The Annals of Statistics
1995, Vol. 23, No. 5, 1749–1778

METHODS FOR THE ANALYSIS OF SAMPLED COHORT DATA IN THE COX PROPORTIONAL HAZARDS MODEL

BY Ø. BORGAN,[1] L. GOLDSTEIN[2] AND B. LANGHOLZ[3]

University of Oslo, University of Southern California and University of Southern California

Methods are provided for regression parameter and cumulative baseline hazard estimation in the Cox proportional hazards model when the cohort is sampled according to a predictable sampling probability law. The key to the methodology is to define counting processes which count joint failure and sampled risk sets occurrences and to derive the appropriate intensities for these counting processes, leading to estimation methods parallel to those for full cohort data. These techniques are illustrated for a number of sampling designs, including three novel designs: counter-matching with additional randomly sampled controls; quota sampling of controls; and nested case-control sampling with number of controls dependent on the failure's exposure status. General asymptotic theory is developed for the maximum partial likelihood estimator and cumulative baseline hazard estimator and is used to derive the asymptotic distributions for estimators from a large class of designs.

1. Introduction. Epidemiologic cohort studies are considered the most reliable method for assessing the variation in rates of morbidity and mortality due to factors present in the population under study. Cohort members are observed over some time period and either "fail" (develop or die from the disease of interest) or are "censored" (are alive at the end of the study period, die of some other cause or are lost to follow-up). Variation in rates are then modeled from information on exposures, confounders and other potential predictors of risk, which we generically call *covariates*, collected on cohort members. If complete covariate information is obtained for all cohort members, a wide range of parametric and semiparametric analytic techniques are available [e.g., Breslow and Day (1987)].

Especially useful has been the Cox regression model in which the process counting failures for subject i is assumed to have intensity process of the form

$$\lambda_i(t) = Y_i(t)\,\alpha_0(t)\exp\left(\beta_0^T\mathbf{Z}_i(t)\right), \tag{1.1}$$

Received January 1993; revised March 1995.
[1]Supported by Norwegian Research Council for Science and the Humanities.
[2]Supported by NSF Grant DMS-90-05833.
[3]Supported by National Cancer Institute Grant CA14089.
AMS 1991 *subject classifications*. 62D05, 62F12, 62G05, 62M99, 62P10.
Key words and phrases. Case-control study, cohort sampling, counting process, Cox's regression model, epidemiology, marked point process, martingale, partial likelihood, survival analysis.

with $\alpha_0(t)$ a nonnegative baseline intensity or hazard function and β_0 a p-dimensional vector of regression parameters [Cox (1972), Andersen and Gill (1982)]. Here $Y_i(t)$ is an indicator process taking the value 1 if the ith subject is on study (at risk) at $t-$ and taking the value 0 otherwise, and $\mathbf{Z}_i(t)$ is a vector of (possibly) time dependent covariates for this individual. The partial likelihood used for estimation of the regression parameters is given as

$$\mathscr{L}(\beta) = \prod_{t_j} \left\{ \frac{\exp\left(\beta^T \mathbf{Z}_{i_j}(t_j)\right)}{\sum_{l \in \mathscr{R}_j} \exp\left(\beta^T \mathbf{Z}_l(t_j)\right)} \right\}, \tag{1.2}$$

where the t_j are the ordered failure times, i_j is the index of the failure at time t_j and $\mathscr{R}_j$ is the set of all those at risk at t_j-, the failure and those on study at that time. Further, nonparametric estimators of the cumulative baseline hazard and related quantities are available [e.g., Andersen, Borgan, Gill and Keiding (1993), Section VII.2].

Typically, because of the rarity of the disease outcome or the complexity of the relationships to be explored, cohort studies require very large numbers of subjects or long periods of follow-up in order to accumulate enough failures to have sufficient statistical power to give reliable answers to the questions of interest. This leads to something of a paradox. If a cohort study is large enough to allow for a meaningful analysis, the cost of collecting high-quality covariate information on all subjects is prohibitively expensive, if not logistically impossible. It would also seem unnecessary. Loosely, if the disease of interest is rare, the contribution of the nonfailures, in terms of the statistical power of the study, will be negligible compared to that of the failures. Thus cohort sampling methods which include all the failures and a portion of the nonfailures are highly desirable. Standard case-control study designs effectively exploit this principle.

The nested case-control design [Thomas (1977)] is of particular interest in the context of this paper because, with the exception of a sampling-with-replacement variant [Robins, Gail and Lubin (1986)], it has been the only cohort sampling method which is analyzed using partial likelihood techniques. In this design, "sampled risk sets" $\tilde{\mathscr{R}}_j$ consist of the "case" (failure) at t_j and $m-1$ "controls" randomly selected from those at risk at t_j. The nested case-control partial likelihood has the same form as that for the full cohort except that $\tilde{\mathbf{R}}_j$ replaces $\mathscr{R}_j$ in (1.2) [Oakes (1981)] and, under suitable conditions, may be treated like an ordinary likelihood [Goldstein and Langholz (1992)]. Recently, Langholz and Borgan (1995) have proposed an extension of the nested case-control design, called counter-matching, where the control sampling is performed within sampling strata.

In this paper, we introduce a large class of cohort sampling designs further generalizing the nested case-control and counter-matching designs. Moreover, we develop methods for the analysis of these designs which parallel those available for full cohort data. Estimation of the regression parameters for a given design is surprisingly simple; it is based on maximizing a partial likelihood, defined precisely in Section 3, of the form

$$(1.3)\qquad \mathscr{L}(\beta) = \prod_{t_j} \left\{ \frac{\exp\left(\beta^T \mathbf{Z}_{i_j}(t_j)\right) w_{i_j}\left(t_j, \tilde{\mathscr{R}}_j\right)}{\sum_{l \in \tilde{\mathscr{R}}_j} \exp\left(\beta^T \mathbf{Z}_l(t_j)\right) w_l\left(t_j, \tilde{\mathscr{R}}_j\right)} \right\}.$$

Here $\tilde{\mathscr{R}}_j$ is the sampled risk set at t_j, and the $w_l(t_j, \tilde{\mathscr{R}}_j)$ are weights which depend on the sampling design. In Section 3, we indicate why (1.3) may be treated like an ordinary likelihood, while in Section 6 we give formal proofs of consistency and asymptotic normality of the maximum partial likelihood estimator $\hat{\beta}$ obtained by maximizing (1.3). Further, estimation of the cumulative baseline hazard is a simple generalization of the Breslow estimator for full cohort data. The estimator for nested case-control sampling was given by Borgan and Langholz (1993). In Section 4, we extend this to our general sampling designs, while in Section 6 we show that the estimator converges weakly to a Gaussian process.

The key point to the development of these methods is to use a marked point process [e.g., Brémaud (1981), Karr (1991)] to model simultaneously the occurrence of failures and the sampling of controls at each failure time. This model construction is presented in Section 2 together with related counting and intensity processes.

Applications of the general methodology to specific sampling designs are given in Section 5. Along with the full cohort data, nested case-control sampling and counter-matching, there are three new, and potentially quite useful, sampling designs: counter-matching with additional randomly sampled controls; quota sampling of controls; and nested case-control sampling with number of controls dependent on the failure's exposure status. Case-cohort sampling [Prentice (1986)] also belongs to this class and is given as an example where the maximum partial likelihood estimator is clearly inefficient. In Section 7, we apply the general asymptotic results of Section 6 to four of the designs presented in Section 5, showing how our conditions are satisfied and giving the asymptotic variance formulas.

