_n, \tilde{\boldsymbol{\beta}}_U\right\} \to \Phi(\mathbf{u})\,,$$

then, because the conditional probability is a bounded random variable in the unconditional space, which converges to a constant, Eq.(26) is implied by the dominated convergence theorem. The same arguments hold for proving Eq.(27), and we skip the proof.

□

Lastly, let us briefly address the computational complexity involved in computing $\tilde{\boldsymbol{\beta}}_{TS}$. Since we are including all observed failure times in the analysis, and since it is assumed that the number of observed failure times constitutes a fixed (small) proportion of the overall sample size, we cannot use big-O notation in order to express the expected reduction in computation time. In the asymptotic big-O tools, which are commonly used to measure algorithm complexity, it holds that $O(\alpha n) = O(n)$, for any fixed α, no matter how small. Therefore, whatever the big-O complexity of the original algorithm is, by taking a fixed proportion of the sample, the resultant big-O complexity is at least of the same order. However, we demonstrate by simulations and real data analysis that in practice our method results in a substantial computation time reduction.

2.5 Variance Estimation

Based on Eq.(26), a natural estimator for the covariance matrix of $\tilde{\boldsymbol{\beta}}_{TS}$ is

$$q^{-1}\mathbb{V}_{\tilde{\boldsymbol{\beta}}}(\mathbf{p}^{opt}, \tilde{\boldsymbol{\beta}}_{TS}) = n^{-1}\boldsymbol{\mathcal{I}}^{-1}(\tilde{\boldsymbol{\beta}}_{TS}) + q^{-1}\boldsymbol{\mathcal{I}}^{-1}(\tilde{\boldsymbol{\beta}}_{TS})\boldsymbol{\varphi}(\mathbf{p}^{opt}, \tilde{\boldsymbol{\beta}}_{TS})\boldsymbol{\mathcal{I}}^{-1}(\tilde{\boldsymbol{\beta}}_{TS})\,.$$

However, calculation of $\boldsymbol{\mathcal{I}}^{-1}(\tilde{\boldsymbol{\beta}}_{TS})$ and $\boldsymbol{\varphi}(\mathbf{p}^{opt}, \tilde{\boldsymbol{\beta}}_{TS})$ involves the full data, and may be avoided by replacing these matrices with their subsampling-based counterparts, $\tilde{\boldsymbol{\mathcal{I}}}^{-1}(\tilde{\boldsymbol{\beta}}_{TS})$ and $\tilde{\boldsymbol{\varphi}}(\mathbf{p}^{opt}, \tilde{\boldsymbol{\beta}}_{TS})$, where

$$\tilde{\boldsymbol{\varphi}}(\mathbf{p}^{opt}, \tilde{\boldsymbol{\beta}}_{TS}) = \frac{1}{n^2}\left\{\frac{1}{q}\sum_{i=1}^{q}\frac{\tilde{\mathbf{a}}_i(\tilde{\boldsymbol{\beta}}_{TS})\tilde{\mathbf{a}}_i(\tilde{\boldsymbol{\beta}}_{TS})^T}{p_i^{*2}} - \frac{1}{q^2}\sum_{i=1}^{q}\frac{\tilde{\mathbf{a}}_i(\tilde{\boldsymbol{\beta}}_{TS})}{p_i^{*}}\left(\sum_{i=1}^{q}\frac{\tilde{\mathbf{a}}_i(\tilde{\boldsymbol{\beta}}_{TS})}{p_i^{*}}\right)^T\right\},$$

where

$$\tilde{\mathbf{a}}_i(\boldsymbol{\beta}) = \int_0^{\tau}\left\{\mathbf{X}_i - \frac{\mathbf{S}_w^{(1)}(\boldsymbol{\beta}, t)}{S_w^{(0)}(\boldsymbol{\beta}, t)}\right\}\frac{Y_i(t)e^{\boldsymbol{\beta}^T\mathbf{X}_i}}{S_w^{(0)}(\boldsymbol{\beta}, t)}dN_{\cdot}(t)\,.$$

The variance estimator for $\hat{\Lambda}_0(t, \tilde{\boldsymbol{\beta}}_{TS})$ is simply $q^{-1}\mathbb{V}_{\hat{\Lambda}_0(t,\tilde{\boldsymbol{\beta}})}(\mathbf{p}^{opt}, \tilde{\boldsymbol{\beta}}_{TS}, t)$. In Section 3 we demonstrate that our proposed variance estimators indeed approximate very well the covariance matrices.

2.6 Cox Model Refinements

Our method can be quite easily extended to more complicated analyses, illustrating its potential utility.

2.6.1 Delayed Entry

For right censored and delayed entry data, the standard riskset correction approach is implemented. Namely, suppose observation i, $i = 1 \ldots, n$, has a delayed entry time $L_i \geq 0$, then the modified at-risk process is given by $Y_i(t) = I(L_i \leq t \leq T_i)$. Observations whose entry and censoring times both occurred between two successive observed failure times, contribute no information to the partial likelihood. The value of $\mathbf{a}_i(\boldsymbol{\beta})$ for these observations is 0 and therefore they are assigned a sampling probability of 0 according to both L-opt and A-opt. Its implication is that our methods remove these observations from the dataset, and do not allow their inclusion in the subsample. Under these modifications and A.6, all of our previous results hold.

2.6.2 Time-Dependent Covariates

For using the R `coxph` function with time-dependent covariates, the data should be organized in "pseudo observations" form, as guided by Therneau et al. (2017). Namely, each observation may be broken into several distinct "pseudo-observations", such that each pseudo-observation has its entrance and exit times, creating non-overlapping intervals, that together can be put up to reconstruct the original time interval. Each pseudo-observation, within its interval, has fixed covariate values, and by treating them as if they were ordinary independent observations, the resultant partial-likelihood is the same as that of the original data. This elegant trick, although providing a convenient tool for analyzing data with time-dependent covariates, also suffers from a computational downside by inflating the number of "observations" inserted to the optimization routine. When the number of observations is large to begin with, the number of "pseudo-observations" can be daunting.

We propose two approaches to use our subsampling method in order to cope with the computation challenge, with the first approach being the one recommended by us.

Approach 1:

This approach simply amounts to sampling from the censored pseudo-observations. In this way, we in fact only use pieces of information from the original censored observations, however the most informative pieces are selected. Since regarding the pseudo-observations as independent results in the original full-data partial-likelihood, our derived optimality results extend to sampling from the pseudo-observations. Similarly to the discussion in the delayed entry subsection, it may be the case now that pseudo-observations are assigned sampling probability of 0. Its implication is that some parts of an observation's trajectory in the sample, may contribute no information to the estimation procedure.

Approach 2:

When splitting the data into pseudo-observations, each one of them is assigned an ID number, such that psuedo-observations originating from the same source, have the same ID. In case one finds themselves uncomfortable with the idea of using fractions of observations, another possibility is to compute $\mathbf{a}(\boldsymbol{\beta})$ for all pseudo-observations, and then add them by observation ID, thus reassembling all $\mathbf{a}_i(\boldsymbol{\beta})$'s of the original censored observations. This alternative has three drawbacks, the first is that sampling an observation results in sampling a block of pseudo-observations, and it is hard to know in advance what the size of the subsampled data will be, as observations vary in the number of their corresponding pseudo-observations. The second, is that for the same computational costs, more informative pseudo-observations could have been selected. The third, is that some additional computational costs are involved in reconstructing the original $\mathbf{a}_i(\boldsymbol{\beta})$, incurred by the ID matching.
Since we are seeking an efficient approximation to the full-data partial-likelihood estimator, we prefer the first option, as it provides a better approximation, even if the idea of using fractions of observations is conceptually new.

2.6.3 Stratified Analysis

In a stratified analysis, each observation's riskset consists of its respective stratum. Therefore, in order to compute sampling probabilities, we suggest deriving all censored observations' $\mathbf{a}_i(\boldsymbol{\beta})$'s in each stratum separately. If the A-opt option is used, the information matrices should be derived for each strata, and then summed, to get the information matrix of the full data. Having both all $\mathbf{a}_i(\boldsymbol{\beta})$'s and the information matrix, the L-opt and A-opt sampling probabilities can be worked out, and the two-step subsampling can be done.

2.6.4 Time-Dependent Coefficients

In a time-dependent coefficient analysis, it is customary to specify some function to model the time-dependent part of a covariate's effect. For instance, it may be assumed that the coefficient of a given covariate is $\beta_1 + \beta_2 t$ or $\beta_1 + \beta_2 \ln t$, so the effect has both a time-dependent and a time-independent part. The `coxph` function supports such an analysis using a user defined function, as explained in Therneau et al. (2017), and during the execution process, the data is transformed into time-dependent form. Specifically, each observation is divided into time intervals between all successive observed failure times to which it was at risk. This transformation heavily inflates the number of pseudo-observations, such that even an innocent looking dataset of 10,000 observations with about 1,000 observed failure times, is transformed into over 4 million pseudo-observations. Our method can be applied here to pick only a small and informative subset of pseudo-observations

to reduce the computational burden.

3 Numerical Results

The simulation study as well as usecase 2 in the UKB analysis, to be described below, were performed on a Dell XPS 15 9500, with an Intel processor i7-10750H CPU. The analysis of usecase 1 in the UKB analysis was performed on an AWS instance r5.8xlarge.

3.1 Simulation Study

In order to evaluate the performance of our proposed methods, compared to the full-data partial-likelihood estimator, and to the NCC and classic CC (uniform subsampling from censored), we carried out a simulation study. 300 samples were generated, each of size $n = 15,000$, using a vector of coefficients $\boldsymbol{\beta}^o$ of size $r = 6$, $\boldsymbol{\beta}^o = (3, -5, 1, -1, 1, -3)^T$. For the time-independent covariates simulations, the censoring times were generated from an exponential distribution with rate 0.2, independently of failure times. The instantaneous baseline hazard rates were set to be

$$\lambda_0(t) = 0.001I(t < 6) + c_{\lambda_0} I(t \geq 6)\,,$$

where c_{λ_0} was different for each setting "A", "B" or "C", as described below. The distribution of the covariates $\mathbf{X}$ also differed between all three settings, as follows.

A. $X_i \sim Unif(0, 4)$, for all $i = 1, \ldots, 6$, and are independent. $c_{\lambda_0} = 0.075$. This is the simple setting of equal variances and no correlation between the covariates.

B. $X_i \sim Unif(0, u_i)$, $u_i = 1, 6, 2, 2, 1, 6$, for $i = 1, \ldots, 6$, respectively, and are independent. This is the setting of unequal variances, but no correlation between the covariates. $c_{\lambda_0} = 0.15$.

C. X_1, X_2, X_3 are independently sampled from $U(0, 4)$. $X_4 = 0.5X_1 + 0.5X_2 + \varepsilon_1$. $X_5 = X_1 + \varepsilon_2$. $X_6 = X_1 + \varepsilon_3$, where $\varepsilon_1 \sim N(0, 0.1)$, $\varepsilon_2 \sim N(0, 1)$, $\varepsilon_3 \sim N(1, 1.5)$ and the ε's are independent. This is the setting of correlated covariates with (mildly) unequal variances. The strongest pairwise Pearson's correlation is about 0.75. $c_{\lambda_0} = 0.05$.

These settings were each replicated with and without delayed entry, and for different numbers of sampled censored observations. For the delayed entry settings, entry times were sampled from $Unif(0, 0.9T_i)$, independently for each observation.

For the time-dependent covariate simulations, we took all previous settings, and added a time-dependent covariate that mimics the number of CRC screening tests that individuals may experience throughout their life. The times between two consecutive tests were sampled from a $Unif(3, 12)$ distribution, and each observation may undergo up to 4 tests. At each time t, the time-dependent covariate holds the cumulative number of tests the individual had up to that moment. The time-dependent covariate's coefficient was set to $\beta^o_{dep} = 0.25$, and in setting B, c_{λ_0} was changed to 0.05. Similarly to the failure times, the censoring times were sampled such that they are dependent on the underlying covariates, and are conditionally independent to the failure times, given those covariates. Specifically, a vector of coefficients $\beta^C = (0.15, -0.1, 0.15, -0.1, 0.15, -0.1)^T$, $\beta^C_{dep} = 0.2$, and $\lambda^C_0(t) = 0.2I(t < 6) + 0.15I(t \geq 6)$ were used.

The different methods were compared on the basis of three metrics. The root mean squared error (RMSE) for the vector of coefficients was derived for each method, averaged over the 300 samples. The RMSE was calculated both with respect to the real vector of coefficients $\boldsymbol{\beta}^o$ and with respect to the full-data partial-likelihood estimator, serving as the "gold standard" we would have obtained had we used the full data.The average runtime and the relative RMSE (RR) with respect to the full-data partial-likelihood estimator are also presented. The RR was defined as the ratio between the RMSE of each method to that of the full-data-based estimator.

For matching controls to cases in the NCC methods, we used the "ccwc" function in the Epi R package (Carstensen et al., 2020). For the estimation procedures of the NCC methods we used a stratified analysis with the `coxph` function for the classic NCC ("NCC-c"), and the R package MultipleNCC for Samuelsen's improved NCC ("NCC-S") (Samuelsen, 1997). We also present the average runtime of the different methods, measured in seconds.

With regards to RMSE and RR, Tables 1–3 show that the L-opt and A-opt methods tend to perform similarly, with some advantage to the A-opt under setting "C". However, this advantage seems to shrink as the number of sampled observations increase. Additionally, our methods outperform the NCC-c, NCC-S and the uniform method under all settings, except for setting "A" when three controls are sampled per observed event, where NCC-S performs similarly to L-opt. With regards to runtime, the L-opt is slightly faster than the A-opt method, and both are considerably faster than the full-data partial likelihood and the NCC methods, but as expected, are slower than the uniform method, which does not require computation of sampling probabilities. The conclusion is that L-opt and A-opt may offer a computational advantage over the full-data partial-likelihood, without sacrificing a lot of efficiency, in contrast the uniform sampling which may potentially incur a high efficiency loss.

The simulation results of the time-dependent covariates setting are summarized in Tables 4–6. It should be kept in mind that sampling pseudo-observations (Approach 1), denoted as "ps" in the tables, and Approach 2 ("L-opt", "A-opt") are not readily comparable. In Approach 1, q reflects the number of pseudo-observations, or rows in the dataset, which are included in the analysis, whereas in Approach 2 q reflects the number of observations, such that each observation may include several rows. Despite the fact that Approach 2 uses overall more pseudo-observations than Approach 1, the latter still has an advantage in setting C, in terms of RMSE and RR. That can be explained by the fact that Approach 1 selects the most informative pseudo-observations, while Approach 2 selects observations as one piece, which is not optimal. In order to grasp this notion more clearly, suppose that for some observation, the integral from time 0 to time t in the vector $\mathbf{a}(\boldsymbol{\beta})$ takes a negative value, while from time t to time τ it takes a positive value. Overall, $\mathbf{a}(\boldsymbol{\beta})$ may be small, due to the two parts canceling out each other, even though each part may be informative on its own, and could have been selected using Approach 2.