Throughout the paper we will, without further reference, use standard results from the theory of multivariate counting processes, local square integrable martingales and stochastic integrals as surveyed, for example, by Fleming and Harrington [(1991), Chapters 1 and 2] and Andersen, Borgan, Gill and Keiding [(1993), Sections II.2–II.4]. We will only consider marked point processes with a finite mark space, so we do not, however, need results on marked point processes beyond those surveyed by Arjas [(1989), Sections 2 and 4] and Andersen, Borgan, Gill and Keiding [(1993), Sections II.4 and II.7].

2. A model for sampled cohort data. We fix throughout the paper a time interval $[0, \tau]$ for a given terminal time τ, $0 < \tau \le \infty$, and assume that the cohort consists of n individuals. First we specify a model for the marked point process $\{(t_j, i_j);\ j \ge 1\}$ recording the times t_j when failures occur and the individuals i_j which fail at these time points, without consideration of the sampling of controls. To this end we assume that the marked point process $\{(t_j, i_j);\ j \geq 1\}$ is defined on a probability space $(\Omega, \mathscr{F}, \mathbf{P})$ and that it is adapted to the filtration $(\mathscr{H}_t)$ generated by the observed events (failures, censorings, etc.) in the cohort [Andersen, Borgan, Gill and Keiding (1993), Section III]. Associated with this marked point process we have the counting processes $N_i(t) = \sum_{j \geq 1} I(t_j \leq t,\ i_j = i)$ counting the number of observed failures for individual i in $[0, t]$, $i = 1, 2, \dots, n$, with intensity processes given by (1.1). A fundamental assumption throughout the paper is that the indicator processes Y_i and the covariate processes $\mathbf{Z}_i$ are left-continuous and adapted; consequently they are predictable and locally bounded.

Now that a model for the failures has been specified, we turn to describe how the sampling of controls may be superimposed onto this model. This is done by sampling at each failure time, according to a distribution $\pi_t(\mathbf{r}|i)$ to be specified below, a set of controls for the failing individual. We let $\tilde{\mathscr{R}}_j$ denote the sampled risk set consisting of these controls together with the individual i_j failing at t_j. Then

$$\left\{\left(t_j, \left(i_j, \tilde{\mathscr{R}}_j\right)\right); j \geq 1\right\} \tag{2.1}$$

will be a marked point process with a finite mark space E which may be specified as follows: Let $\mathscr{P}$ be the power set of $\{1, 2, \dots, n\}$, that is, the set of all subsets of $\{1, 2, \dots, n\}$, and let $\mathscr{P}_i = \{\mathbf{r}: \mathbf{r} \in \mathscr{P},\ i \in \mathbf{r}\}$. Then the mark space of (2.1) is given as

$$E = \left\{(i, \mathbf{r}): i \in \{1, 2, \dots, n\}, \mathbf{r} \in \mathscr{P}_i\right\} = \left\{(i, \mathbf{r}): \mathbf{r} \in \mathscr{P}, i \in \mathbf{r}\right\}.$$

The introduction of the sampling into the model will bring in some extra random variation, so the marked point process (2.1) will not be adapted to the filtration $(\mathscr{H}_t)$ generated by the available data from the cohort. Thus we now have to work with the enlarged family of sub-σ-algebras $(\mathscr{F}_t)_{t \in [0, \tau]}$ of $\mathscr{F}$ given by $\mathscr{F}_t = \mathscr{H}_t \vee \sigma\{\tilde{\mathscr{R}}_j;\ t_j \leq t\}$, that is, $(\mathscr{F}_t)$ is generated by the observed events in the cohort together with the sampled risk sets.

Corresponding to the marked point process (2.1) we now have, for each $(i, \mathbf{r}) \in E$, the counting process

$$N_{(i, \mathbf{r})}(t) = \sum_{j \geq 1} I\left(t_j \leq t, \left(i_j, \tilde{\mathscr{R}}_j\right) = (i, \mathbf{r})\right) \tag{2.2}$$

counting the observed number of failures for the ith individual in $[0, t]$ with associated sampled risk set $\mathbf{r}$. Since the mark space E is finite, the marked point process (2.1) is, in fact, equivalent to the multivariate counting process $(N_{(i, \mathbf{r})};\ (i, \mathbf{r}) \in E)$. We denote the intensity process of $N_{(i, \mathbf{r})}$ by $\lambda_{(i, \mathbf{r})}$. From (2.2) we may recover the counting process N_i, registering the observed failures for the ith individual, by $N_i = \sum_{\mathbf{r} \in \mathscr{P}_i} N_{(i, \mathbf{r})}$, and a similar relation holds for the intensity processes λ_i and $\lambda_{(i, \mathbf{r})}$.

The fact that we have to augment the original filtration may have the consequence that the $(\mathcal{F}_t)$-intensity processes of the N_i may differ from their $(\mathcal{H}_t)$-intensity processes (1.1). For instance, in a prevention trial, this will be the case if individuals selected as controls change their behavior in such a way that their risk of failure is different from similar individuals who have not been selected as controls. To rule out such possibilities, we introduce the concept of *independent sampling* analogous to the usual assumption of independent censoring [Andersen, Borgan, Gill and Keiding (1993), Section III.2.2]. Formally, we will say that *the sampling is independent provided that the* $(\mathcal{F}_t)$*-intensity processes of the counting processes* N_i *are the same as their* $(\mathcal{H}_t)$*-intensity processes*. Under independent sampling, which will be tacitly assumed below, the $(\mathcal{F}_t)$-intensity processes of the N_i are given by (1.1). In the sequel we will consider intensity processes, martingales, and so on with respect to the filtration $(\mathcal{F}_t)$ and not the "cohort history" $(\mathcal{H}_t)$.

Then, given $\pi_t(\mathbf{r}|i)$, the conditional probability of selecting the sampled risk set $\mathbf{r} \in \mathcal{P}_i$ at time t given $\mathcal{F}_{t-}$ and the fact that the ith individual fails at t, the intensity processes $\lambda_{(i,\mathbf{r})}$ for the counting processes (2.2) is given by $\lambda_{(i,\mathbf{r})}(t) = \lambda_i(t)\pi_t(\mathbf{r}|i)$. Thus, by (1.1),

$$\lambda_{(i,\mathbf{r})}(t) = Y_i(t)\alpha_0(t)\exp\big(\beta_0^T\mathbf{Z}_i(t)\big)\pi_t(\mathbf{r}|i), \tag{2.3}$$

and it follows that a model for cohort sampling is given by specifying, for each t and each i with $Y_i(t) = 1$, the sampling distributions $\pi_t(\cdot|i)$ over sets $\mathbf{r}$ in $\mathcal{P}_i$. [For notational convenience we set $\pi_t(\mathbf{r}|i) = 0$ if $Y_i(t) = 0$.] This specification must be based on information available just before time t; formally, considered as processes in t, the $\pi_t(\mathbf{r}|i)$ are throughout the paper assumed to be left-continuous and adapted (and hence predictable and locally bounded). In particular this rules out selection of controls depending on events in the future, for example, one may not exclude as potential controls for a current case individuals that *subsequently* fail [Lubin and Gail (1984)]. We will give examples of specific sampling designs in Section 5.

For ease of presentation, we have previously assumed that censoring is a part of the cohort history $(\mathcal{H}_t)$ only, and that no extra censoring is introduced by the sampling of controls. This may easily be extended, however, along the lines of Andersen, Borgan, Gill and Keiding [(1993), Section III.2] to include extra (independent) censoring depending on the previous sampling history. For example, one may censor individuals after they have been picked as controls.