For demonstrating the finite sample performance of our proposed variance estimators, we estimated the covariance matrices of the coefficients, under setting "C", without delayed entry, and with three censored observations per observed failure time, for all 300 samples, and averaged them. Table 7 shows the empirical SD and average estimated SE of each estimated coefficient, as well as the average Frobenius distance between the empirical and estimated covariance matrices. For the cumulative hazard function variance estimator, we took a grid of time points, ranging from 0 to 10, with jumps of size 0.2, and calculated the empirical SD and estimated SE at these points. In Figure 1 the empirical SD at each time point is plotted against its estimated counterpart. Both Table 7 and Figure 1 show excellent agreement between the estimated and empirical variances.

3.2 UK Biobank Colorectal Cancer Data

The UKB is a large-scale health resource containing rich information of about 500,000 participants in the UK, recruited at ages 40–69. All participants provided consent for follow-up through linkage to their health-related records. The collected data include genetic, medical and environmental information measured in various ways, such as blood, urine and saliva samples as well as personal questionnaires. Out of 484,918 UKB participants with available genetic and environmental data, there is a total number of 2,792 incident CRC cases during the follow-up time. As a result of the ascertainment procedure, the UKB data are characterized by right censoring with delayed entry, and our analyses are conducted accordingly.

The following analyses include the well-established environmental CRC risk factors: body mass index (BMI), smoking status (yes/no), history of CRC in

family (yes/no), physical activity (yes/no), sex (2 levels: female/male) , alcohol consumption (3 levels: "non or occasional","light frequent drinker", "very frequent drinker"), education (5 levels: "prefer not to answer", "lower than highschool", "highschool", "higher vocational education", "collage or university graduate"), drug use (3 levels: "none", "Aspirin", "Ibuprofen"), post-menopausal hormones (yes/no); as well as 72 SNPs that have been linked to CRC by GWAS (Jeon et al., 2018). The SNPs were standardized to have mean zero and unit variance.

In the following sections, we describe two use cases of the CRC UKB data, for which a subsampling approach can be of great practical value.

3.2.1 Time-Dependent Coefficients

Our goal in this section is to build a prediction model for CRC, based on the UKB data, using Cox regression. The model includes all 72 identified SNPs, environmental risk factors, and the interactions between sex and physical activity, BMI, smoking status and alcohol consumption. The full data analysis takes 93.4 seconds, while the runtimes for the uniform, L-opt, A-opt, NCC-c and NCC-S, with 4 sampled censored observations per observed failure time, were 1.7, 4.4, 9.9, 41.6 and 1765.5 seconds, respectively. The results of these analyses are provided in Tables 8–9 and Figure 2. Table 8 demonstrates that the L-opt and A-opt methods perform substantially better than the other methods, considering both the runtime and the distance to the full-data partial-likelihood estimator, measured by the RMSE (for the coefficients) and the Forbenius norm (for the variance matrix). However, since 93.4 seconds is not too prohibitive a time to wait, the subsampling methods are not direly needed. It is now required to check that the proportional-hazards assumption indeed holds, and in case it is violated, the model can be refined using time-dependent coefficients. A useful function in the R Survival package for detection of departures from the proportional-hazards assumption is `cox.zph`. This function provides Schoenfeld-residuals-based graphical and statistical tests for evaluating the necessity of time-dependent coefficients.

The graph and statistical test produced for the interaction between sex and physical activity seem suspicious, as observed in Figure 3 and Table 10. The graphs and statistical tests for the rest of the covariates are provided in Figures 4–5 and Table 10. Based on these diagnostics, it is reasonable to fit a model such that the interaction has a time-dependent coefficient of the form $\beta_1 + \beta_2 \ln(t)$, where the ln function is chosen to shrink the time scale. For fitting such a model, one could use the "`tt`" argument in the `coxph` function, or equivalently, transform a dataset into time-dependent form by means of the `survSplit` function in the Survival package, with the sorted vector of distinct failure times as the cut points. More details about these mechanisms are provided in Therneau et al. (2017). The more distinct observed failure times there are, the bigger the time-

dependent dataset will become, so all observed failure times are rounded to their first decimal digit, producing a maximal rounding error of about 18 days, quite negligible for cancer research with long follow-up duration. The resultant UKB-based time-dependent dataset comprises about 33 million pseudo-observations, with a median of 71 pseudo-observations per one original observation. Including both time-dependent and independent coefficients for the sex and physical activity interaction term, induced severe multicollinearity (Pearson's correlation of 0.9993), so the time-independent coefficient was removed from the model. Considering the long runtime required by the NCC methods, we only considered the subsampling approaches in this analysis with $q = 15, 30, 71$, whose results are provided in Tables 11 – 14. One can see that the L-opt method offers a good balance of runtime and accuracy. It substantially reduces the runtime from the full-data partial-likelihood's 25.7 minutes to approximately 6 minutes, when $q = 71$, while maintaining an advantage over the uniform subsampling in terms of the RMSE and Frobenius norm metrics. When $q = 15, 30$ the L-opt's Frobenius norm metric is about one third of the uniform's, and when $q = 71$, it is about one half, and therefore we advocate using L-opt. As to the propotional-hazards assumption, upon examining the time-dependent coefficient under consideration, it turns out to be non-significant under all methods.

3.2.2 SNP Marginal Analysis

Another important application for the subsampling approach, is the case of SNP marginal analysis. As mentioned, for example, in Yang et al. (2012), SNPs are usually tested for associations with a trait based on a one-SNP-at-a-time model, while controlling for other clinical and environmental covariates (Bush and Moore, 2012). Large electronic health records (EHR) data and biobanks promote discoveries of new genetic variants, thus advancing precision medicine research (Denny et al., 2018; Hughey et al., 2019). The Cox regression model was found to increase statistical power in the detection of phenotype-associated SNPs, over the traditionally used logistic regression model (Van Der Net et al., 2008; Staley et al., 2017; Hughey et al., 2019), hence it would be beneficial to alleviate its computational burden while losing minimal efficiency. The number of SNPs to be tested for an association with a trait of interest can range from thousands to millions (Abed and Belzile, 2019), implying that a similar number of models need be fitted. In the UKB, there is a total number of about 96 million testable variants (Bycroft et al., 2018), and about 5 millions of them are expected to be common SNPs. Hughey et al. (2019) ran a Cox-regression-based GWAS analysis with 795,850 common SNPs, for 50 different phenotypes, and a sample size of almost 50,000. The overall runtime, pararellized on 36 cores, was 7.1 days. Additionally, the number of observed failure times for the different phenotypes ranged from 104 to 7972, hence

our methods are well suited for these phenotypes, as the rare event assumption seems to hold. We conducted a marginal analysis for the UKB data with the 72 SNPs, and the environmental risk factors described above. The results of this analysis are provided in Tables 15–16. Even though the analysis included environmental risk factors, Table 16 contains the estimates of the 72 SNPs. It should be understood that we performed the analysis only on 72 SNPs for the sake of demonstration. In practice, one would perform the analysis on millions of SNPs. The runtime required by the full-data partial likelihood was about 11 minutes, and only about 0.9, 1.2 and 0.2 minutes by the L-opt, A-opt and uniform, respectively. Therefore, our methods are adequate for analyzing millions of SNPs, while maintaining statistical efficiency.

4 Discussion and Future Research

We have proposed efficient and fast subsampling-based estimators for the proportional-hazards model, motivated by two optimality criteria, under the settings of rare events. These estimators help alleviating the computational burden involved in the process of analyzing massive datasets, without compromising on estimation efficiency. In this work all observed failure times were included in the analysis, while the censored observations were downsampled using optimal sampling probabilities. Under the rare events settings this approach is natural as the number of censored observations far surpasses the number of observed failure times. In a future work, one could extend these methods to the case of non-rare events, such that the observed failure times also need to be downsampled. This work also stands as a cornerstone for the development of subsampling methods for other Cox-regression-based methods, such as (semi-)competing risks, frailty models, correlated data, recurrent events, multistate models and more. Existing methods for analyzing these more intricate models are usually more computationally intensive, and therefore our work is an important first step towards reducing their computational burden.

We would like to emphasize that the `coxph` function in R is very comprehensive, and includes many appealing features and abilities. Our methods allow users to keep using the `coxph` function, thus enabling access to all of its features. If one wanted to try and reduce the computational burden by using other optimization routines, such as quasi-Newton or gradient descent, the programming will have to be carried out from scratch, and usage of `coxph` will be blocked.

Another research direction worth exploring, is whether Cox regression can be adopted as a working model for computing sampling probabilities for other computationally intensive models, where tailor-made analytic derivation may be intractable. In particular, it would make sense to consider partial-likelihood and log-rank-based random forests (Ishwaran et al., 2008), gradient boosting (Binder

et al., 2009), among others. Moreover, our method can be evaluated against other NCC and CC methods, on the premise of missing covariates. Namely, when some covariates are too expensive to measure, one can apply our method on all fully observed data and obtain a subset of observations to undergo measuring. It should be interesting to examine under what circumstances our method brings an improvement upon the classically used NCC and CC.

Lastly, it is assumed in this work that the researches determine q according to their computational resources. Another interesting research question is quantifying the efficiency loss in terms of q.

Acknowledgments

The author MG gratefully acknowledges support from the U.S.-Israel Binational Science Foundation (BSF, 2016126), and the Israel Science Foundation (ISF, 1067/17) in carrying out this work. We would like to thank Asaf Ben-Arie for assistance with the programming involved in this work and Li Hsu for fruitful discussions. This research has been conducted using the UK Biobank Resource, under project number 56885.

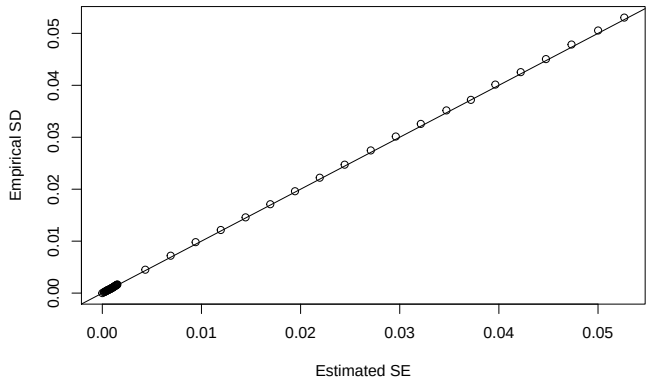

Figure 1: Simulation results of right-censored data without delayed entry, under setting "C", for the A-opt method, with three censored observations per observed failure time: estimated SE and empirical SD for the cumulative hazard function at various time points.

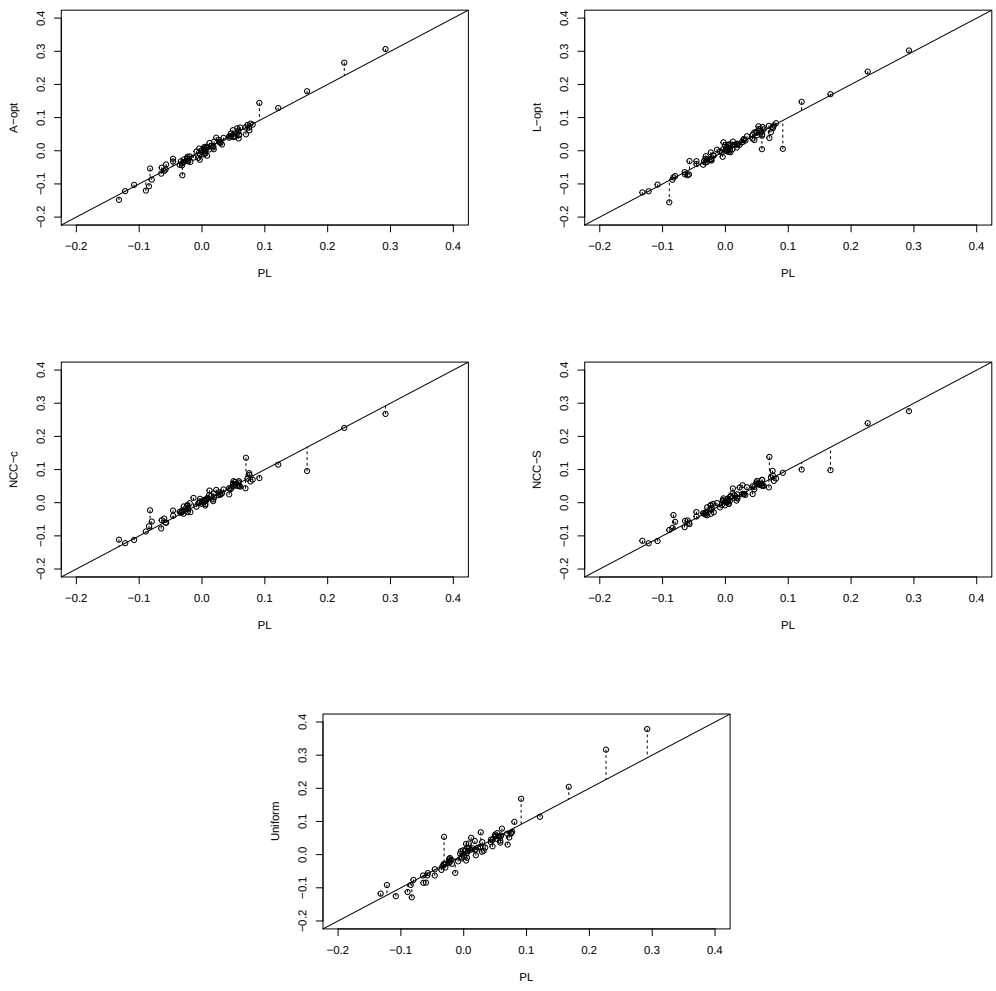

Figure 2: UKB CRC data analysis: comparison between the estimated regression coefficients of each method (L-opt, A-opt, NCC-c, NCC-S, uniform) against the partial-likelihood-based estimator (PL).

Table 1: Simulation results of right-censored data without and with delayed entry: the value of q was set equal to the number of failures, hence the mean, $\bar{q}$, and standard deviation of q, SD(q), are reported.