3. A partial likelihood and estimation of the regression parameters. For cohort data, inference on β_0 is based on the partial likelihood (1.2). Here we will use the general ideas of Arjas [(1989), Section 4] to derive a similar partial likelihood for the Cox model based on sampled cohort data. To this end we first introduce the reduced marked point process $\{(t_j, \tilde{\mathcal{R}}_j);\ j \geq 1\}$ derived from (2.1) by disregarding the information about which individuals fail at the various time points. Corresponding to this marked point process we have the counting processes $N_{\mathbf{r}}(t) = \sum_{i \in \mathbf{r}} N_{(i,\mathbf{r})}(t)$ counting the number of times the sampled risk set equals $\mathbf{r}$ in $[0, t]$. By (2.3) these have intensity processes

$$\lambda_{\mathbf{r}}(t) = \sum_{i \in \mathbf{r}} \lambda_{(i,\mathbf{r})}(t) = \sum_{i \in \mathbf{r}} Y_i(t)\alpha_0(t)\exp(\beta_0^T \mathbf{Z}_i(t))\pi_t(\mathbf{r}|i). \tag{3.1}$$

We then factorize the intensity processes $\lambda_{(i,\mathbf{r})}$, not as in (2.3) but as $\lambda_{(i,\mathbf{r})}(t) = \lambda_{\mathbf{r}}(t)\pi_t(i|\mathbf{r}; \beta_0)$, where

$$\pi_t(i|\mathbf{r}; \beta_0) = \frac{\lambda_{(i,\mathbf{r})}(t)}{\lambda_{\mathbf{r}}(t)} = \frac{Y_i(t)\exp(\beta_0^T \mathbf{Z}_i(t))\pi_t(\mathbf{r}|i)}{\sum_{l \in \mathbf{r}} Y_l(t)\exp(\beta_0^T \mathbf{Z}_l(t))\pi_t(\mathbf{r}|l)} \tag{3.2}$$

is the conditional probability of the ith individual failing at t, given $\mathscr{F}_{t-}$ and that there is a failure among individuals in the set $\mathbf{r}$ at t.

Statistical inference on the regression parameters may therefore be based on the partial likelihood

$$\begin{aligned} \mathscr{L}_\tau(\beta) &= \prod_{t_j} \pi_{t_j}(i_j|\tilde{\mathscr{R}}_j; \beta) \\ &= \prod_{t_j} \left\{ \frac{Y_{i_j}(t_j)\exp(\beta^T \mathbf{Z}_{i_j}(t_j))\pi_{t_j}(\tilde{\mathscr{R}}_j|i_j)}{\sum_{l \in \tilde{\mathscr{R}}_j} Y_l(t_j)\exp(\beta^T \mathbf{Z}_l(t_j))\pi_{t_j}(\tilde{\mathscr{R}}_j|l)} \right\} \\ &= \prod_{u \in [0,\tau]} \prod_{\mathbf{r} \in \mathscr{P}} \prod_{i \in \mathbf{r}} \left\{ \frac{Y_i(u)\exp(\beta^T \mathbf{Z}_i(u))\pi_u(\mathbf{r}|i)}{\sum_{l \in \mathbf{r}} Y_l(u)\exp(\beta^T \mathbf{Z}_l(u))\pi_u(\mathbf{r}|l)} \right\}^{\Delta N_{(i,\mathbf{r})}(u)} \end{aligned} \tag{3.3}$$

obtained by only using the information contained in the conditional distributions of the failing individuals i_j given the sampled risk sets $\tilde{\mathscr{R}}_j$, and thereby disregarding the information on β contained in the reduced marked point process $\{(t_j, \tilde{\mathscr{R}}_j);\ j \geq 1\}$. This generalizes the partial likelihood of Oakes (1981) for nested case-control designs with simple random sampling of the controls (Example 2). Note that in the denominator of (3.3) each individual is weighted with the probability of selecting the sampled risk set had the individual been the failure. In particular, an observed failure is weighted no differently than the controls.

The estimator $\hat{\beta}$ obtained by maximizing the partial likelihood (3.3) has similar properties as a maximum likelihood estimator. At this point we will only show that (3.3) has "basic likelihood properties," that is, that the score vector has expectation zero and that its covariance matrix equals the expected information matrix (implicitly assuming the necessary regularity conditions to hold) and return to a detailed study in Section 6. The notation and derivations are similar to those for the full cohort data reviewed by Andersen, Borgan, Gill and Keiding [(1993), Section VII.2.1].

We introduce

(3.4) $$S_{\mathbf{r}}^{(\gamma)}(\beta,t) = \sum_{i\in\mathbf{r}} Y_i(t)\mathbf{Z}_i(t)^{\otimes\gamma}\exp(\beta^T\mathbf{Z}_i(t))\pi_t(\mathbf{r}|i), \quad \gamma=0,1,2,$$

(3.5) $$\mathbf{E}_{\mathbf{r}}(\beta,t) = \frac{S_{\mathbf{r}}^{(1)}(\beta,t)}{S_{\mathbf{r}}^{(0)}(\beta,t)},$$

(3.6) $$\mathbf{V}_{\mathbf{r}}(\beta,t) = \frac{S_{\mathbf{r}}^{(2)}(\beta,t)}{S_{\mathbf{r}}^{(0)}(\beta,t)} - \mathbf{E}_{\mathbf{r}}(\beta,t)^{\otimes 2},$$

where, for a vector $\mathbf{a}$, $\mathbf{a}^{\otimes 0}=1$, $\mathbf{a}^{\otimes 1}=\mathbf{a}$ and $\mathbf{a}^{\otimes 2}=\mathbf{a}\mathbf{a}^T$. Note that the two quantities $\mathbf{E}_{\mathbf{r}}(\beta,t)$ and $\mathbf{V}_{\mathbf{r}}(\beta,t)$ are the expectation and the covariance matrix, respectively, of the covariate vector $\mathbf{Z}_i(t)$ if an individual is selected with probability $\pi_t(i|\mathbf{r};\beta)$ [cf. (3.2)].

Then the vector of score functions may be written as

(3.7) $$\mathbf{U}_\tau(\beta) = \frac{\partial}{\partial\beta}\log\mathscr{L}_\tau(\beta) = \int_0^\tau \sum_{\mathbf{r}\in\mathscr{P}}\sum_{i\in\mathbf{r}}\{\mathbf{Z}_i(u)-\mathbf{E}_{\mathbf{r}}(\beta,u)\}\,dN_{(i,\mathbf{r})}(u),$$

while the observed information matrix takes the form

(3.8) $$\mathscr{I}_\tau(\beta) = -\frac{\partial^2}{\partial\beta^2}\log\mathscr{L}_\tau(\beta) = \int_0^\tau \sum_{\mathbf{r}\in\mathscr{P}} \mathbf{V}_{\mathbf{r}}(\beta,u)\,dN_r(u).$$

By the interpretation of $\mathbf{V}_{\mathbf{r}}(\beta,t)$ as a covariance matrix given just below (3.6), it follows that $\mathscr{I}_\tau(\beta)$ is positive semidefinite, and hence that the log partial likelihood is concave.