Setting	$\bar{q}$	SD(q)	Method	RMSE $\boldsymbol{\beta}^o$	RR	RMSE $\hat{\boldsymbol{\beta}}_{PL}$	Runtime (sec.)
No delayed entry							
A	792	27.9	PL	0.0724	1.0000	0.0000	0.0530
			L-opt	0.0921	1.2717	0.0571	0.0180
			A-opt	0.0936	1.2928	0.0588	0.0186
			Uniform	0.1914	2.6446	0.1742	0.0068
			NCC-c	0.1221	1.6872	0.0949	0.3143
			NCC-S	0.1054	1.4558	0.0732	0.5017
B	495	23.2	PL	0.2358	1.0000	0.0000	0.0534
			L-opt	0.3130	1.3276	0.2117	0.0149
			A-opt	0.3059	1.2975	0.1909	0.0153
			Uniform	0.7278	3.0867	0.7004	0.0062
			NCC-c	0.4053	1.7189	0.3354	0.1948
			NCC-S	0.3718	1.5767	0.2894	0.3278
C	458	21.4	PL	0.4967	1.0000	0.0000	0.0559
			L-opt	0.6621	1.3332	0.4774	0.0172
			A-opt	0.6300	1.2686	0.3650	0.0191
			Uniform	1.4687	2.9572	1.3659	0.0080
			NCC-c	0.7977	1.6061	0.6154	0.1890
			NCC-S	0.7488	1.5076	0.5585	0.3206
With delayed entry							
A	792	28.5	PL	0.0778	1.0000	0.0000	0.0719
			L-opt	0.0952	1.2245	0.0576	0.0252
			A-opt	0.0962	1.2367	0.0567	0.0271
			Uniform	0.1846	2.3738	0.1709	0.0088
			NCC-c	0.1195	1.5368	0.0913	0.3096
			NCC-S	0.1063	1.3671	0.0728	2.2043
B	494	21.6	PL	0.2387	1.0000	0.0000	0.0708
			L-opt	0.3129	1.3109	0.2029	0.0188
			A-opt	0.3061	1.2821	0.1917	0.0200
			Uniform	0.7010	2.9367	0.6552	0.0075
			NCC-c	0.3949	1.6541	0.3228	0.1930
			NCC-S	0.3667	1.5360	0.2832	1.7583
C	454	21.9	PL	0.4394	1.0000	0.0000	0.0833
			L-opt	0.6694	1.5236	0.4645	0.0405
			A-opt	0.5540	1.2610	0.3846	0.0274
			Uniform	1.4151	3.2208	1.3546	0.0078
			NCC-c	0.7215	1.6421	0.5956	0.2007
			NCC-S	0.6642	1.5118	0.5247	2.0368

Table 2: Simulation results of right-censored data without and with delayed entry: the value of q was set equal to the number of failures$\times 2$, hence the mean, $\bar{q}$, and standard deviation of q, SD(q), are reported.

Setting	$\bar{q}$	SD(q)	Method	RMSE $\boldsymbol{\beta}^o$	RR	RMSE $\hat{\boldsymbol{\beta}}_{PL}$	Runtime (sec.)
No delayed entry							
A	1592	56.7	PL	0.0737	1.0000	0.0000	0.0539
			L-opt	0.0833	1.1305	0.0411	0.0221
			A-opt	0.0845	1.1470	0.0405	0.0243
			Uniform	0.1482	2.0113	0.1260	0.0098
			NCC-c	0.1010	1.3708	0.0681	0.3196
			NCC-S	0.0871	1.1825	0.0488	0.5080
B	987	44.8	PL	0.2338	1.0000	0.0000	0.0557
			L-opt	0.2725	1.1653	0.1454	0.0181
			A-opt	0.2732	1.1685	0.1336	0.0185
			Uniform	0.5401	2.3101	0.4785	0.0075
			NCC-c	0.3423	1.4641	0.2388	0.1985
			NCC-S	0.3017	1.2904	0.1862	0.3281
C	913	43.5	PL	0.4844	1.0000	0.0000	0.0536
			L-opt	0.5971	1.2329	0.3151	0.0188
			A-opt	0.5399	1.1146	0.2552	0.0187
			Uniform	1.0050	2.0750	0.8994	0.0077
			NCC-c	0.6847	1.4136	0.4474	0.1838
			NCC-S	0.6364	1.3139	0.3673	0.3087
With delayed entry							
A	1582	54.5	PL	0.0775	1.0000	0.0000	0.0701
			L-opt	0.0883	1.1392	0.0415	0.0306
			A-opt	0.0879	1.1339	0.0393	0.0310
			Uniform	0.1362	1.7573	0.1169	0.0127
			NCC-c	0.0997	1.2864	0.0648	0.3139
			NCC-S	0.0901	1.1620	0.0470	2.2400
B	991	41.4	PL	0.2315	1.0000	0.0000	0.0767
			L-opt	0.2726	1.1776	0.1446	0.0352
			A-opt	0.2689	1.1616	0.1349	0.0384
			Uniform	0.5299	2.2891	0.4599	0.0110
			NCC-c	0.3419	1.4768	0.2370	0.2187
			NCC-S	0.3057	1.3207	0.1852	2.2119
C	910	43.6	PL	0.5139	1.0000	0.0000	0.0762
			L-opt	0.5679	1.1050	0.3049	0.0244
			A-opt	0.5492	1.0687	0.2487	0.0256
			Uniform	1.1116	2.1630	0.9301	0.0097
			NCC-c	0.6850	1.3329	0.4192	0.1849
			NCC-S	0.6583	1.2809	0.3510	1.8030

Table 3: Simulation results of right-censored data without and with delayed entry: the value of q was set equal to the number of failures$\times$3, hence the mean, $\bar{q}$, and standard deviation of q, SD(q), are reported.

Setting	$\bar{q}$	SD(q)	Method	RMSE $\boldsymbol{\beta}^o$	RR	RMSE $\hat{\boldsymbol{\beta}}_{PL}$	Runtime (sec.)
No delayed entry							
A	2378	87.9	PL	0.0733	1.0000	0.0000	0.0537
			L-opt	0.0808	1.1024	0.0331	0.0292
			A-opt	0.0803	1.0952	0.0328	0.0302
			Uniform	0.1257	1.7141	0.1012	0.0133
			NCC-c	0.0897	1.2232	0.0536	0.3218
			NCC-S	0.0804	1.0960	0.0345	0.5142
B	1493	63.7	PL	0.2347	1.0000	0.0000	0.0554
			L-opt	0.2564	1.0923	0.1211	0.0214
			A-opt	0.2596	1.1060	0.1099	0.0220
			Uniform	0.4602	1.9603	0.3847	0.0093
			NCC-c	0.3069	1.3074	0.1963	0.1988
			NCC-S	0.2766	1.1782	0.1456	0.3297
C	1370	62.5	PL	0.4950	1.0000	0.0000	0.0507
			L-opt	0.5595	1.1303	0.2568	0.0227
			A-opt	0.5378	1.0864	0.1976	0.0222
			Uniform	0.8779	1.7735	0.7350	0.0098
			NCC-c	0.6053	1.2229	0.3519	0.1849
			NCC-S	0.5915	1.1949	0.2886	0.3151
With delayed entry							
A	2380	78.2	PL	0.0794	1.0000	0.0000	0.0707
			L-opt	0.0864	1.0882	0.0320	0.0369
			A-opt	0.0867	1.0909	0.0328	0.0386
			Uniform	0.1258	1.5840	0.0984	0.0165
			NCC-c	0.0951	1.1973	0.0515	0.3193
			NCC-S	0.0877	1.1036	0.0351	2.2356
B	1479	63.3	PL	0.2273	1.0000	0.0000	0.0866
			L-opt	0.2537	1.1160	0.1186	0.0447
			A-opt	0.2518	1.1076	0.1055	0.0447
			Uniform	0.4504	1.9812	0.3885	0.0130
			NCC-c	0.3059	1.3458	0.1891	0.2591
			NCC-S	0.2758	1.2131	0.1489	2.5273
C	1361	61.4	PL	0.4789	1.0000	0.0000	0.0732
			L-opt	0.5578	1.1646	0.2648	0.0328
			A-opt	0.5209	1.0876	0.2025	0.0293
			Uniform	0.9243	1.9298	0.7661	0.0110
			NCC-c	0.6093	1.2723	0.3545	0.1877
			NCC-S	0.5887	1.2292	0.3066	1.7834

Table 4: Simulation results of right-censored data with time dependent covariates: the value of q was set equal to the number of failures, hence the mean, $\bar{q}$, and standard deviation of q, SD(q), are reported.

Setting	$\bar{q}$	SD(q)	Method	RMSE $\boldsymbol{\beta}^o$	RR	RMSE $\hat{\boldsymbol{\beta}}_{PL}$	Runtime (sec.)
A	462	22.8	PL	0.1153	1.0000	0.0000	0.1188
			L-opt	0.1485	1.2882	0.0964	0.0628
			A-opt	0.1479	1.2834	0.0928	0.0753
			Uniform	0.3484	3.0229	0.3304	0.0189
			L-opt-ps	0.1622	1.4075	0.1108	0.0325
			A-opt-ps	0.1605	1.3923	0.1084	0.0331
B	423	19	PL	0.2702	1.0000	0.0000	0.1890
			L-opt	0.3700	1.3693	0.2470	0.0793
			A-opt	0.3689	1.3654	0.2491	0.0909
			Uniform	0.7350	2.7201	0.6694	0.0225
			L-opt-ps	0.3847	1.4237	0.2736	0.0419
			A-opt-ps	0.3704	1.3708	0.2538	0.0439
C	330	18.6	PL	0.5537	1.0000	0.0000	0.1418
			L-opt	0.8747	1.5797	0.6297	0.0561
			A-opt	0.9146	1.6517	0.6266	0.0652
			Uniform	2.1622	3.9047	2.0481	0.0130
			L-opt-ps	0.9674	1.7470	0.6585	0.0317
			A-opt-ps	0.7293	1.3171	0.5409	0.0333

Table 5: Simulation results of right-censored data with time dependent covariates: the value of q was set equal to the number of failures$\times 2$, hence the mean, $\bar{q}$, and standard deviation of q, SD(q), are reported.

Setting	$\bar{q}$	SD(q)	Method	RMSE $\boldsymbol{\beta}^o$	RR	RMSE $\hat{\boldsymbol{\beta}}_{PL}$	Runtime (sec.)
A	916	44.9	PL	0.1066	1.0000	0.0000	0.1250
			L-opt	0.1261	1.1823	0.0634	0.0753
			A-opt	0.1231	1.1546	0.0652	0.0821
			Uniform	0.2558	2.3986	0.2268	0.0213
			L-opt-ps	0.1348	1.2640	0.0704	0.0421
			A-opt-ps	0.1277	1.1973	0.0739	0.0456
B	850	39.8	PL	0.2548	1.0000	0.0000	0.1960
			L-opt	0.3144	1.2342	0.1716	0.0923
			A-opt	0.2994	1.1751	0.1603	0.1068
			Uniform	0.5399	2.1194	0.5001	0.0294
			L-opt-ps	0.3107	1.2194	0.1884	0.0475
			A-opt-ps	0.3089	1.2127	0.1666	0.0520
C	660	36.5	PL	0.5862	1.0000	0.0000	0.1336
			L-opt	0.7091	1.2097	0.3865	0.0638
			A-opt	0.6638	1.1324	0.3826	0.0729
			Uniform	1.3374	2.2814	1.3229	0.0159
			L-opt-ps	0.6546	1.1167	0.3944	0.0330
			A-opt-ps	0.6578	1.1221	0.3603	0.0367

Table 6: Simulation results of right-censored data with time dependent covariates: the value of q was set equal to the number of failures$\times 3$, hence the mean, $\bar{q}$, and standard deviation of q, SD(q), are reported.

Setting	$\bar{q}$	SD(q)	Method	RMSE $\boldsymbol{\beta}^o$	RR	RMSE $\hat{\boldsymbol{\beta}}_{PL}$	Runtime (sec.)
A	1368	58.5	PL	0.1123	1.0000	0.0000	0.1353
			L-opt	0.1257	1.1195	0.0522	0.0906
			A-opt	0.1257	1.1196	0.0543	0.0965
			Uniform	0.2174	1.9364	0.1823	0.0247
			L-opt-ps	0.1282	1.1418	0.0601	0.0443
			A-opt-ps	0.1250	1.1133	0.0573	0.0467
B	1262	64.3	PL	0.2762	1.0000	0.0000	0.1972
			L-opt	0.3144	1.1384	0.1351	0.0983
			A-opt	0.3019	1.0930	0.1346	0.1160
			Uniform	0.4757	1.7222	0.3871	0.0289
			L-opt-ps	0.3233	1.1705	0.1504	0.0505
			A-opt-ps	0.3064	1.1094	0.1449	0.0551
C	986	52.2	PL	0.6121	1.0000	0.0000	0.1491
			L-opt	0.6765	1.1052	0.3057	0.0819
			A-opt	0.6940	1.1337	0.2996	0.0913
			Uniform	1.1109	1.8149	1.0383	0.0226
			L-opt-ps	0.6658	1.0877	0.3467	0.0381
			A-opt-ps	0.6560	1.0716	0.2974	0.0436

Table 7: Simulation results of right-censored data without delayed entry, under setting "C", with three censored observations per observed failure time: empirical SE (Emp.), estimated SEs based on the full sample (Full) and on the subsample (Sub) for each estimated regression coefficient. The Frobenius norm was used to calculate the distance between the empirical and estimated variance matrices.

	Uniform			L-opt			A-opt		
	Emp.	Full	Sub	Emp.	Full	Sub	Emp.	Full	Sub
β_1	0.4621	0.4622	0.4370	0.2874	0.2787	0.2794	0.2728	0.2666	0.2667
β_2	0.4463	0.4493	0.4243	0.2820	0.2720	0.2727	0.2604	0.2597	0.2598
β_3	0.0793	0.0768	0.0728	0.0457	0.0459	0.0460	0.0498	0.0488	0.0488
β_4	0.8901	0.8852	0.8359	0.5513	0.5345	0.5357	0.5144	0.5089	0.5091
β_5	0.0832	0.0886	0.0838	0.0548	0.0529	0.0530	0.0563	0.0556	0.0556
β_6	0.0598	0.0607	0.0560	0.0372	0.0354	0.0354	0.0377	0.0378	0.0379
$\|.\|_F$	0.0000	0.0112	0.1339	0.0000	0.0297	0.0277	0.0000	0.0108	0.0106

Table 8: UKB CRC data analysis: metrics for evaluating the different methods. The RMSE and Frobenius distance are with respect to the full data partial-likelihood estimator. 4 censored observations were sampled for each observed failure time.

	RMSE	$\|.\|_F$	Time (seconds)
PL	0.0000	0.0000	93.4000
L-opt	0.1871	0.0369	4.4200
A-opt	0.1738	0.0196	9.9000
Uniform	0.3547	0.0664	1.7000
NCC-c	0.2017	0.0374	41.6100
NCC-S	0.2137	0.0609	1765.5000

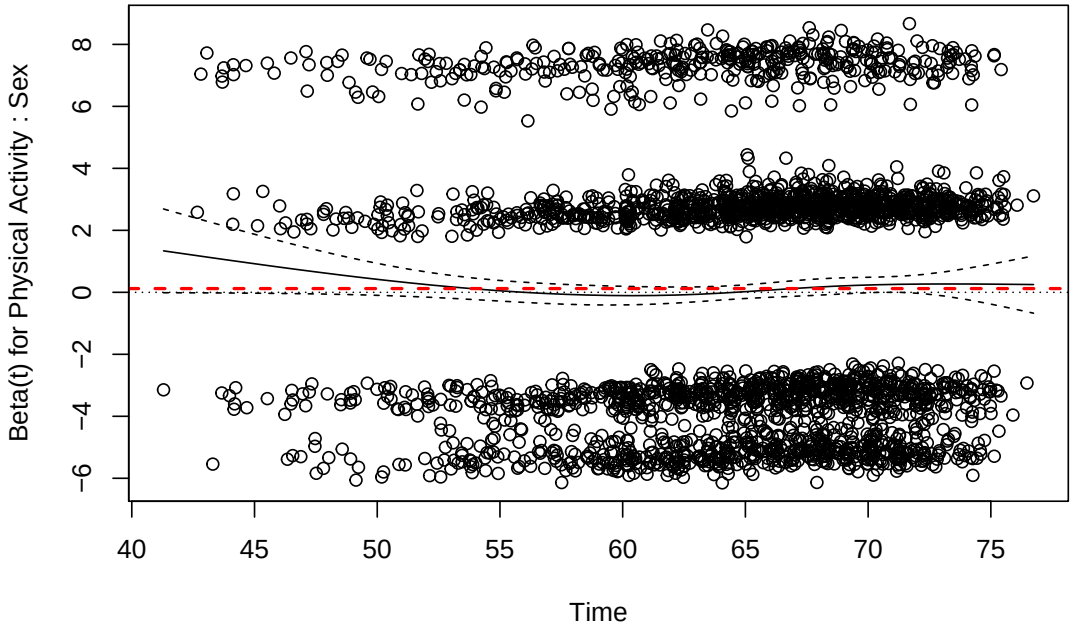

Figure 3: UKB CRC data analysis: Diagnostic graph for evaluating necessity of a time-dependent coefficient for the interaction term of physical activity and sex. The red dashed line shows the time-fixed coefficient for physical activity : sex.