By standard counting process theory $M_{(i,\mathbf{r})}(t) = N_{(i,\mathbf{r})}(t) - \int_0^t \lambda_{(i,\mathbf{r})}(u)\,du$, $(i,\mathbf{r})\in E$, are orthogonal local square integrable martingales. It follows that the score, evaluated at the true parameter value β_0, equal the (vector-valued) stochastic integral

(3.9) $$\mathbf{U}_\tau(\beta_0) = \int_0^\tau \sum_{\mathbf{r}\in\mathscr{P}}\sum_{i\in\mathbf{r}}\{\mathbf{Z}_i(u)-\mathbf{E}_{\mathbf{r}}(\beta_0,u)\}\,dM_{(i,\mathbf{r})}(u).$$

In particular, the expected score is zero. We let $\mathbf{U}_t(\beta_0)$ be defined by (3.9), but with the integral taken over $[0,t]$ instead of $[0,\tau]$, and we note that $\mathbf{U}.(\beta_0)$ is a (vector-valued) local square integrable martingale. The matrix of predictable covariation processes of this martingale, evaluated at τ, becomes

(3.10) $$\langle\mathbf{U}.(\beta_0)\rangle(\tau) = \int_0^\tau \sum_{\mathbf{r}\in\mathscr{P}} \mathbf{V}_{\mathbf{r}}(\beta_0,u)S_{\mathbf{r}}^{(0)}(\beta_0,u)\alpha_0(u)\,du.$$

Moreover, the observed information matrix, evaluated at β_0, may be written as

(3.11) $$\mathscr{I}_\tau(\beta_0) = \langle\mathbf{U}.(\beta_0)\rangle(\tau) + \int_0^\tau \sum_{\mathbf{r}\in\mathscr{P}} \mathbf{V}_{\mathbf{r}}(\beta_0,u)\,dM_{\mathbf{r}}(u),$$

with $M_{\mathbf{r}}(t) = N_{\mathbf{r}}(t) - \int_0^t \lambda_{\mathbf{r}}(u)\,du$, $\mathbf{r} \in \mathscr{P}$, being orthogonal local square integrable martingales. Thus the observed information matrix equals the predictable variation of the score plus a local square integrable martingale. In particular, by taking expectations, it follows that the expected information matrix equals the covariance matrix of the score.

4. Estimation of the integrated baseline hazard. We then turn to the problem of estimating the integrated baseline hazard $A_0(t) = \int_0^t \alpha_0(t)\,dt$. To this end we first factorize $\pi_t(\cdot|i)$, for each t and each i with $Y_i(t) = 1$, as

$$\pi_t(\mathbf{r}|i) = \pi_t(\mathbf{r})w_i(t,\mathbf{r}), \tag{4.1}$$

where

$$\pi_t(\mathbf{r}) = n(t)^{-1}\sum_{l\in\mathbf{r}}\pi_t(\mathbf{r}|l) \quad\text{and}\quad w_i(t,\mathbf{r}) = \frac{\pi_t(\mathbf{r}|i)}{n(t)^{-1}\sum_{l\in\mathbf{r}}\pi_t(\mathbf{r}|l)}, \tag{4.2}$$

with

$$n(t) = \sum_{\mathbf{r}\in\mathscr{P}}\sum_{l\in\mathbf{r}}\pi_t(\mathbf{r}|l) = \sum_{l=1}^{n}\sum_{\mathbf{r}\in\mathscr{P}_l}\pi_t(\mathbf{r}|l) = \sum_{l=1}^{n}Y_l(t) \tag{4.3}$$

being the number at risk at $t-$. Note that $\pi_t(\cdot)$ is a sampling distribution over sets $\mathbf{r}$ in $\mathscr{P}$ defined for each t for which there is at least one individual at risk.

We then suggest the estimator

$$\begin{aligned}\hat{A}_0(t;\hat{\beta}) &= \sum_{t_j\le t}\frac{1}{\sum_{l\in\tilde{\mathscr{R}}_j}Y_l(t_j)\exp\big(\hat{\beta}^T\mathbf{Z}_l(t_j)\big)w_l\big(t_j,\tilde{\mathscr{R}}_j\big)}\\ &= \int_0^t\sum_{\mathbf{r}\in\mathscr{P}}\frac{dN_{\mathbf{r}}(u)}{\sum_{l\in\mathbf{r}}Y_l(u)\exp\big(\hat{\beta}^T\mathbf{Z}_l(u)\big)w_l(u,\mathbf{r})}\end{aligned} \tag{4.4}$$

for the integrated baseline hazard. This estimator may be motivated by the same "method of moments" argument as for the full cohort Nelson–Aalen estimator [Andersen, Borgan, Gill and Keiding (1993), Section IV.1.1]. To be specific, let $J(t) = I(n(t) > 0)$ be the predictable indicator process which equals 1 if someone is at risk at $t-$ and equals 0 otherwise, and interpret $0/0$ as 0 and $\pi_t(\mathbf{r})$ as 0 when $J(t) = 0$. Then use (3.1) and (4.1) to find that (4.4), with $\hat{\beta}$ replaced by the true value β_0, may be written as

$$\begin{aligned}\hat{A}_0(t;\beta_0) &= \int_0^t\sum_{\mathbf{r}\in\mathscr{P}}\frac{J(u)\,dN_{\mathbf{r}}(u)}{\sum_{l\in\mathbf{r}}Y_l(u)\exp\big(\beta_0^T\mathbf{Z}_l(u)\big)w_l(u,\mathbf{r})}\\ &= \int_0^t\sum_{\mathbf{r}\in\mathscr{P}}\alpha_0(u)\pi_u(\mathbf{r})\,du + \hat{W}(t)\\ &= \int_0^t J(u)\,\alpha_0(u)\,du + \hat{W}(t),\end{aligned}$$

where

$$\hat{W}(t) = \int_0^t \sum_{\mathbf{r}\in\mathscr{P}} \frac{J(u)\, dM_{\mathbf{r}}(u)}{\sum_{l\in\mathbf{r}} Y_l(u)\exp(\beta_0^T \mathbf{Z}_l(u)) w_l(u,\mathbf{r})} \tag{4.5}$$

is a local square integrable martingale. This shows that $\hat{A}_0(t;\beta_0)$ is almost unbiased for $A_0(t)$ (the small bias only being due to the possibility of having no one at risk), thereby giving a justification for the proposed estimator (4.4).

Using Theorems 2 and 3 and Proposition 2 of Section 6, one may copy the arguments of Andersen, Borgan, Gill and Keiding [(1993), Corollary VII.2.4] to get that $\sqrt{n}(\hat{A}_0(\cdot;\hat{\beta}) - A_0(\cdot))$ converges weakly, as $n \to \infty$, to a mean-zero Gaussian process. The covariance function of this limiting process may be estimated uniformly consistently by $n\hat{\sigma}^2(s,t)$, where

$$\hat{\sigma}^2(s,t) = \hat{\omega}^2(s\wedge t;\hat{\beta}) + \hat{\mathbf{B}}(s;\hat{\beta})^T \mathscr{I}(\hat{\beta})^{-1}\hat{\mathbf{B}}(t;\hat{\beta}), \tag{4.6}$$

with

$$\hat{\omega}^2(t;\beta) = \sum_{t_j\le t} \frac{1}{\left\{\sum_{l\in\tilde{\mathscr{R}}_j} Y_l(t_j)\exp(\beta^T\mathbf{Z}_l(t_j)) w_l(t_j,\tilde{\mathscr{R}}_j)\right\}^2} \tag{4.7}$$

and

$$\hat{\mathbf{B}}(t;\beta) = \sum_{t_j\le t} \frac{\sum_{l\in\tilde{\mathscr{R}}_j} Y_l(t_j)\mathbf{Z}_l(t_j)\exp(\beta^T\mathbf{Z}_l(t_j)) w_l(t_j,\tilde{\mathscr{R}}_j)}{\left\{\sum_{l\in\tilde{\mathscr{R}}_j} Y_l(t_j)\exp(\beta^T\mathbf{Z}_l(t_j)) w_l(t_j,\tilde{\mathscr{R}}_j)\right\}^2}. \tag{4.8}$$

Moreover, by the arguments in Andersen, Borgan, Gill and Keiding [(1993), Corollary VII.2.6 and Section VII.2.3], one may also derive the asymptotic distribution of the integrated hazard and the survival function for an individual with a specified vector of convariates $\mathbf{Z}_0$ fixed over time.

5. Examples of specific sampling schemes. The methodology we have presented in the preceding sections provides analytic tools for a very large class of sampling schemes. We illustrate its flexibility by deriving the partial likelihood for a few diverse designs. In each situation considered, the sometimes complex $\pi_t(\mathbf{r}|i)$ in the partial likelihood (3.3) are not needed for the actual analysis of the data. Factorization as in (4.1) and cancellation of common terms lead to considerable reduction yielding simple weights $w_i(t,\mathbf{r})$ as in (1.3). The same weights will (typically) also be used in the estimator (4.4) for the integrated baseline hazard and its variance (4.6).

We note that, while the partial likelihood provides a method of valid estimation of the regression parameters, there is no guarantee that the estimator $\hat{\beta}$ will be efficient relative to "the best" method of analysis. This is illustrated in Example 7 and further discussed in Section 8. We let $|\mathbf{r}|$ denote the number of elements in the set $\mathbf{r}$.

EXAMPLE 1 (Full cohort). The full cohort partial likelihood is a special case in which the entire risk set $\mathscr{R}(t) = \{i: Y_i(t) = 1\}$ is sampled with probability 1.

In our notation then, $\pi_t(\mathbf{r}|i) = I(\mathbf{r} = \mathscr{R}(t))$ for all $i \in \mathscr{R}(t)$, and the usual Cox partial likelihood for the full data set is recovered. Noting that (4.1) and (4.2) are fulfilled with $\pi_t(\mathscr{R}(t)) \equiv w_i(t, \mathscr{R}(t)) \equiv 1$, (4.4) reduces to the usual Breslow estimator of the integrated baseline hazard function.

EXAMPLE 2 (Nested case-control sampling). The most common type of cohort sampling technique is nested case-control sampling, in which $m - 1$ controls are randomly sampled, without replacement from those at risk at the failure's failure time. Here we assume that $m > 1$ is fixed. Letting $n(t) = \sum_{i=1}^{n} Y(t) = |\mathscr{R}(t)|$ denote the number at risk at time t, this sampling scheme is specified by

$$\pi_t(\mathbf{r}|i) = \binom{n(t) - 1}{m - 1}^{-1} I(i \in \mathbf{r}, \mathbf{r} \subset \mathscr{R}(t), |\mathbf{r}| = m),$$

which is the same for each $i \in \mathbf{r}$ and thus drops out of (3.3) leaving the usual partial likelihood [Oakes (1981)] for nested case-control sampling. Further, (4.1) and (4.2) are satisfied with

$$\pi_t(\mathbf{r}) = \binom{n(t)}{m}^{-1} I(\mathbf{r} \subset \mathscr{R}(t), |\mathbf{r}| = m) \quad \text{and} \quad w_i(t, \mathbf{r}) = \frac{n(t)}{m},$$

and, from (4.4) an estimator of the cumulative hazard function is

$$\hat{A}_0(t; \hat{\beta}) = \sum_{t_j \le t} \frac{1}{\sum_{l \in \tilde{\mathscr{R}}_j} \exp\left(\hat{\beta}^T \mathbf{Z}_l(t_j)\right) n(t_j)/m},$$

with variance estimator from (4.6) [cf. Borgan and Langholz (1993), equations (7) and (10)].

EXAMPLE 3 (Counter-matching). In this extension of nested case-control sampling [Langholz and Borgan (1995)], control sampling is performed within sampling strata. In general, let $C_i(t)$ be $(\mathscr{F}_t)$-predictable sampling strata indicators with $C_i(t) \in \mathscr{C}$, a (small) finite set of indices. Define $\mathscr{R}_l(t) = \{i: Y_i(t) = 1, C_i(t) = l\}$ to be sampling stratum l with $n_l(t) = |\mathscr{R}_l(t)|$. If a subject (say, i) fails at time t, then m_l controls are randomly sampled without replacement from $\mathscr{R}_l(t)$ except for the failure's stratum $\mathscr{R}_{C_i(t)}(t)$, from which $m_{C_i(t)} - 1$ are sampled from the $n_{C_i(t)}(t) - 1$ nonfailures. The probability of picking a given set depends on the sampling stratum of the case and is given by

$$\pi_t(\mathbf{r}|i) = \left[\prod_{l \in \mathscr{C}} \binom{n_l(t)}{m_l}\right]^{-1} \frac{n_{C_i(t)}(t)}{m_{C_i(t)}}$$
$$\times I\left(i \in \mathbf{r}, \mathbf{r} \subset \mathscr{R}(t), \left|\mathbf{r} \cap \mathscr{R}_l(t)\right| = m_l; l \in \mathscr{C}\right).$$

These sampling probabilities satisfy (4.1) and (4.2) with

$$\pi_t(\mathbf{r}) = \left[\prod_{l\in\mathscr{C}} \binom{n_l(t)}{m_l}\right]^{-1} I(\mathbf{r} \subset \mathscr{R}(t), |\mathbf{r} \cap \mathscr{R}_l(t)| = m_l; l \in \mathscr{C})$$

and

$$w_i(t,\mathbf{r}) = \frac{n_{C_i(t)}(t)}{m_{C_i(t)}}.$$

Thus, in the partial likelihood (1.3), the relative risk for a subject from a given stratum is weighted by the inverse of the proportion of the stratum sampled. Moreover, the same weights are used in the cumulative baseline hazard estimator (4.4) and its variance estimator (4.6).

Because $\mathscr{F}_t$ contains failure, censoring, covariate and sampling histories up to time t, the sampling strata may be defined in some quite diverse ways. As discussed in Langholz and Borgan (1955), the $C_i(t)$ can be based on factors specific to subject i, such as absolute exposure level, or relative to the distribution of exposures, such as grouping based on some empirical quantile levels of exposure. Another interesting possibility is to base the counter-matching on the sampling history. As a specific example, which ensures that each sample risk set adds m_0 new subjects to the sample, one may sample m_0 from the set of those not sampled in any previous risk set (stratum 0) and m_1 from those who have been sampled (stratum 1).

EXAMPLE 4 (Quota sampling: negative hypergeometric sampling). Consider processes $C_i(t) \in \mathscr{C} = \{0,1\}$ which partition the risk set into "target" and "nontarget" subjects indicated by $C_i(t) = 1$ or 0, respectively. Let $\mathscr{R}_1(t) = \{i\colon Y_i(t) = 1, C_i(t) = 1\}$ and $n_1(t) = |\mathscr{R}_1(t)|$. In this sampling method, if a subject i fails at time t, controls are sampled sequentially until a predetermined "quota" of m_1 subjects are selected from $\mathscr{R}_1(t)$. As before, a failure in the target group is counted as one of the m_1.

The probability of sampling a particular set depends on the size of the set and $C_i(t)$ for the failure. Specifically, the size of the sampled risk set has a negative hypergeometric distribution [Schuster and Sype (1987)], while the probability of choosing a particular $\mathbf{r}$ given a fixed size is as in the counter-matching example. Simplification leads to

$$\pi_t(\mathbf{r}|i) = \frac{m_1 - C_i(t)}{\binom{n(t)-1}{|\mathbf{r}|-2}} \times \frac{1}{n(t) - |\mathbf{r}| + 1} I(i \in \mathbf{r}, \mathbf{r} \subset \mathscr{R}(t), |\mathbf{r} \cap \mathscr{R}_1(t)| = m_1).$$

It is seen that (4.1) and (4.2) are fulfilled with

$$\pi_t(\mathbf{r}) = \frac{m_1}{\binom{n(t)}{|\mathbf{r}|-1}} \frac{1}{n(t) - |\mathbf{r}| + 1} I(\mathbf{r} \subset \mathscr{R}(t), |\mathbf{r} \cap \mathscr{R}_1(t)| = m_1),$$

and weights

$$w_i(t, \mathbf{r}) = \frac{m_1 - C_i(t)}{m_1} \frac{n(t)}{|\mathbf{r}| - 1}.$$

These weights would then be included in the partial likelihood (1.3), as well as in (4.4) and (4.6), to construct estimators of the cumulative baseline hazard and its variance.