Table 9: UKB CRC data analysis: estimated regression coefficients and standard errors of each method. 4 censored observations were sampled for each observed failure time.

	PL	PL se	L-opt	L se	A-opt	A se	Unif	Unif se	NCC-c	NCC-c se	NCC-S	NCC-S se
BMI	0.0005	0.0059	0.0000	0.0066	0.0024	0.0065	-0.0034	0.0068	0.0037	0.0064	0.0032	0.0064
Smoking	0.0700	0.0595	0.0387	0.0666	0.0496	0.0670	0.0303	0.0678	0.1355	0.0651	0.1380	0.0664
Family	0.2265	0.0635	0.2389	0.0718	0.2656	0.0715	0.3165	0.0803	0.2254	0.0735	0.2397	0.0765
Activity	-0.0800	0.0624	-0.0762	0.0697	-0.0869	0.0698	-0.0764	0.0717	-0.0568	0.0685	-0.0576	0.0703
Sex	-0.7535	0.2580	-0.8085	0.2943	-0.6431	0.2885	-1.0261	0.3085	-0.6301	0.2879	-0.6260	0.2886
Alcohol2	-0.0135	0.0743	0.0033	0.0836	-0.0207	0.0832	-0.0550	0.0856	0.0145	0.0819	-0.0009	0.0834
Alcohol3	-0.0827	0.0696	-0.0801	0.0782	-0.0535	0.0784	-0.1287	0.0803	-0.0230	0.0767	-0.0373	0.0784
Education1	-0.2622	0.1597	-0.2967	0.1819	-0.2699	0.1646	-0.2708	0.2016	-0.2255	0.1840	-0.2038	0.2008
Education2	-0.2515	0.1582	-0.3036	0.1802	-0.2580	0.1626	-0.2587	0.1996	-0.2317	0.1826	-0.2081	0.1991
Education3	-0.2551	0.1629	-0.3184	0.1856	-0.2825	0.1686	-0.2722	0.2056	-0.2279	0.1882	-0.2177	0.2040
Education4	-0.3020	0.1592	-0.3232	0.1815	-0.3319	0.1640	-0.2935	0.2008	-0.2903	0.1837	-0.2741	0.2001
Aspirin	0.0581	0.0497	0.0048	0.0568	0.0606	0.0570	0.0422	0.0606	0.0606	0.0558	0.0690	0.0583
Ibuprofen	-0.0893	0.0658	-0.1554	0.0743	-0.1201	0.0742	-0.1126	0.0766	-0.0868	0.0731	-0.0822	0.0751
Hormones	-0.0312	0.0596	-0.0163	0.0667	-0.0740	0.0668	0.0536	0.0681	-0.0232	0.0650	-0.0359	0.0672
BMI:Sex	0.0295	0.0082	0.0325	0.0093	0.0235	0.0092	0.0378	0.0098	0.0263	0.0092	0.0268	0.0091
Smoking:Sex	0.1672	0.0789	0.1706	0.0897	0.1795	0.0900	0.2045	0.0925	0.0956	0.0879	0.0982	0.0907
Sex:Alcohol2	0.0914	0.1072	0.0059	0.1221	0.1442	0.1211	0.1684	0.1266	0.0743	0.1194	0.0902	0.1230
Sex:Alcohol3	0.2921	0.0962	0.3024	0.1096	0.3066	0.1086	0.3786	0.1137	0.2677	0.1074	0.2761	0.1107
Activity:Sex	0.1214	0.0821	0.1476	0.0929	0.1289	0.0932	0.1140	0.0971	0.1148	0.0918	0.1001	0.0949
SNP1	-0.0601	0.0200	-0.0738	0.0229	-0.0605	0.0230	-0.0843	0.0234	-0.0477	0.0221	-0.0535	0.0226
SNP2	-0.0642	0.0191	-0.0713	0.0218	-0.0510	0.0219	-0.0849	0.0227	-0.0529	0.0213	-0.0546	0.0220
SNP3	0.0459	0.0187	0.0552	0.0214	0.0438	0.0215	0.0251	0.0226	0.0441	0.0210	0.0473	0.0217
SNP4	0.0039	0.0188	0.0167	0.0214	0.0116	0.0218	-0.0177	0.0225	0.0085	0.0214	0.0126	0.0218
SNP5	0.0502	0.0191	0.0542	0.0218	0.0412	0.0221	0.0607	0.0230	0.0503	0.0214	0.0567	0.0223
SNP6	0.0038	0.0190	0.0031	0.0216	0.0071	0.0216	0.0326	0.0226	0.0065	0.0211	0.0040	0.0222
SNP7	0.0433	0.0183	0.0485	0.0209	0.0393	0.0207	0.0452	0.0222	0.0250	0.0207	0.0260	0.0216
SNP8	-0.0231	0.0190	-0.0282	0.0218	-0.0177	0.0216	-0.0220	0.0228	-0.0193	0.0215	-0.0211	0.0222
SNP9	-0.0220	0.0191	-0.0290	0.0217	-0.0317	0.0218	-0.0101	0.0229	-0.0145	0.0210	-0.0169	0.0217
SNP10	0.0271	0.0100	0.0259	0.0104	0.0266	0.0108	0.0674	0.0189	0.0314	0.0143	0.0527	0.0210
SNP11	0.0503	0.0189	0.0467	0.0215	0.0433	0.0216	0.0554	0.0225	0.0648	0.0213	0.0656	0.0220
SNP12	-0.0008	0.0189	0.0168	0.0215	0.0001	0.0216	0.0132	0.0227	-0.0035	0.0211	-0.0081	0.0218
SNP13	-0.0088	0.0193	-0.0018	0.0220	-0.0021	0.0223	-0.0202	0.0230	-0.0121	0.0217	-0.0141	0.0225
SNP14	0.0119	0.0269	0.0278	0.0305	0.0233	0.0308	0.0508	0.0322	0.0366	0.0303	0.0431	0.0316
SNP15	0.0583	0.0269	0.0454	0.0306	0.0371	0.0308	0.0364	0.0322	0.0511	0.0303	0.0511	0.0314
SNP16	0.0806	0.0184	0.0832	0.0210	0.0787	0.0211	0.0989	0.0224	0.0696	0.0209	0.0731	0.0216
SNP17	0.0054	0.0187	-0.0039	0.0205	-0.0004	0.0207	0.0209	0.0226	-0.0040	0.0210	-0.0035	0.0217
SNP18	-0.0044	0.0208	-0.0182	0.0214	0.0067	0.0230	0.0106	0.0217	-0.0003	0.0230	0.0046	0.0249
SNP19	0.0317	0.0191	0.0309	0.0216	0.0188	0.0219	0.0120	0.0228	0.0279	0.0213	0.0237	0.0220
SNP20	-0.0290	0.0194	-0.0267	0.0220	-0.0265	0.0222	-0.0276	0.0230	-0.0112	0.0217	-0.0137	0.0221
SNP21	0.0024	0.0190	-0.0009	0.0216	-0.0112	0.0219	0.0119	0.0228	0.0022	0.0210	0.0004	0.0218
SNP22	0.0342	0.0191	0.0438	0.0217	0.0389	0.0219	0.0216	0.0227	0.0400	0.0211	0.0467	0.0217
SNP23	-0.0235	0.0191	-0.0256	0.0217	-0.0227	0.0218	-0.0180	0.0231	-0.0093	0.0214	-0.0087	0.0222
SNP24	0.0263	0.0190	0.0256	0.0218	0.0325	0.0216	0.0239	0.0226	0.0229	0.0215	0.0257	0.0222
SNP25	0.0081	0.0189	-0.0043	0.0214	-0.0148	0.0217	0.0335	0.0225	0.0136	0.0213	0.0183	0.0218
SNP26	-0.0185	0.0216	-0.0145	0.0246	-0.0338	0.0248	-0.0276	0.0256	-0.0284	0.0239	-0.0288	0.0247
SNP27	0.0692	0.0178	0.0752	0.0204	0.0704	0.0205	0.0632	0.0220	0.0435	0.0201	0.0462	0.0211
SNP28	-0.0035	0.0213	0.0048	0.0243	-0.0271	0.0245	-0.0115	0.0255	0.0039	0.0238	0.0077	0.0247
SNP29	-0.1322	0.0192	-0.1255	0.0219	-0.1478	0.0219	-0.1175	0.0228	-0.1115	0.0215	-0.1149	0.0222
SNP30	0.0226	0.0188	0.0181	0.0215	0.0395	0.0218	0.0216	0.0226	0.0386	0.0213	0.0459	0.0223
SNP31	-0.0843	0.0194	-0.0874	0.0220	-0.1072	0.0223	-0.0916	0.0231	-0.0710	0.0217	-0.0761	0.0224
SNP32	0.0048	0.0189	0.0089	0.0216	-0.0088	0.0216	0.0047	0.0225	0.0063	0.0213	0.0088	0.0218
SNP33	-0.0460	0.0195	-0.0401	0.0223	-0.0325	0.0223	-0.0441	0.0229	-0.0394	0.0217	-0.0417	0.0228
SNP34	0.0771	0.0191	0.0764	0.0219	0.0817	0.0220	0.0690	0.0228	0.0641	0.0214	0.0657	0.0219
SNP35	0.0492	0.0184	0.0567	0.0209	0.0620	0.0210	0.0553	0.0220	0.0573	0.0206	0.0611	0.0215
SNP36	0.0080	0.0189	0.0192	0.0213	0.0087	0.0217	0.0196	0.0225	0.0116	0.0214	0.0200	0.0222
SNP37	0.0526	0.0183	0.0738	0.0211	0.0426	0.0208	0.0478	0.0222	0.0627	0.0211	0.0580	0.0218
SNP38	0.0432	0.0185	0.0364	0.0211	0.0436	0.0211	0.0376	0.0221	0.0435	0.0211	0.0495	0.0222
SNP39	-0.0230	0.0191	-0.0246	0.0217	-0.0301	0.0219	-0.0136	0.0229	-0.0282	0.0214	-0.0334	0.0226
SNP40	0.0052	0.0189	0.0192	0.0216	0.0027	0.0216	-0.0093	0.0225	-0.0086	0.0211	-0.0025	0.0215
SNP41	0.0608	0.0190	0.0648	0.0216	0.0702	0.0218	0.0785	0.0226	0.0485	0.0213	0.0504	0.0224
SNP42	-0.0580	0.0186	-0.0713	0.0211	-0.0567	0.0213	-0.0627	0.0222	-0.0598	0.0211	-0.0604	0.0216
SNP43	-0.0292	0.0189	-0.0335	0.0216	-0.0340	0.0217	-0.0387	0.0224	-0.0334	0.0213	-0.0381	0.0223
SNP44	0.0144	0.0186	0.0178	0.0211	0.0146	0.0214	0.0189	0.0224	0.0203	0.0211	0.0229	0.0214
SNP45	0.0193	0.0195	0.0186	0.0221	0.0277	0.0222	-0.0019	0.0233	0.0164	0.0220	0.0127	0.0224
SNP46	-0.0233	0.0281	-0.0049	0.0331	-0.0251	0.0319	-0.0258	0.0320	-0.0073	0.0312	-0.0067	0.0304
SNP47	-0.0317	0.0192	-0.0240	0.0218	-0.0440	0.0221	-0.0284	0.0231	-0.0282	0.0216	-0.0315	0.0224
SNP48	0.0586	0.0192	0.0494	0.0218	0.0472	0.0219	0.0552	0.0230	0.0649	0.0215	0.0686	0.0224
SNP49	0.0187	0.0190	0.0213	0.0216	0.0045	0.0215	0.0160	0.0229	0.0277	0.0213	0.0254	0.0219
SNP50	0.0113	0.0186	0.0051	0.0211	0.0037	0.0212	0.0146	0.0221	0.0231	0.0208	0.0279	0.0221
SNP51	-0.0057	0.0189	-0.0052	0.0217	-0.0214	0.0216	0.0029	0.0226	-0.0020	0.0214	-0.0065	0.0222
SNP52	0.0534	0.0189	0.0657	0.0215	0.0518	0.0218	0.0648	0.0226	0.0540	0.0212	0.0546	0.0219
SNP53	-0.0354	0.0190	-0.0421	0.0216	-0.0424	0.0217	-0.0460	0.0227	-0.0273	0.0213	-0.0315	0.0222
SNP54	0.0724	0.0179	0.0559	0.0203	0.0781	0.0207	0.0516	0.0217	0.0727	0.0204	0.0763	0.0214
SNP55	0.0460	0.0190	0.0325	0.0216	0.0522	0.0219	0.0448	0.0227	0.0415	0.0213	0.0405	0.0219
SNP56	-0.0031	0.0190	0.0256	0.0219	-0.0099	0.0216	-0.0066	0.0226	0.0118	0.0213	0.0130	0.0223
SNP57	-0.0461	0.0192	-0.0314	0.0220	-0.0242	0.0222	-0.0629	0.0230	-0.0238	0.0216	-0.0280	0.0223
SNP58	-0.0337	0.0190	-0.0301	0.0216	-0.0319	0.0218	-0.0351	0.0227	-0.0308	0.0213	-0.0340	0.0223
SNP59	0.0180	0.0188	0.0099	0.0212	0.0138	0.0216	0.0144	0.0226	0.0099	0.0209	0.0151	0.0216
SNP60	0.0290	0.0179	0.0352	0.0204	0.0269	0.0207	0.0087	0.0219	0.0239	0.0203	0.0256	0.0212
SNP61	0.0071	0.0193	0.0011	0.0220	0.0219	0.0221	0.0113	0.0233	0.0077	0.0215	0.0054	0.0227
SNP62	-0.1080	0.0200	-0.1021	0.0226	-0.1027	0.0228	-0.1252	0.0240	-0.1122	0.0224	-0.1156	0.0231
SNP63	-0.0571	0.0209	-0.0310	0.0239	-0.0412	0.0239	-0.0558	0.0248	-0.0605	0.0233	-0.0644	0.0246
SNP64	0.0755	0.0194	0.0682	0.0221	0.0617	0.0225	0.0659	0.0231	0.0842	0.0218	0.0817	0.0227
SNP65	-0.0317	0.0370	-0.0355	0.0419	-0.0381	0.0392	-0.0302	0.0382	-0.0272	0.0397	-0.0287	0.0405
SNP66	0.0177	0.0189	0.0196	0.0213	0.0099	0.0215	0.0412	0.0227	0.0047	0.0213	0.0063	0.0219
SNP67	0.0560	0.0188	0.0551	0.0215	0.0678	0.0215	0.0557	0.0226	0.0505	0.0212	0.0532	0.0218
SNP68	0.0593	0.0186	0.0717	0.0213	0.0485	0.0215	0.0576	0.0224	0.0505	0.0210	0.0520	0.0215
SNP69	-0.0196	0.0190	-0.0139	0.0215	-0.0172	0.0218	-0.0156	0.0225	-0.0014	0.0213	-0.0039	0.0219
SNP70	-0.1222	0.0196	-0.1222	0.0223	-0.1216	0.0225	-0.0917	0.0232	-0.1220	0.0218	-0.1225	0.0229
SNP71	-0.0649	0.0191	-0.0645	0.0215	-0.0690	0.0219	-0.0630	0.0227	-0.0776	0.0215	-0.0737	0.0225
SNP72	0.0751	0.0196	0.0720	0.0226	0.0701	0.0225	0.0634	0.0233	0.0891	0.0220	0.0959	0.0226

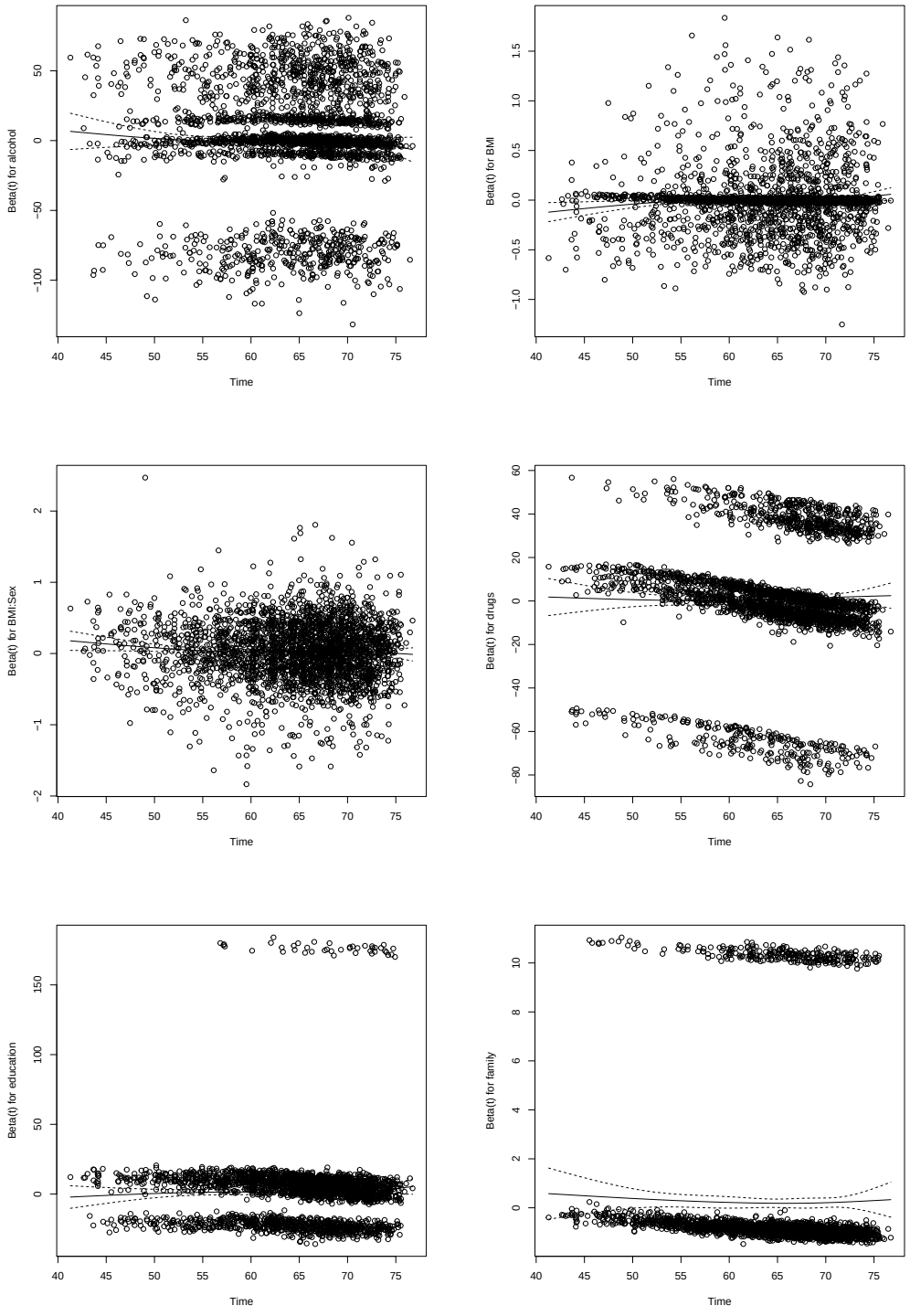

Figure 4: UKB CRC data analysis: Graphical tests for evaluating the necessity of time-dependent coefficients, as produced by the `cox.zph` function - part 1.

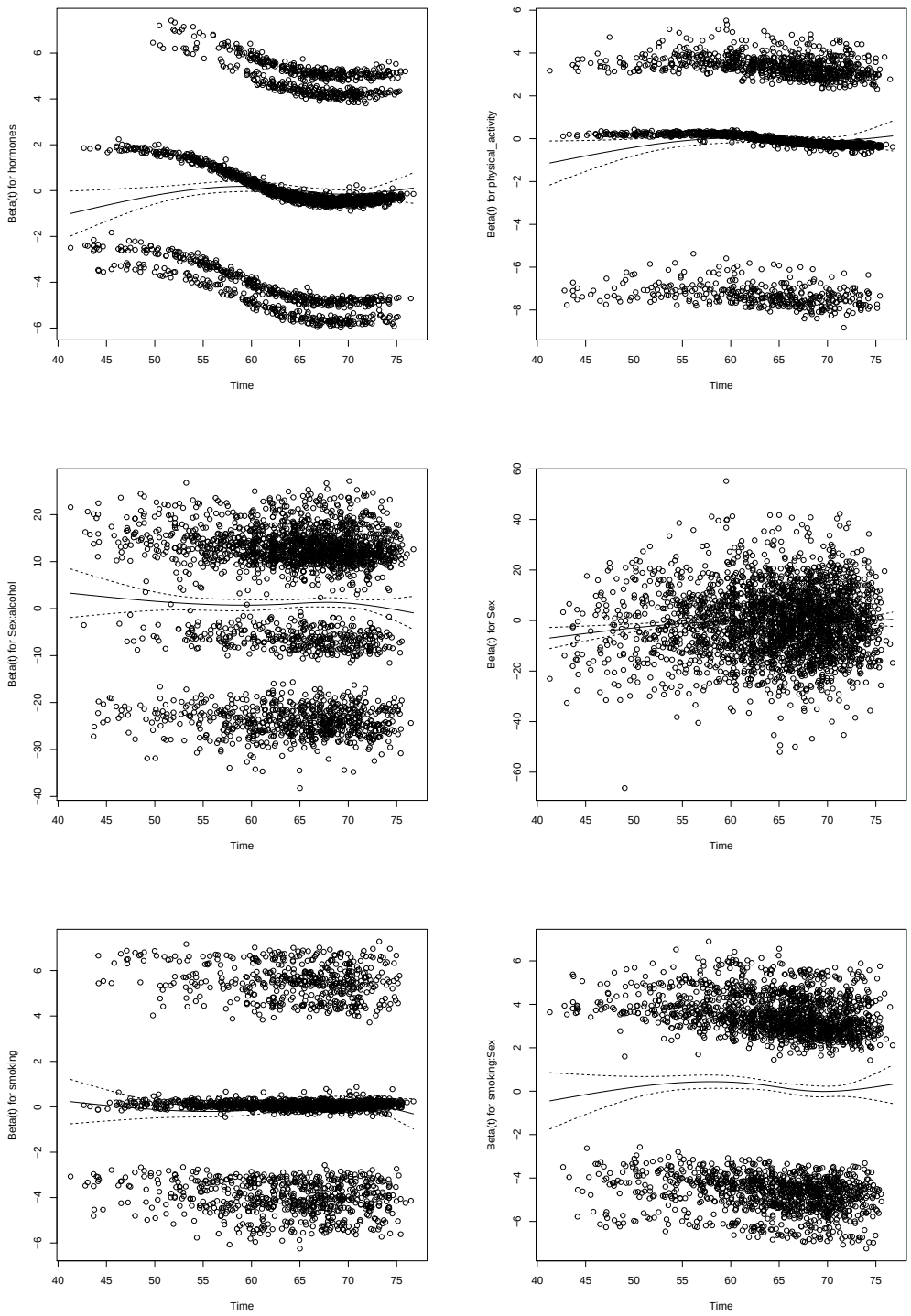

Figure 5: UKB CRC data analysis: Graphical tests for evaluating the necessity of time-dependent coefficients, as produced by the `cox.zph` function - part 2.

Table 10: UKB CRC data analysis: Statistical tests for evaluating the necessity of time-dependent coefficients, as produced by the `cox.zph` function.

	Chisq	df	P-value
BMI	2.5101	1.0000	0.1131
Smoking	3.6420	1.0000	0.0563
Family	0.3278	1.0000	0.5670
Activity	1.9921	1.0000	0.1581
Sex	6.6049	1.0000	0.0102
Alcohol	0.1426	2.0000	0.9312
Education	6.0541	4.0000	0.1951
Drugs	1.8821	2.0000	0.3902
Hormones	4.8073	1.0000	0.0283
BMI:Sex	6.1406	1.0000	0.0132
Smoking:Sex	3.6072	1.0000	0.0575
Sex:alcohol	2.4712	2.0000	0.2907
Activity:Sex	7.3212	1.0000	0.0068

Table 11: UKB CRC time-dependent coefficient data analysis: metrics for evaluating the different methods. 15, 30 and 71 censored observations were sampled for each observed failure time. The RMSE and Frobenius distance are with respect to the full data partial-likelihood estimator.

	RMSE	$\|.\|_F$	Runtime (seconds)
PL	0.0000	0.0000	1542.0000
15			
L-opt	0.1680	0.0087	342.5340
A-opt	0.1069	0.0052	890.0220
Uniform	0.2547	0.0220	164.9380
30			
L-opt	0.1067	0.0041	343.9890
A-opt	0.0867	0.0021	892.4160
Uniform	0.2529	0.0112	170.9510
71			
L-opt	0.0467	0.0022	358.7080
A-opt	0.0513	0.0011	906.9360
Uniform	0.0722	0.0045	194.4480

Table 12: UKB CRC time-dependent coefficient data analysis: estimated regression coefficients and standard errors of each method. 15 censored observations were sampled for each observed failure time.