Note that if $m_1 = 1$, it is not possible to estimate the regression parameters since all target subjects, that is, those with $C_i(t) = 1$, are weighted by zero. This is because if the failure is a target subject, the sampled risk set consists only of that failure, making estimation impossible. One possible solution is to (simple) randomly sample one control, without regard to the target status of the failure, before starting the quota sampling. This would assure that target failures are almost always matched to a nontarget control (assuming that the target group is rare) and that there would be one (or rarely two) target controls for a nontarget failure.

EXAMPLE 5 (Counter-matching with additional randomly sampled controls). This is a hybrid design in which the controls consist of some which are counter-matched and others which are a simple random sample. Specifically, consider first counter-matching $\overline{m}$ individuals (Example 3) with $\overline{m}_l$ from sampling stratum l, $\overline{m} = \sum_{l \in \mathscr{C}} \overline{m}_l$, and then randomly sampling (Example 2) $\tilde{m}$ from the remaining $n(t) - \overline{m}$ subjects in the risk set. Let $\tilde{m}_l$ be the number of the randomly sampled controls which happen to be picked from $\mathscr{R}_l(t)$. The probability of picking a particular counter-matched set and particular randomly sampled set is the product of probabilities of the forms given in the previous examples. This is then multiplied by the number of such combinations which would yield a given set $\mathbf{r}$ which, after some simplification, becomes

$$\pi_t(\mathbf{r}|i) = \left[\frac{\prod_l \binom{n_l(t) - \overline{m}_l}{\tilde{m}_l}}{\binom{n(t) - \overline{m}}{\tilde{m}}} \right] \left[\prod_l \binom{n_l(t)}{\overline{m}_l + \tilde{m}_l} \right]^{-1}$$

$$\times \frac{n_{C_i(t)}(t)}{\overline{m}_{C_i(t)} + \tilde{m}_{C_i(t)}} I(i \in \mathbf{r}, \mathbf{r} \subset \mathscr{R}(t), |\mathbf{r} \cap \mathscr{R}_l(t)| = \overline{m}_l + \tilde{m}_l; l \in \mathscr{C}).$$

This yields $w_i(t, \mathbf{r})$ of the same form as for counter-matching with $\overline{m}_l + \tilde{m}_l$ replacing m_l.

EXAMPLE 6 (Nested case-control sampling with variable matching ratio). Instead of fixed m in the nested case-control design of Example 2, consider using a variable matching ratio, possibly depending upon characteristics of the case. Specifically, in this class of sampling designs, first the size of the sampled risk set is (randomly) determined and then a nested case-control sample of this size is selected. Let $m(t)$ be the size of the sampled risk set if there is a failure at time t. We assume that $m(t)$ is random with a (predictable) probability distribution on $\{1, \dots, n(t)\}$ which may depend on who failed at that time. We may then specify the sampling probabilities as functions of the size of $\mathbf{r}$ with

$$\pi_t(\mathbf{r}|i) = \binom{n(t)-1}{|\mathbf{r}|-1}^{-1} I(i \in \mathbf{r}, \mathbf{r} \subset \mathscr{R}(t))\mathbb{P}(m(t) = |\mathbf{r}|\,|i, \mathscr{F}_{t-}).$$

Becomes the binomial coefficient is common to all $i \in \mathbf{r}$, $w_i(t, \mathbf{r})$ involves only the $\mathbb{P}(m(t) = |\mathbf{r}|\,|i, \mathscr{F}_{t-})$ and $m(t)$.

EXAMPLE 7 (Case-cohort sampling). Prentice (1986) presents case-cohort sampling in which a subcohort $\tilde{C}$ is randomly sampled from the full cohort at $t = 0$. He shows heuristically that the partial likelihood for this design does not make use of the non-subcohort failures in the estimation of the regression parameters and proposes a "pseudolikelihood" approach. In our formulation, since $\tilde{C} \in \mathscr{F}_0$,

$$\pi_t(\mathbf{r}|i) = I\big(\mathbf{r} = \big(\tilde{C} \cap \mathscr{R}(t)\big) \cup \{i\}\big).$$

Thus, if $i \in \tilde{C}$ fails at t_j, then $\tilde{\mathscr{R}}_j = \tilde{C} \cap \mathscr{R}(t_j)$ and $\pi_t(\tilde{\mathscr{R}}_j|l) = 1$ for all $l \in \tilde{\mathscr{R}}_j$, but if $i \notin \tilde{C}$, $\tilde{\mathscr{R}}_j = (\tilde{C} \cap \mathscr{R}(t_j)) \cup \{i\} \neq \tilde{C} \cap \mathscr{R}(t_j)$ and $\pi_t(\tilde{\mathscr{R}}_j|l) = I(l = i)$ since this sampled risk set would occur with probability zero if a subcohort member failed. Thus, if i is a non-subcohort failure, the partial likelihood weights subcohort members by zero, leaving a contribution of one for that sampled risk set, confirming Prentice's conclusion. This is an example where the partial likelihood (3.3) is clearly inefficient for the design.

We note that in Examples 2–5 the support of the sampling distribution $\pi_t(\mathbf{r})$ can be partitioned into "blocks" of sets so that $\pi_t(\mathbf{r})$ is uniform, conditional on the sampled risk set being chosen from a block of the partition. In counter-matching, Example 3, a case where the partition has only a single block, the sampling is uniform over all sets with m_l individuals of stratum l. For quota sampling, Example 4, the sampling becomes uniform when conditioned on the number sampled; here the blocks are composed of sets of common size. This property becomes very useful in proving asymptotic results for these schemes under iid assumptions. We give a general formulation of this partition structure at the end of Section 6 and use it in Section 7 to study the asymptotic properties of Examples 2–5.

6. Asymptotic properties of the estimators. In order to discuss asymptotic behavior, we need to consider a sequence of models of the form defined in Section 2 with processes $N_{(i,\mathbf{r})}^{(n)}$, $Y_i^{(n)}$, $\mathbf{Z}_i^{(n)}$, $\pi^{(n)}(\mathbf{r}|i)$ and so on in the nth model, $n = 1, 2, \dots$. For ease of notation we will drop the superscript (n), but the reader should keep in mind that these quantities depend on n whereas the true parameters β_0 and α_0 are the same in all models. Further, we remind the reader of definitions (3.4)–(3.6) and write $\mathbf{U}_t(\beta)$ and $\mathscr{I}_t(\beta)$ for (3.7) and (3.8), respectively, when the integral is taken over $[0, t]$ instead of $[0, \tau]$. Finally, we denote the jth component of the vector $\mathbf{U}_t(\beta)$ by $U_t^j(\beta)$, and the (j, k)th element of the matrix $\mathscr{I}_t(\beta)$ by $\mathscr{I}_t^{jk}(\beta)$.

Before we specify the conditions needed in a study of the large-sample properties, let us mention why Andersen–Gill type conditions [Andersen and Gill (1982) Condition B, Page 1105; Andersen, Borgan, Gill and Keiding (1993), Condition VII.2.1.a] cannot be applied in the present situation, and discuss which conditions replace them. As mentioned in Example 1, the full-cohort situation is the special case of our general setup in which the entire risk set $\mathscr{R}(t)$ is sampled with probability 1. For this particular situation, the main condition of Andersen and Gill (1982) is that $n^{-1}S_{\mathscr{R}(t)}^{(\gamma)}(\beta, t)$, for $\gamma = 0, 1, 2$, converge uniformly in probability to corresponding limiting functions $s^{(\gamma)}(\beta, t)$. From this assumption follows the convergence in probability of $n^{-1}\langle \mathbf{U}.(\beta_0)\rangle(\tau)$, which is key in proving weak convergence of the score $\mathbf{U}_\tau(\beta_0)$, and, hence, also of the maximum partial likelihood estimator $\hat{\beta}$.

In most situations of interest to us, the size of the sampled risk sets will not increase with n (e.g., Examples 2–5). Thus we cannot assume uniform convergence in probability of $n^{-1}S_{\mathbf{r}}^{(\gamma)}(\beta, t)$ for each $\mathbf{r}$. However, since the number of possible sampled risk sets increases as n grows, it is quite reasonable to assume that averages over these sets will converge in probability. This is the essential content of Condition 2 which implies convergence in probability of $1/n$ times the integrand of (3.10). In addition we need a boundedness condition (Condition 4) which ensures convergence in probability of the corresponding integral, that is, of $n^{-1}\langle \mathbf{U}.(\beta_0)\rangle(\tau)$. In Section 7 we illustrate how these general conditions are fulfilled for Examples 2–5 under an iid model.

As just indicated, in the proofs we will have to infer convergence in probability of certain integrals based on pointwise convergence in probability of the integrands. For this we will use the following version of the theorem of dominated convergence [Hjort and Pollard (1993)].

PROPOSITION 1. *Suppose $A_0(\tau) < \infty$ and let $0 \le bU_n(s) \le D_n(s)$ be left-continuous random processes on the interval $[0, \tau]$. Suppose $D_n(s) \to_{\mathbb{P}} D(s)$ and $U_n(s) \to_{\mathbb{P}} U(s)$ for almost all s, as $n \to \infty$, and that $\int_0^\tau D_n(s)\alpha_0(s)\,ds \to_{\mathbb{P}} \int_0^\tau D(s)\alpha_0(s)\,ds < \infty$. Then $\int_0^t U_n(s)\alpha_0(s)\,ds \to_{\mathbb{P}} \int_0^t U(s)\alpha_0(s)\,ds$ for all $t \in [0, \tau]$ as $n \to \infty$.*

We are now in a position to formulate our conditions. Here and below the norm of a vector $\mathbf{a} = (a_i)$ or a matrix $\mathbf{A} = \{a_{ij}\}$ is $\|\mathbf{a}\| = \sup_i |a_i|$ and $\|\mathbf{A}\| = \sup_{i,j} |a_{ij}|$, respectively.

CONDITION 1. The hazard $A_0(\tau) < \infty$.

CONDITION 2. For $(\rho, \gamma) = (0,2)$ and $(\rho, \gamma) = (2,0)$ there exist functions $q^{(\rho,\gamma)}$ such that, for all $t \in [0, \tau]$ as $n \to \infty$,

$$(6.1)\quad Q^{(\rho,\gamma)}(\beta_0, t) = \frac{1}{n} \sum_{\mathbf{r} \in \mathscr{P}} \mathbf{E}_{\mathbf{r}}(\beta_0, t)^{\otimes \rho} S_{\mathbf{r}}^{(\gamma)}(\beta_0, t) \to_{\mathbb{P}} q^{(\rho,\gamma)}(\beta_0, t).$$

CONDITION 3. *The* $p \times p$ *matrix* $\mathbf{\Sigma} = \{\sigma_{jk}\}$ *given by*

$$\mathbf{\Sigma} = \int_0^\tau \left[q^{(0,2)}(\beta_0, t) - q^{(2,0)}(\beta_0, t) \right] \alpha_0(t)\, dt$$

is positive definite.

CONDITION 4. For any n and each $\mathbf{r} \in \mathscr{P}$ there exists a locally bounded predictable process $X_{\mathbf{r}}$ such that, for all $t \in [0, \tau]$,

$$(6.2)\qquad \|\mathbf{Z}_i(t)\| \le X_{\mathbf{r}}(t) \quad \text{for all } i \in \mathbf{r}.$$

Moreover, there exist a $b_0 > 3\|\beta_0\|$ and a function D such that, with $\pi_t(\mathbf{r})$ defined by (4.2),

$$(6.3)\qquad D_n(t) = \sum_{\mathbf{r} \in \mathscr{P}} \pi_t(\mathbf{r}) \exp(b_0 X_{\mathbf{r}}(t)) \to_{\mathbb{P}} D(t),$$

for all $t \in [0, \tau]$ as $n \to \infty$, and

$$(6.4)\qquad \int_0^\tau D_n(t)\alpha_0(t)\, dt \to_{\mathbb{P}} \int_0^\tau D(t)\alpha_0(t)\, dt < \infty.$$

For the special case of bounded covariates, $D_n(t)$ in Condition 4 may be chosen as a constant independent of n. Thus (6.3) and (6.4) are trivially fulfilled in this situation. We have chosen, however, to formulate Condition 4 such that it also covers, for example, normally distributed covariates (cf. Section 7).

Before we derive the asymptotic properties of the maximum partial likelihood estimator $\hat{\beta}$, we will show how Conditions 1–4 imply convergence in probability of $n^{-1}\langle \mathbf{U}_{\cdot}(\beta_0)\rangle(\tau)$ to the matrix $\mathbf{\Sigma}$ defined in Condition 3; we will also state some other consequences of the conditions which will be useful in the proofs.

Let $\mathscr{B}_0$ be an open neighborhood of β_0 with $\sup\{\|\beta\|: \beta \in \mathscr{B}_0\} \le b_0/3$. Then by (3.4), (4.2) and (6.2) we have, for $\beta \in \mathscr{B}_0$,

$$(6.5)\qquad \begin{aligned} &n(t)\pi_t(\mathbf{r})\exp\left(-\left(\frac{b_0}{3}\right)X_{\mathbf{r}}(t)\right) \\ &\quad \le S_{\mathbf{r}}^{(0)}(\beta, t) \le n(t)\pi_t(\mathbf{r}) \exp\left(\left(\frac{b_0}{3}\right)X_{\mathbf{r}}(t)\right). \end{aligned}$$

Moreover, for $\gamma = 1, 2$,

$$\left\|S_{\mathbf{r}}^{(\gamma)}(\beta, t)\right\| \le X_{\mathbf{r}}(t)^{\gamma} S_{\mathbf{r}}^{(0)}(\beta, t),$$

and, by (3.5), (3.6) and (6.1),

$$\|\mathbf{E}_{\mathbf{r}}(\beta, t)\| \le X_{\mathbf{r}}(t), \tag{6.6}$$

$$\|\mathbf{V}_{\mathbf{r}}(\beta, t)\| \le X_{\mathbf{r}}(t)^2 \tag{6.7}$$

and

$$\|Q^{(\rho,\gamma)}(\beta_0, t)\| \le \frac{1}{n} \sum_{\mathbf{r}\in\mathscr{P}} X_{\mathbf{r}}(t)^{\rho+\gamma} S_{\mathbf{r}}^{(0)}(\beta_0, t). \tag{6.8}$$

Further, by (6.5), the right-hand side of (6.8) is bounded by a constant times $D_n(t)$, and it follows by Proposition 1 and Conditions 2 and 4 that, for all $t \in [0, \tau]$,

$$\int_0^t Q^{(\rho,\gamma)}(\beta_0, u)\alpha_0(u)\,du \to_{\mathbb{P}} \int_0^t q^{(\rho,\gamma)}(\beta_0, u)\alpha_0(u)\,du \tag{6.9}$$

as $n \to \infty$. In particular, by (3.6), (3.10) and Condition 3,

$$n^{-1}\langle \mathbf{U}_{\cdot}(\beta_0)\rangle(\tau) = \frac{1}{n}\int_0^\tau \sum_{\mathbf{r}\in\mathscr{P}} \mathbf{V}_{\mathbf{r}}(\beta_0, u) S_{\mathbf{r}}^{(0)}(\beta_0, u)\alpha_0(u)\,du \to_{\mathbb{P}} \mathbf{\Sigma}. \tag{6.10}$$

We now prove that the estimator $\hat{\beta}$ for the regression parameters is consistent.