	PL	PL se	L-opt	L-opt se	A-opt	A-opt se	Unif	Unif se
BMI	0.0005	0.0059	0.0005	0.0061	-0.0001	0.0061	0.0002	0.0062
Smoking	0.0699	0.0595	0.0728	0.0616	0.0548	0.0617	0.0675	0.0624
Family	0.2265	0.0635	0.2284	0.0659	0.2299	0.0659	0.2894	0.0689
Activity	-0.0863	0.0621	-0.0721	0.0644	-0.1038	0.0645	-0.0821	0.0652
Sex	-0.7631	0.2578	-0.8341	0.2672	-0.8083	0.2663	-0.8397	0.2771
Alcohol2	-0.0136	0.0743	-0.0159	0.0767	-0.0468	0.0770	-0.0234	0.0781
Alcohol3	-0.0827	0.0696	-0.0872	0.0719	-0.0953	0.0721	-0.0886	0.0730
Education1	-0.2621	0.1597	-0.1769	0.1651	-0.2682	0.1611	-0.3787	0.1745
Education2	-0.2513	0.1582	-0.1814	0.1637	-0.2393	0.1595	-0.3626	0.1726
Education3	-0.2551	0.1629	-0.1999	0.1685	-0.2255	0.1646	-0.3652	0.1776
Education4	-0.3019	0.1592	-0.2383	0.1647	-0.3050	0.1606	-0.4119	0.1737
Aspirin	0.0581	0.0497	0.0786	0.0516	0.0545	0.0519	0.0807	0.0535
Ibuprofen	-0.0895	0.0658	-0.1100	0.0683	-0.1000	0.0683	-0.0874	0.0693
Hormones	-0.0304	0.0597	-0.0472	0.0616	-0.0348	0.0619	-0.0329	0.0627
Sex:Activity:ln(time)	0.0318	0.0196	0.0297	0.0204	0.0389	0.0204	0.0375	0.0209
BMI:Sex	0.0296	0.0082	0.0312	0.0085	0.0289	0.0085	0.0317	0.0089
Smoking:Sex	0.1671	0.0789	0.1694	0.0819	0.1912	0.0823	0.1875	0.0836
Sex:Alcohol2	0.0914	0.1072	0.1148	0.1114	0.1516	0.1114	0.0754	0.1138
Sex:Alcohol3	0.2920	0.0962	0.2972	0.1000	0.3025	0.1000	0.2998	0.1019
SNP1	-0.0601	0.0200	-0.0734	0.0208	-0.0651	0.0208	-0.0546	0.0211
SNP2	-0.0642	0.0191	-0.0677	0.0198	-0.0561	0.0200	-0.0619	0.0204
SNP3	0.0459	0.0187	0.0460	0.0195	0.0454	0.0195	0.0509	0.0200
SNP4	0.0039	0.0188	-0.0032	0.0196	0.0035	0.0195	-0.0006	0.0201
SNP5	0.0502	0.0191	0.0527	0.0199	0.0462	0.0200	0.0480	0.0204
SNP6	0.0038	0.0190	-0.0019	0.0198	0.0003	0.0200	-0.0041	0.0202
SNP7	0.0433	0.0183	0.0374	0.0189	0.0388	0.0192	0.0422	0.0195
SNP8	-0.0231	0.0190	-0.0266	0.0197	-0.0207	0.0199	-0.0258	0.0203
SNP9	-0.0220	0.0191	-0.0195	0.0198	-0.0290	0.0200	-0.0169	0.0203
SNP10	0.0272	0.0100	0.0262	0.0100	0.0290	0.0105	0.0428	0.0149
SNP11	0.0503	0.0189	0.0486	0.0197	0.0528	0.0197	0.0555	0.0202
SNP12	-0.0008	0.0189	-0.0014	0.0196	-0.0043	0.0196	-0.0041	0.0200
SNP13	-0.0088	0.0193	-0.0010	0.0202	-0.0105	0.0202	-0.0087	0.0206
SNP14	0.0119	0.0269	0.0126	0.0281	0.0179	0.0282	0.0048	0.0287
SNP15	0.0583	0.0269	0.0607	0.0279	0.0490	0.0280	0.0705	0.0288
SNP16	0.0806	0.0184	0.0829	0.0193	0.0835	0.0193	0.0843	0.0198
SNP17	0.0054	0.0187	0.0120	0.0194	0.0129	0.0195	0.0094	0.0200
SNP18	-0.0044	0.0208	-0.0013	0.0210	-0.0059	0.0211	0.0092	0.0220
SNP19	0.0317	0.0191	0.0311	0.0199	0.0338	0.0200	0.0454	0.0204
SNP20	-0.0290	0.0194	-0.0326	0.0202	-0.0317	0.0203	-0.0161	0.0206
SNP21	0.0025	0.0190	0.0037	0.0198	0.0069	0.0199	-0.0021	0.0203
SNP22	0.0342	0.0191	0.0421	0.0198	0.0348	0.0199	0.0368	0.0203
SNP23	-0.0235	0.0191	-0.0275	0.0199	-0.0118	0.0199	-0.0291	0.0202
SNP24	0.0263	0.0190	0.0192	0.0197	0.0287	0.0199	0.0217	0.0203
SNP25	0.0081	0.0189	0.0100	0.0196	0.0098	0.0197	-0.0021	0.0202
SNP26	-0.0186	0.0216	-0.0141	0.0223	-0.0256	0.0225	-0.0262	0.0229
SNP27	0.0692	0.0178	0.0760	0.0185	0.0655	0.0186	0.0605	0.0191
SNP28	-0.0035	0.0213	-0.0010	0.0222	0.0001	0.0222	-0.0125	0.0228
SNP29	-0.1322	0.0192	-0.1386	0.0199	-0.1379	0.0200	-0.1327	0.0204
SNP30	0.0226	0.0188	0.0271	0.0196	0.0179	0.0197	0.0229	0.0201
SNP31	-0.0843	0.0194	-0.0827	0.0202	-0.0849	0.0202	-0.0743	0.0206
SNP32	0.0048	0.0189	0.0064	0.0197	0.0024	0.0198	0.0142	0.0202
SNP33	-0.0460	0.0195	-0.0597	0.0203	-0.0339	0.0203	-0.0493	0.0207
SNP34	0.0771	0.0191	0.0811	0.0199	0.0681	0.0200	0.0814	0.0204
SNP35	0.0492	0.0184	0.0468	0.0191	0.0417	0.0192	0.0490	0.0197
SNP36	0.0080	0.0189	0.0093	0.0197	0.0050	0.0197	0.0010	0.0202
SNP37	0.0526	0.0183	0.0499	0.0191	0.0522	0.0192	0.0552	0.0196
SNP38	0.0432	0.0185	0.0443	0.0193	0.0454	0.0193	0.0346	0.0197
SNP39	-0.0230	0.0191	-0.0281	0.0198	-0.0209	0.0198	-0.0175	0.0204
SNP40	0.0052	0.0189	0.0101	0.0196	0.0130	0.0195	0.0138	0.0201
SNP41	0.0608	0.0190	0.0637	0.0198	0.0569	0.0200	0.0491	0.0203
SNP42	-0.0580	0.0186	-0.0646	0.0194	-0.0553	0.0195	-0.0617	0.0200
SNP43	-0.0292	0.0189	-0.0310	0.0196	-0.0382	0.0198	-0.0260	0.0201
SNP44	0.0144	0.0186	0.0145	0.0193	0.0127	0.0195	0.0264	0.0200
SNP45	0.0193	0.0195	0.0235	0.0201	0.0141	0.0204	0.0197	0.0206
SNP46	-0.0233	0.0281	-0.0164	0.0287	-0.0256	0.0292	-0.0086	0.0291
SNP47	-0.0317	0.0192	-0.0367	0.0199	-0.0341	0.0201	-0.0387	0.0204
SNP48	0.0586	0.0192	0.0685	0.0199	0.0669	0.0200	0.0686	0.0205
SNP49	0.0187	0.0190	0.0164	0.0197	0.0248	0.0197	0.0210	0.0202
SNP50	0.0113	0.0186	0.0105	0.0192	0.0127	0.0194	0.0053	0.0198
SNP51	-0.0057	0.0189	-0.0045	0.0197	-0.0041	0.0198	-0.0098	0.0202
SNP52	0.0534	0.0189	0.0581	0.0196	0.0626	0.0198	0.0611	0.0201
SNP53	-0.0354	0.0190	-0.0287	0.0198	-0.0436	0.0198	-0.0345	0.0203
SNP54	0.0724	0.0179	0.0670	0.0187	0.0787	0.0187	0.0628	0.0193
SNP55	0.0460	0.0190	0.0479	0.0196	0.0504	0.0199	0.0401	0.0202
SNP56	-0.0031	0.0190	0.0032	0.0198	0.0005	0.0198	-0.0042	0.0203
SNP57	-0.0461	0.0192	-0.0522	0.0200	-0.0414	0.0201	-0.0568	0.0205
SNP58	-0.0337	0.0190	-0.0274	0.0199	-0.0392	0.0199	-0.0314	0.0202
SNP59	0.0180	0.0188	0.0157	0.0195	0.0262	0.0196	0.0189	0.0201
SNP60	0.0291	0.0179	0.0271	0.0185	0.0276	0.0186	0.0306	0.0192
SNP61	0.0071	0.0193	0.0111	0.0201	0.0179	0.0201	0.0140	0.0207
SNP62	-0.1080	0.0200	-0.1104	0.0208	-0.1093	0.0209	-0.1152	0.0214
SNP63	-0.0571	0.0209	-0.0523	0.0218	-0.0565	0.0219	-0.0524	0.0222
SNP64	0.0755	0.0194	0.0655	0.0201	0.0751	0.0203	0.0753	0.0208
SNP65	-0.0317	0.0370	-0.0464	0.0384	-0.0239	0.0378	-0.0377	0.0382
SNP66	0.0177	0.0189	0.0205	0.0196	0.0229	0.0196	0.0164	0.0201
SNP67	0.0560	0.0188	0.0540	0.0195	0.0503	0.0196	0.0527	0.0199
SNP68	0.0593	0.0186	0.0566	0.0194	0.0544	0.0195	0.0575	0.0198
SNP69	-0.0196	0.0190	-0.0124	0.0198	-0.0200	0.0199	-0.0195	0.0202
SNP70	-0.1222	0.0196	-0.1268	0.0203	-0.1157	0.0205	-0.1180	0.0208
SNP71	-0.0649	0.0191	-0.0669	0.0198	-0.0672	0.0199	-0.0596	0.0203
SNP72	0.0751	0.0196	0.0654	0.0204	0.0777	0.0205	0.0694	0.0209

Table 13: UKB CRC time-dependent coefficient data analysis: estimated regression coefficients and standard errors of each method. 30 censored observations were sampled for each observed failure time.

	PL	PL se	L-opt	L-opt se	A-opt	A-opt se	Unif	Unif se
BMI	0.0005	0.0059	0.0014	0.0059	0.0008	0.0059	-0.0018	0.0060
Smoking	0.0699	0.0595	0.0525	0.0606	0.0704	0.0605	0.0739	0.0609
Family	0.2265	0.0635	0.2167	0.0647	0.2284	0.0647	0.2479	0.0662
Activity	-0.0863	0.0621	-0.0962	0.0633	-0.1005	0.0633	-0.0866	0.0637
Sex	-0.7631	0.2578	-0.7162	0.2615	-0.6907	0.2611	-0.8819	0.2674
Alcohol	-0.0136	0.0743	-0.0275	0.0755	-0.0050	0.0756	-0.0271	0.0762
Alcohol	-0.0827	0.0696	-0.0878	0.0707	-0.0843	0.0709	-0.0945	0.0714
Education1	-0.2621	0.1597	-0.2144	0.1624	-0.2798	0.1604	-0.3697	0.1674
Education2	-0.2513	0.1582	-0.2109	0.1609	-0.2571	0.1589	-0.3654	0.1658
Education3	-0.2551	0.1629	-0.2192	0.1657	-0.2458	0.1637	-0.3588	0.1706
Education4	-0.3019	0.1592	-0.2577	0.1620	-0.2960	0.1599	-0.4076	0.1668
Aspirin	0.0581	0.0497	0.0623	0.0506	0.0470	0.0507	0.0582	0.0516
Ibuprofen	-0.0895	0.0658	-0.1005	0.0671	-0.0778	0.0671	-0.0752	0.0676
Hormones	-0.0304	0.0597	-0.0249	0.0608	-0.0356	0.0608	-0.0335	0.0611
Sex:Activity:ln(time)	0.0318	0.0196	0.0349	0.0200	0.0310	0.0200	0.0364	0.0202
BMI:Sex	0.0296	0.0082	0.0283	0.0083	0.0269	0.0083	0.0330	0.0085
Smoking:Sex	0.1671	0.0789	0.1745	0.0805	0.1694	0.0805	0.1654	0.0812
Sex:Alcohol2	0.0914	0.1072	0.1058	0.1093	0.0909	0.1093	0.0903	0.1106
Sex:Alcohol3	0.2920	0.0962	0.2797	0.0981	0.2819	0.0981	0.3129	0.0991
SNP1	-0.0601	0.0200	-0.0640	0.0203	-0.0577	0.0204	-0.0590	0.0206
SNP2	-0.0642	0.0191	-0.0634	0.0195	-0.0623	0.0195	-0.0658	0.0197
SNP3	0.0459	0.0187	0.0427	0.0191	0.0387	0.0192	0.0536	0.0194
SNP4	0.0039	0.0188	0.0058	0.0191	0.0082	0.0192	-0.0039	0.0195
SNP5	0.0502	0.0191	0.0520	0.0195	0.0515	0.0196	0.0465	0.0197
SNP6	0.0038	0.0190	0.0044	0.0193	0.0079	0.0194	0.0050	0.0197
SNP7	0.0433	0.0183	0.0420	0.0187	0.0486	0.0187	0.0390	0.0189
SNP8	-0.0231	0.0190	-0.0188	0.0194	-0.0210	0.0194	-0.0230	0.0197
SNP9	-0.0220	0.0191	-0.0220	0.0195	-0.0170	0.0196	-0.0204	0.0197
SNP10	0.0272	0.0100	0.0264	0.0101	0.0283	0.0103	0.0371	0.0131
SNP11	0.0503	0.0189	0.0500	0.0193	0.0512	0.0194	0.0536	0.0195
SNP12	-0.0008	0.0189	0.0008	0.0192	0.0038	0.0193	-0.0003	0.0194
SNP13	-0.0088	0.0193	-0.0132	0.0197	-0.0023	0.0197	-0.0048	0.0200
SNP14	0.0119	0.0269	0.0159	0.0275	0.0228	0.0275	0.0082	0.0279
SNP15	0.0583	0.0269	0.0566	0.0274	0.0532	0.0275	0.0575	0.0279
SNP16	0.0806	0.0184	0.0775	0.0188	0.0823	0.0188	0.0809	0.0191
SNP17	0.0054	0.0187	0.0067	0.0190	0.0104	0.0191	0.0083	0.0193
SNP18	-0.0044	0.0208	-0.0058	0.0208	-0.0067	0.0210	0.0009	0.0212
SNP19	0.0317	0.0191	0.0333	0.0195	0.0403	0.0196	0.0379	0.0197
SNP20	-0.0290	0.0194	-0.0298	0.0197	-0.0300	0.0198	-0.0241	0.0201
SNP21	0.0025	0.0190	-0.0002	0.0194	0.0056	0.0195	-0.0047	0.0197
SNP22	0.0342	0.0191	0.0328	0.0194	0.0420	0.0195	0.0394	0.0197
SNP23	-0.0235	0.0191	-0.0232	0.0194	-0.0206	0.0195	-0.0254	0.0197
SNP24	0.0263	0.0190	0.0270	0.0194	0.0254	0.0196	0.0188	0.0197
SNP25	0.0081	0.0189	0.0031	0.0193	0.0083	0.0193	-0.0060	0.0196
SNP26	-0.0186	0.0216	-0.0209	0.0220	-0.0143	0.0221	-0.0228	0.0222
SNP27	0.0692	0.0178	0.0654	0.0182	0.0673	0.0182	0.0632	0.0185
SNP28	-0.0035	0.0213	-0.0027	0.0218	-0.0030	0.0217	-0.0075	0.0220
SNP29	-0.1322	0.0192	-0.1336	0.0195	-0.1316	0.0197	-0.1313	0.0198
SNP30	0.0226	0.0188	0.0204	0.0193	0.0200	0.0193	0.0246	0.0195
SNP31	-0.0843	0.0194	-0.0921	0.0198	-0.0836	0.0199	-0.0837	0.0200
SNP32	0.0048	0.0189	0.0014	0.0193	0.0025	0.0193	0.0125	0.0196
SNP33	-0.0460	0.0195	-0.0465	0.0198	-0.0481	0.0199	-0.0523	0.0201
SNP34	0.0771	0.0191	0.0773	0.0195	0.0720	0.0195	0.0845	0.0198
SNP35	0.0492	0.0184	0.0488	0.0187	0.0480	0.0187	0.0446	0.0191
SNP36	0.0080	0.0189	-0.0006	0.0192	0.0079	0.0192	0.0010	0.0195
SNP37	0.0526	0.0183	0.0586	0.0186	0.0492	0.0187	0.0513	0.0190
SNP38	0.0432	0.0185	0.0401	0.0189	0.0340	0.0189	0.0393	0.0191
SNP39	-0.0230	0.0191	-0.0186	0.0195	-0.0269	0.0194	-0.0245	0.0197
SNP40	0.0052	0.0189	0.0108	0.0192	-0.0007	0.0193	0.0078	0.0195
SNP41	0.0608	0.0190	0.0578	0.0194	0.0614	0.0195	0.0528	0.0196
SNP42	-0.0580	0.0186	-0.0602	0.0190	-0.0634	0.0190	-0.0616	0.0194
SNP43	-0.0292	0.0189	-0.0276	0.0193	-0.0332	0.0193	-0.0294	0.0195
SNP44	0.0144	0.0186	0.0150	0.0190	0.0116	0.0191	0.0237	0.0193
SNP45	0.0193	0.0195	0.0194	0.0198	0.0209	0.0199	0.0179	0.0201
SNP46	-0.0233	0.0281	-0.0276	0.0288	-0.0197	0.0284	-0.0117	0.0288
SNP47	-0.0317	0.0192	-0.0342	0.0195	-0.0290	0.0196	-0.0379	0.0198
SNP48	0.0586	0.0192	0.0613	0.0195	0.0601	0.0196	0.0633	0.0199
SNP49	0.0187	0.0190	0.0187	0.0192	0.0215	0.0194	0.0183	0.0196
SNP50	0.0113	0.0186	0.0102	0.0189	0.0171	0.0190	0.0094	0.0192
SNP51	-0.0057	0.0189	0.0036	0.0194	-0.0023	0.0194	-0.0094	0.0196
SNP52	0.0534	0.0189	0.0546	0.0193	0.0527	0.0194	0.0563	0.0196
SNP53	-0.0354	0.0190	-0.0370	0.0194	-0.0387	0.0195	-0.0386	0.0197
SNP54	0.0724	0.0179	0.0708	0.0183	0.0751	0.0183	0.0646	0.0186
SNP55	0.0460	0.0190	0.0492	0.0193	0.0440	0.0194	0.0491	0.0196
SNP56	-0.0031	0.0190	0.0012	0.0193	-0.0027	0.0194	-0.0003	0.0196
SNP57	-0.0461	0.0192	-0.0470	0.0197	-0.0432	0.0197	-0.0520	0.0198
SNP58	-0.0337	0.0190	-0.0376	0.0194	-0.0352	0.0193	-0.0292	0.0196
SNP59	0.0180	0.0188	0.0193	0.0191	0.0203	0.0192	0.0159	0.0194
SNP60	0.0291	0.0179	0.0317	0.0182	0.0235	0.0182	0.0337	0.0185
SNP61	0.0071	0.0193	0.0108	0.0197	0.0095	0.0198	0.0014	0.0201
SNP62	-0.1080	0.0200	-0.1134	0.0204	-0.1075	0.0205	-0.1120	0.0207
SNP63	-0.0571	0.0209	-0.0544	0.0214	-0.0644	0.0214	-0.0584	0.0215
SNP64	0.0755	0.0194	0.0738	0.0198	0.0760	0.0199	0.0778	0.0201
SNP65	-0.0317	0.0370	-0.0344	0.0373	-0.0349	0.0373	-0.0432	0.0392
SNP66	0.0177	0.0189	0.0201	0.0192	0.0111	0.0193	0.0132	0.0194
SNP67	0.0560	0.0188	0.0551	0.0191	0.0585	0.0191	0.0565	0.0194
SNP68	0.0593	0.0186	0.0619	0.0190	0.0591	0.0191	0.0554	0.0193
SNP69	-0.0196	0.0190	-0.0193	0.0194	-0.0159	0.0194	-0.0187	0.0196
SNP70	-0.1222	0.0196	-0.1170	0.0200	-0.1252	0.0200	-0.1221	0.0202
SNP71	-0.0649	0.0191	-0.0691	0.0195	-0.0662	0.0195	-0.0632	0.0197
SNP72	0.0751	0.0196	0.0743	0.0200	0.0730	0.0200	0.0776	0.0202

Table 14: UKB CRC time-dependent coefficient data analysis: estimated regression coefficients and standard errors of each method. 71 censored observations were sampled for each observed failure time.

	PL	PL se	L-opt	L-opt se	A-opt	A-opt se	Unif	Unif se
BMI	0.0005	0.0059	-0.0002	0.0059	0.0003	0.0059	-0.0000	0.0059
Smoking	0.0699	0.0595	0.0675	0.0599	0.0740	0.0600	0.0752	0.0601
Family	0.2265	0.0635	0.2327	0.0640	0.2390	0.0640	0.2283	0.0647
Activity	-0.0863	0.0621	-0.1010	0.0626	-0.0938	0.0626	-0.0873	0.0628
Sex	-0.7631	0.2578	-0.7472	0.2605	-0.7691	0.2595	-0.8063	0.2620
Alcohol	-0.0136	0.0743	-0.0139	0.0748	0.0059	0.0748	-0.0177	0.0751
Alcohol	-0.0827	0.0696	-0.0755	0.0701	-0.0952	0.0702	-0.0816	0.0703
Education1	-0.2621	0.1597	-0.2400	0.1610	-0.2480	0.1600	-0.2901	0.1628
Education2	-0.2513	0.1582	-0.2402	0.1595	-0.2424	0.1585	-0.2810	0.1612
Education3	-0.2551	0.1629	-0.2424	0.1642	-0.2431	0.1632	-0.2707	0.1660
Education4	-0.3019	0.1592	-0.2898	0.1605	-0.2866	0.1595	-0.3292	0.1622
Aspirin	0.0581	0.0497	0.0483	0.0501	0.0610	0.0501	0.0533	0.0505
Ibuprofen	-0.0895	0.0658	-0.0952	0.0663	-0.0973	0.0663	-0.0854	0.0665
Hormones	-0.0304	0.0597	-0.0303	0.0601	-0.0270	0.0601	-0.0305	0.0603
Sex:Activity:ln(time)	0.0318	0.0196	0.0344	0.0198	0.0330	0.0198	0.0358	0.0199
BMI:Sex	0.0296	0.0082	0.0289	0.0083	0.0297	0.0082	0.0305	0.0083
Smoking:Sex	0.1671	0.0789	0.1758	0.0795	0.1697	0.0796	0.1649	0.0798
Sex:Alcohol2	0.0914	0.1072	0.0924	0.1081	0.0835	0.1081	0.0903	0.1086
Sex:Alcohol3	0.2920	0.0962	0.2809	0.0970	0.3146	0.0970	0.3017	0.0974
SNP1	-0.0601	0.0200	-0.0655	0.0201	-0.0603	0.0202	-0.0625	0.0202
SNP2	-0.0642	0.0191	-0.0630	0.0192	-0.0599	0.0192	-0.0657	0.0193
SNP3	0.0459	0.0187	0.0430	0.0189	0.0472	0.0189	0.0510	0.0190
SNP4	0.0039	0.0188	0.0045	0.0189	0.0071	0.0190	0.0002	0.0191
SNP5	0.0502	0.0191	0.0528	0.0193	0.0492	0.0193	0.0526	0.0194
SNP6	0.0038	0.0190	0.0080	0.0191	0.0055	0.0192	0.0036	0.0193
SNP7	0.0433	0.0183	0.0437	0.0185	0.0421	0.0185	0.0400	0.0186
SNP8	-0.0231	0.0190	-0.0198	0.0192	-0.0230	0.0192	-0.0239	0.0193
SNP9	-0.0220	0.0191	-0.0203	0.0193	-0.0226	0.0193	-0.0223	0.0194
SNP10	0.0272	0.0100	0.0270	0.0101	0.0266	0.0102	0.0315	0.0111
SNP11	0.0503	0.0189	0.0509	0.0190	0.0554	0.0190	0.0526	0.0192
SNP12	-0.0008	0.0189	-0.0043	0.0190	-0.0025	0.0190	0.0016	0.0191
SNP13	-0.0088	0.0193	-0.0057	0.0195	-0.0044	0.0195	-0.0063	0.0196
SNP14	0.0119	0.0269	0.0101	0.0272	0.0062	0.0271	0.0121	0.0275
SNP15	0.0583	0.0269	0.0582	0.0272	0.0680	0.0272	0.0591	0.0274
SNP16	0.0806	0.0184	0.0813	0.0186	0.0825	0.0186	0.0809	0.0187
SNP17	0.0054	0.0187	0.0060	0.0189	0.0048	0.0189	0.0060	0.0189
SNP18	-0.0044	0.0208	-0.0042	0.0209	-0.0048	0.0208	-0.0008	0.0210
SNP19	0.0317	0.0191	0.0287	0.0193	0.0332	0.0193	0.0351	0.0194
SNP20	-0.0290	0.0194	-0.0298	0.0195	-0.0319	0.0195	-0.0295	0.0197
SNP21	0.0025	0.0190	0.0016	0.0192	0.0056	0.0192	-0.0037	0.0193
SNP22	0.0342	0.0191	0.0353	0.0192	0.0332	0.0192	0.0341	0.0193
SNP23	-0.0235	0.0191	-0.0189	0.0192	-0.0250	0.0192	-0.0277	0.0194
SNP24	0.0263	0.0190	0.0261	0.0192	0.0263	0.0192	0.0258	0.0193
SNP25	0.0081	0.0189	0.0092	0.0191	0.0097	0.0192	0.0044	0.0192
SNP26	-0.0186	0.0216	-0.0200	0.0217	-0.0137	0.0218	-0.0164	0.0218
SNP27	0.0692	0.0178	0.0689	0.0180	0.0674	0.0180	0.0671	0.0182
SNP28	-0.0035	0.0213	-0.0001	0.0215	0.0005	0.0216	-0.0042	0.0216
SNP29	-0.1322	0.0192	-0.1273	0.0194	-0.1375	0.0193	-0.1383	0.0195
SNP30	0.0226	0.0188	0.0211	0.0190	0.0206	0.0190	0.0227	0.0191
SNP31	-0.0843	0.0194	-0.0843	0.0195	-0.0833	0.0196	-0.0825	0.0197
SNP32	0.0048	0.0189	0.0001	0.0191	0.0058	0.0191	0.0071	0.0192
SNP33	-0.0460	0.0195	-0.0430	0.0196	-0.0471	0.0196	-0.0470	0.0197
SNP34	0.0771	0.0191	0.0788	0.0193	0.0793	0.0193	0.0744	0.0194
SNP35	0.0492	0.0184	0.0502	0.0185	0.0457	0.0186	0.0456	0.0187
SNP36	0.0080	0.0189	0.0103	0.0190	0.0039	0.0191	0.0060	0.0191
SNP37	0.0526	0.0183	0.0529	0.0184	0.0512	0.0185	0.0515	0.0186
SNP38	0.0432	0.0185	0.0468	0.0186	0.0419	0.0187	0.0421	0.0187
SNP39	-0.0230	0.0191	-0.0218	0.0192	-0.0258	0.0193	-0.0246	0.0194
SNP40	0.0052	0.0189	0.0060	0.0190	0.0051	0.0190	0.0054	0.0192
SNP41	0.0608	0.0190	0.0606	0.0192	0.0616	0.0192	0.0549	0.0193
SNP42	-0.0580	0.0186	-0.0569	0.0188	-0.0546	0.0188	-0.0574	0.0189
SNP43	-0.0292	0.0189	-0.0279	0.0190	-0.0296	0.0191	-0.0295	0.0192
SNP44	0.0144	0.0186	0.0130	0.0188	0.0125	0.0189	0.0167	0.0189
SNP45	0.0193	0.0195	0.0202	0.0196	0.0221	0.0197	0.0196	0.0197
SNP46	-0.0233	0.0281	-0.0214	0.0282	-0.0264	0.0284	-0.0189	0.0283
SNP47	-0.0317	0.0192	-0.0318	0.0194	-0.0295	0.0194	-0.0300	0.0194
SNP48	0.0586	0.0192	0.0572	0.0193	0.0591	0.0194	0.0570	0.0195
SNP49	0.0187	0.0190	0.0199	0.0191	0.0209	0.0192	0.0200	0.0192
SNP50	0.0113	0.0186	0.0132	0.0187	0.0115	0.0188	0.0091	0.0189
SNP51	-0.0057	0.0189	-0.0096	0.0191	-0.0082	0.0191	-0.0064	0.0192
SNP52	0.0534	0.0189	0.0536	0.0191	0.0529	0.0191	0.0557	0.0192
SNP53	-0.0354	0.0190	-0.0380	0.0192	-0.0333	0.0192	-0.0410	0.0193
SNP54	0.0724	0.0179	0.0706	0.0181	0.0741	0.0181	0.0719	0.0182
SNP55	0.0460	0.0190	0.0471	0.0191	0.0472	0.0192	0.0456	0.0193
SNP56	-0.0031	0.0190	-0.0031	0.0191	-0.0023	0.0192	-0.0029	0.0193
SNP57	-0.0461	0.0192	-0.0437	0.0194	-0.0457	0.0194	-0.0505	0.0195
SNP58	-0.0337	0.0190	-0.0349	0.0191	-0.0383	0.0191	-0.0324	0.0192
SNP59	0.0180	0.0188	0.0174	0.0189	0.0161	0.0189	0.0176	0.0190
SNP60	0.0291	0.0179	0.0251	0.0180	0.0316	0.0180	0.0278	0.0182
SNP61	0.0071	0.0193	0.0053	0.0195	0.0048	0.0196	0.0048	0.0197
SNP62	-0.1080	0.0200	-0.1082	0.0201	-0.1069	0.0202	-0.1071	0.0203
SNP63	-0.0571	0.0209	-0.0602	0.0211	-0.0614	0.0211	-0.0571	0.0212
SNP64	0.0755	0.0194	0.0759	0.0196	0.0794	0.0196	0.0747	0.0197
SNP65	-0.0317	0.0370	-0.0317	0.0372	-0.0299	0.0372	-0.0353	0.0375
SNP66	0.0177	0.0189	0.0183	0.0190	0.0193	0.0191	0.0180	0.0191
SNP67	0.0560	0.0188	0.0579	0.0189	0.0587	0.0189	0.0594	0.0191
SNP68	0.0593	0.0186	0.0591	0.0188	0.0601	0.0188	0.0578	0.0189
SNP69	-0.0196	0.0190	-0.0217	0.0191	-0.0192	0.0192	-0.0208	0.0192
SNP70	-0.1222	0.0196	-0.1284	0.0197	-0.1223	0.0198	-0.1218	0.0199
SNP71	-0.0649	0.0191	-0.0632	0.0192	-0.0659	0.0193	-0.0602	0.0193
SNP72	0.0751	0.0196	0.0737	0.0197	0.0744	0.0197	0.0722	0.0199

Table 15: UKB CRC marginal analysis: metrics for evaluating the different methods. The RMSE and Frobenius distance are with respect to the full data partial-likelihood estimator. 4 censored observations were sampled for each observed failure time.

	RMSE	$\|.\|_F$	Runtime (seconds)
PL	0.000	0.000	658.280
L-opt	0.194	0.027	54.680
A-opt	0.156	0.016	74.070
Uniform	0.426	0.058	11.750

References

Abed, A. and F. Belzile (2019). Comparing single-snp, multi-snp, and haplotype-based approaches in association studies for major traits in barley. *The Plant Genome 12*(3), 190036.

Ai, M., J. Yu, H. Zhang, and H. Wang (2018). Optimal subsampling algorithms for big data generalized linear models. *arXiv preprint arXiv:1806.06761*.

Binder, H., A. Allignol, M. Schumacher, and J. Beyersmann (2009). Boosting for high-dimensional time-to-event data with competing risks. *Bioinformatics 25*(7), 890–896.

Breslow, N. E. (1972). Contribution to discussion of paper by dr cox. *J. Roy. Statist. Soc., Ser. B 34*, 216–217.

Bush, W. S. and J. H. Moore (2012). Genome-wide association studies. *PLoS Comput Biol 8*(12), e1002822.

Bycroft, C., C. Freeman, D. Petkova, G. Band, L. T. Elliott, K. Sharp, A. Motyer, D. Vukcevic, O. Delaneau, J. O'Connell, et al. (2018). The uk biobank resource with deep phenotyping and genomic data. *Nature 562*(7726), 203–209.

Carstensen, B., M. Plummer, E. Laara, and M. Hills (2020). *Epi: A Package for Statistical Analysis in Epidemiology*. R package version 2.42.

Chen, K. and S.-H. Lo (1999). Case-cohort and case-control analysis with cox's model. *Biometrika 86*(4), 755–764.

Cox, D. R. (1972). Regression models and life-tables. *Journal of the Royal Statistical Society: Series B (Methodological) 34*(2), 187–202.

Davidson-Pilon, C., J. Kalderstam, N. Jacobson, sean reed, B. Kuhn, P. Zivich, M. Williamson, AbdealiJK, D. Datta, A. Fiore-Gartland, A. Parij, D. WIlson, Gabriel, L. Moneda, K. Stark, A. Moncada-Torres, H. Gadgil, Jona, K. Singaravelan, L. Besson, M. S. Peña, S. Anton, A. Klintberg, J. Noorbakhsh, M. Begun, R. Kumar, S. Hussey, D. Golland, jlim13, and A. Flaxman (2020, August). Camdavidsonpilon/lifelines: v0.25.1.

Denny, J. C., S. L. Van Driest, W.-Q. Wei, and D. M. Roden (2018). The influence of big (clinical) data and genomics on precision medicine and drug development. *Clinical Pharmacology & Therapeutics 103*(3), 409–418.

Dhillon, P., Y. Lu, D. P. Foster, and L. Ungar (2013). New subsampling algorithms for fast least squares regression. In *Advances in neural information processing systems*, pp. 360–368.

Fleming, T. R. and D. P. Harrington (2011). *Counting processes and survival analysis*, Volume 169. John Wiley & Sons.

Gorfine, M., N. Keret, A. Ben Arie, D. Zucker, and L. Hsu (2020). Marginalized frailty-based illness-death model: Application to the uk-biobank survival data. *Journal of the American Statistical Association*, 1–13.

Hughey, J. J., S. D. Rhoades, D. Y. Fu, L. Bastarache, J. C. Denny, and Q. Chen (2019). Cox regression increases power to detect genotype-phenotype associations in genomic studies using the electronic health record. *BMC genomics 20*(1), 1–7.

Ishwaran, H., U. B. Kogalur, E. H. Blackstone, M. S. Lauer, et al. (2008). Random survival forests. *The annals of applied statistics 2*(3), 841–860.

Jeon, J., M. Du, R. E. Schoen, M. Hoffmeister, P. A. Newcomb, S. I. Berndt, B. Caan, P. T. Campbell, A. T. Chan, J. Chang-Claude, et al. (2018). Determining risk of colorectal cancer and starting age of screening based on lifestyle, environmental, and genetic factors. *Gastroenterology 154*(8), 2152–2164.

Johansson, A. L., T. M.-L. Andersson, C.-C. Hsieh, S. Cnattingius, P. W. Dickman, and M. Lambe (2015). Family history and risk of pregnancy-associated breast cancer (pabc). *Breast Cancer Research and Treatment 151*(1), 209–217.

Keogh, R. H. and I. R. White (2013). Using full-cohort data in nested case–control and case–cohort studies by multiple imputation. *Statistics in medicine 32*(23), 4021–4043.

Kim, S., J. Cai, and W. Lu (2013). More efficient estimators for case-cohort studies. *Biometrika 100*(3), 695–708.

Klein, J. P. and M. L. Moeschberger (2006). *Survival analysis: techniques for censored and truncated data*. Springer Science & Business Media.

Kulich, M. and D. Lin (2004). Improving the efficiency of relative-risk estimation in case-cohort studies. *Journal of the American Statistical Association 99*(467), 832–844.

Langholz, B. and Ø. R. Borgan (1995). Counter-matching: a stratified nested case-control sampling method. *Biometrika 82*(1), 69–79.

Liddell, F., J. McDonald, and D. Thomas (1977). Methods of cohort analysis: appraisal by application to asbestos mining. *Journal of the Royal Statistical Society: Series A (General) 140*(4), 469–483.

Ma, P., M. W. Mahoney, and B. Yu (2015). A statistical perspective on algorithmic leveraging. *The Journal of Machine Learning Research 16*(1), 861–911.

Mittal, S., D. Madigan, R. S. Burd, and M. A. Suchard (2014). High-dimensional, massive sample-size cox proportional hazards regression for survival analysis. *Biostatistics 15*(2), 207–221.

Pang, Y., C. Kartsonaki, Y. Guo, Y. Chen, L. Yang, Z. Bian, F. Bragg, I. Y. Millwood, E. Mao, Y. Li, et al. (2018). Adiposity and risks of colorectal and small intestine cancer in chinese adults: A prospective study of 0.5 million people. *British journal of cancer 119*(2), 248–250.

Prentice, R. L. (1986). A case-cohort design for epidemiologic cohort studies and disease prevention trials. *Biometrika 73*(1), 1–11.

Pukelsheim, F. (2006). *Optimal design of experiments*. SIAM.

R Core Team (2013). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.

Raghupathi, W. and V. Raghupathi (2014). Big data analytics in healthcare: promise and potential. *Health information science and systems 2*(1), 3.

Samuelsen, S. O. (1997). A psudolikelihood approach to analysis of nested case-control studies. *Biometrika 84*(2), 379–394.

Staley, J. R., E. Jones, S. Kaptoge, A. S. Butterworth, M. J. Sweeting, A. M. Wood, and J. M. Howson (2017). A comparison of cox and logistic regression for use in genome-wide association studies of cohort and case-cohort design. *European Journal of Human Genetics 25*(7), 854–862.

Therneau, T., C. Crowson, and E. Atkinson (2017). Using time dependent covariates and time dependent coefficients in the cox model. *Survival Vignettes*.

Therneau, T. M. (2020). *A Package for Survival Analysis in R*. R package version 3.2-3.

Therneau, T. M., P. M. Grambsch, and T. R. Fleming (1990). Martingale-based residuals for survival models. *Biometrika 77*(1), 147–160.

Tsai, W.-Y. (1990). Testing the assumption of independence of truncation time and failure time. *Biometrika 77*(1), 169–177.

UK-Biobank (2016). *UK Biobank Cancer Numbers Summary Report*.

Van Der Net, J. B., A. C. J. Janssens, M. J. Eijkemans, J. J. Kastelein, E. J. Sijbrands, and E. W. Steyerberg (2008). Cox proportional hazards models have more statistical power than logistic regression models in cross-sectional genetic association studies. *European Journal of Human Genetics 16*(9), 1111–1116.

van der Vaart, A. W. (1998). *Asymptotic Statistics*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press.

Wang, H. and Y. Ma (2020). Optimal subsampling for quantile regression in big data. *arXiv preprint arXiv:2001.10168*.

Wang, H., R. Zhu, and P. Ma (2018). Optimal subsampling for large sample logistic regression. *Journal of the American Statistical Association 113*(522), 829–844.

Yang, J., T. Ferreira, A. P. Morris, S. E. Medland, P. A. Madden, A. C. Heath, N. G. Martin, G. W. Montgomery, M. N. Weedon, R. J. Loos, et al. (2012). Conditional and joint multiple-snp analysis of gwas summary statistics identifies additional variants influencing complex traits. *Nature genetics 44*(4), 369–375.

Yu, J., H. Wang, M. Ai, and H. Zhang (2020). Optimal distributed subsampling for maximum quasi-likelihood estimators with massive data. *Journal of the American Statistical Association* (just-accepted), 1–29.

Zheng, Y., M. Brown, A. Lok, and T. Cai (2017). Improving efficiency in biomarker incremental value evaluation under two-phase designs. *The annals of applied statistics 11*(2), 638.

Zuo, L., H. Zhang, H. Wang, and L. Liu (2020). Sampling-based estimation for massive survival data with additive hazards model. *Statistics in Medicine*.

Table 16: UKB CRC marginal analysis: estimated regression coefficients and standard errors of each method for all 72 SNPs. 4 censored observations were sampled for each observed failure time.

	PL	PL se	L-opt	L se	A-opt	A se	Unif	Unif se
SNP1	-0.0574	0.0199	-0.0563	0.0224	-0.0522	0.0230	-0.0524	0.0227
SNP2	-0.0606	0.0191	-0.0628	0.0215	-0.0561	0.0217	-0.0675	0.0221
SNP3	0.0465	0.0187	0.0442	0.0212	0.0569	0.0216	0.0245	0.0219
SNP4	0.0046	0.0188	0.0190	0.0211	0.0058	0.0215	0.0216	0.0220
SNP5	0.0532	0.0191	0.0478	0.0215	0.0492	0.0220	0.0785	0.0222
SNP6	0.0031	0.0190	-0.0020	0.0214	0.0037	0.0218	-0.0110	0.0219
SNP7	0.0439	0.0183	0.0408	0.0208	0.0426	0.0209	0.0400	0.0214
SNP8	-0.0261	0.0189	-0.0226	0.0214	-0.0136	0.0219	-0.0330	0.0222
SNP9	-0.0184	0.0191	-0.0214	0.0215	-0.0215	0.0220	-0.0121	0.0220
SNP10	0.0264	0.0102	0.0254	0.0109	0.0265	0.0101	0.0396	0.0183
SNP11	0.0535	0.0188	0.0534	0.0212	0.0653	0.0216	0.0450	0.0219
SNP12	0.0005	0.0188	0.0065	0.0214	-0.0104	0.0217	-0.0063	0.0219
SNP13	-0.0107	0.0193	-0.0052	0.0217	-0.0165	0.0221	-0.0061	0.0223
SNP14	0.0531	0.0189	0.0404	0.0214	0.0484	0.0218	0.0466	0.0220
SNP15	0.0668	0.0189	0.0625	0.0214	0.0658	0.0217	0.0797	0.0221
SNP16	0.0818	0.0184	0.0769	0.0209	0.0746	0.0212	0.0913	0.0214
SNP17	0.0055	0.0186	0.0091	0.0211	-0.0009	0.0212	-0.0106	0.0212
SNP18	-0.0051	0.0207	-0.0142	0.0221	-0.0023	0.0221	0.0029	0.0237
SNP19	0.0330	0.0191	0.0321	0.0215	0.0393	0.0222	0.0389	0.0224
SNP20	-0.0279	0.0194	-0.0319	0.0219	-0.0352	0.0223	-0.0322	0.0225
SNP21	-0.0002	0.0190	0.0076	0.0214	-0.0067	0.0216	-0.0002	0.0221
SNP22	0.0297	0.0190	0.0249	0.0215	0.0298	0.0219	0.0403	0.0223
SNP23	-0.0196	0.0190	-0.0072	0.0214	-0.0245	0.0219	-0.0202	0.0222
SNP24	0.0220	0.0190	0.0274	0.0214	0.0246	0.0218	0.0189	0.0220
SNP25	0.0098	0.0189	0.0101	0.0213	0.0186	0.0216	0.0138	0.0219
SNP26	-0.0273	0.0193	-0.0319	0.0218	-0.0247	0.0220	-0.0191	0.0225
SNP27	0.0677	0.0176	0.0668	0.0200	0.0621	0.0202	0.0653	0.0209
SNP28	0.0116	0.0192	0.0272	0.0216	0.0096	0.0221	0.0033	0.0224
SNP29	-0.1272	0.0191	-0.1309	0.0216	-0.1305	0.0218	-0.1309	0.0221
SNP30	0.0239	0.0188	0.0103	0.0214	0.0131	0.0217	0.0340	0.0219
SNP31	-0.0807	0.0194	-0.0917	0.0218	-0.0924	0.0220	-0.0736	0.0223
SNP32	0.0064	0.0189	0.0118	0.0211	0.0003	0.0219	0.0131	0.0219
SNP33	-0.0464	0.0195	-0.0603	0.0219	-0.0303	0.0225	-0.0507	0.0225
SNP34	0.0779	0.0191	0.0910	0.0216	0.1078	0.0220	0.0798	0.0222
SNP35	0.0510	0.0184	0.0382	0.0206	0.0480	0.0210	0.0579	0.0215
SNP36	0.0093	0.0189	0.0088	0.0212	-0.0053	0.0217	-0.0114	0.0219
SNP37	0.0507	0.0182	0.0413	0.0206	0.0485	0.0210	0.0549	0.0216
SNP38	0.0431	0.0185	0.0502	0.0208	0.0463	0.0211	0.0539	0.0217
SNP39	-0.0196	0.0190	-0.0026	0.0215	-0.0125	0.0215	-0.0112	0.0221
SNP40	0.0068	0.0189	0.0019	0.0215	0.0110	0.0216	0.0046	0.0219
SNP41	0.0646	0.0190	0.0507	0.0215	0.0694	0.0216	0.0599	0.0221
SNP42	-0.0581	0.0186	-0.0390	0.0209	-0.0550	0.0215	-0.0472	0.0218
SNP43	-0.0268	0.0189	-0.0170	0.0213	-0.0054	0.0216	-0.0184	0.0221
SNP44	0.0159	0.0186	0.0144	0.0210	0.0156	0.0214	0.0113	0.0215
SNP45	0.0216	0.0194	0.0201	0.0220	0.0239	0.0225	0.0194	0.0226
SNP46	-0.0409	0.0273	-0.0499	0.0304	-0.0512	0.0296	-0.0478	0.0291
SNP47	-0.0309	0.0192	-0.0622	0.0216	-0.0409	0.0220	-0.0176	0.0222
SNP48	0.0507	0.0191	0.0536	0.0214	0.0477	0.0219	0.0497	0.0222
SNP49	0.0194	0.0190	0.0143	0.0214	0.0253	0.0218	0.0132	0.0219
SNP50	0.0145	0.0186	0.0095	0.0211	0.0162	0.0212	0.0142	0.0215
SNP51	-0.0050	0.0189	-0.0190	0.0215	-0.0077	0.0218	-0.0159	0.0219
SNP52	0.0559	0.0189	0.0581	0.0214	0.0302	0.0217	0.0330	0.0219
SNP53	-0.0330	0.0189	-0.0180	0.0213	-0.0514	0.0218	-0.0133	0.0218
SNP54	0.0720	0.0178	0.0708	0.0200	0.0888	0.0207	0.0769	0.0212
SNP55	0.0517	0.0189	0.0486	0.0214	0.0760	0.0219	0.0393	0.0220
SNP56	-0.0016	0.0190	-0.0073	0.0212	0.0036	0.0216	0.0069	0.0223
SNP57	-0.0452	0.0192	-0.0415	0.0217	-0.0434	0.0220	-0.0285	0.0223
SNP58	-0.0316	0.0189	-0.0131	0.0212	-0.0156	0.0217	-0.0539	0.0217
SNP59	0.0162	0.0188	0.0217	0.0211	0.0237	0.0215	0.0127	0.0218
SNP60	0.0286	0.0178	0.0265	0.0203	0.0219	0.0204	0.0151	0.0209
SNP61	-0.0228	0.0185	-0.0238	0.0207	-0.0281	0.0211	-0.0354	0.0216
SNP62	-0.1080	0.0191	-0.1216	0.0216	-0.1173	0.0221	-0.1046	0.0223
SNP63	-0.0637	0.0207	-0.0650	0.0233	-0.0700	0.0237	-0.0747	0.0237
SNP64	0.0744	0.0194	0.0694	0.0218	0.0780	0.0222	0.0690	0.0225
SNP65	-0.0434	0.0367	-0.0305	0.0402	-0.0344	0.0389	-0.0430	0.0373
SNP66	0.0167	0.0189	0.0150	0.0214	0.0164	0.0215	0.0077	0.0219
SNP67	0.0578	0.0187	0.0621	0.0212	0.0537	0.0214	0.0514	0.0219
SNP68	0.0559	0.0186	0.0655	0.0209	0.0659	0.0214	0.0576	0.0218
SNP69	-0.0176	0.0190	-0.0139	0.0213	-0.0100	0.0217	-0.0102	0.0222
SNP70	-0.1252	0.0195	-0.1363	0.0222	-0.1357	0.0225	-0.1145	0.0227
SNP71	-0.0676	0.0191	-0.0561	0.0215	-0.0721	0.0219	-0.0626	0.0222
SNP72	0.0787	0.0195	0.0874	0.0218	0.0711	0.0223	0.0556	0.0225