THEOREM 1. *Assume Conditions* 1–4. *Then the estimator* $\hat{\beta}$ *maximizing* (3.3) *is consistent for* β_0.

PROOF. The proof is similar to the one of Theorem VI.1.1 in Andersen, Borgan, Gill and Keiding (1993). It is sufficient to show that

$$n^{-1} U_\tau^j(\beta_0) \to_{\mathbb{P}} 0, \tag{6.11}$$

$$n^{-1} \mathscr{I}_\tau^{jk}(\beta_0) \to_{\mathbb{P}} \sigma_{jk}, \tag{6.12}$$

as $n \to \infty$, for all j, k, and that there exists a finite constant K such that

$$\lim_{n\to\infty} \mathbb{P}\big(|n^{-1} R_\tau^{jkl}(\beta)| \le K \text{ for all } j,k,l \text{ and all } \beta \in \mathscr{B}_0\big) = 1, \tag{6.13}$$

where the $R_\tau^{jkl}(\beta)$ are the third-order partial derivatives of the logarithm of (3.3). First, note that (6.11) is an immediate consequence of Lenglart's inequality and (6.10). Second, to prove (6.12), note that by (6.7) the predictable variation of the second term on the right-hand side of (3.11) is bounded by $\int_0^\tau \sum_{\mathbf{r}\in\mathscr{P}} X_{\mathbf{r}}(u)^4 S_{\mathbf{r}}^{(0)}(\beta_0, u)\alpha_0(u)\,du$. From this (6.12) follows using Lenglart's inequality, (6.5), Condition 4, (3.11) and (6.10). Finally, to prove (6.13), first note that, for any n and all j, k, l and $\beta \in \mathscr{B}_0$,

$$|n^{-1} R_\tau^{jkl}(\beta)| \le \frac{6}{n}\int_0^\tau \sum_{\mathbf{r}\in\mathscr{P}} X_{\mathbf{r}}(u)^3\,dN_{\mathbf{r}}(u).$$

Then by Lenglart's inequality, (3.1) and (3.4) we have, for any $C, K > 0$,

$$\mathbb{P}\left(\frac{6}{n}\int_0^\tau \sum_{\mathbf{r}\in\mathscr{P}} X_{\mathbf{r}}(u)^3\, dN_{\mathbf{r}}(u) \geq K\right) \leq \frac{C}{K} + \mathbb{P}\left(\frac{6}{n}\int_0^\tau \sum_{\mathbf{r}\in\mathscr{P}} X_{\mathbf{r}}(u)^3 S_{\mathbf{r}}^{(0)}(\beta_0, u)\alpha_0(u)\,du \geq C\right). \tag{6.14}$$

By (6.5) and Condition 4 the second term on the right-hand side tends to zero as $n \to \infty$ if C is chosen large enough. Thus the right-hand side can be made arbitrarily small for n and K large enough, and (6.13) is proved. □

We now demonstrate the asymptotic normality of the maximum partial likelihood estimator.

THEOREM 2. *Assume Conditions* 1–4 *and let* $\hat\beta$ *be the estimator maximizing* (3.3). *Then*

$$\sqrt{n}\left(\hat\beta - \beta_0\right) \to_{\mathscr{D}} \mathscr{N}(\mathbf{0}, \boldsymbol{\Sigma}^{-1})$$

as $n \to \infty$, *where* $\boldsymbol{\Sigma}$ *defined in Condition* 3 *may be estimated consistently by* $n^{-1}\mathscr{I}_\tau(\hat\beta)$.

PROOF. The proof is similar to the one of Theorem VI.1.2 in Andersen, Borgan, Gill and Keiding (1993). The only point which needs attention is the Lindeberg condition of the martingale central limit theorem when this is used to show asymptotic normality of the score. To this end, we introduce $E_{\mathbf{r}}^j(\beta, t)$ for the jth component of (3.5). By a Chebychev-type inequality we then have, for all j and any $\varepsilon > 0$,

$$\begin{aligned}
&\frac{1}{n}\int_0^\tau \sum_{\mathbf{r}\in\mathscr{P}} \sum_{i\in\mathbf{r}} \left\{Z_{ij}(u) - E_{\mathbf{r}}^j(\beta_0, u)\right\}^2 \\
&\qquad \times I\left\{n^{-1/2}\left|Z_{ij}(u) - E_{\mathbf{r}}^j(\beta_0, u)\right| > \varepsilon\right\}\lambda_{(i,\mathbf{r})}(u)\,du \\
&\qquad\leq \frac{1}{\varepsilon n^{3/2}}\int_0^\tau \sum_{\mathbf{r}\in\mathscr{P}} \sum_{i\in\mathbf{r}} \left|Z_{ij}(u) - E_{\mathbf{r}}^j(\beta_0, u)\right|^3 \lambda_{(i,\mathbf{r})}(u)\,du \\
&\qquad\leq \frac{8}{\varepsilon n^{3/2}}\int_0^\tau \sum_{\mathbf{r}\in\mathscr{P}} X_{\mathbf{r}}(u)^3 S_{\mathbf{r}}^{(0)}(\beta_0, u)\alpha_0(u)\,du,
\end{aligned}$$

where the last inequality follows by (3.1), (3.4), (6.2) and (6.6). However, by (6.5) and Condition 4 the right-hand side tends to zero in probability as $n \to \infty$, and the Lindeberg condition is proved. □

We now turn to a study of the large-sample properties of the estimator $\hat A_0(t; \hat\beta)$ of (4.4) for the baseline hazard, and its variance estimator (4.6) given in terms of $\hat{\mathbf{B}}(t; \beta)$ defined in (4.8). To this end we have to impose the following extra conditions.

CONDITION 5. Let $n(t)$ be defined by (4.3). Then $n(t)/n$ is uniformly bounded away from zero in probability as $n \to \infty$.

CONDITION 6. There exist functions e and ϕ such that, for all $t \in [0, \tau]$ as $n \to \infty$,

$$\sum_{\mathbf{r} \in \mathscr{P}} \pi_t(\mathbf{r}) \mathbf{E}_{\mathbf{r}}(\beta_0, t) \to_{\mathbb{P}} e(\beta_0, t) \tag{6.15}$$

and

$$n \sum_{\mathbf{r} \in \mathscr{P}} \pi_t(\mathbf{r})^2 \{S_{\mathbf{r}}^{(0)}(\beta_0, t)\}^{-1} \to_{\mathbb{P}} \phi(\beta_0, t). \tag{6.16}$$

We first prove the following proposition.

PROPOSITION 2. *Let* $\hat{\mathbf{B}}(t; \beta)$ *be given by* (4.8), *and assume that Conditions* 1–6 *hold. Then, for any* $\beta^* \to_{\mathbb{P}} \beta_0$, *we have*

$$\sup_{t \in [0, \tau]} \left\| \hat{\mathbf{B}}(t; \beta^*) - \mathbf{B}(t; \beta_0) \right\| \to_{\mathbb{P}} 0$$

as $n \to \infty$, *with*

$$\mathbf{B}(t; \beta_0) = \int_0^t e(\beta_0, u) \alpha_0(u) \, du, \tag{6.17}$$

and $e(\beta_0, u)$ *defined in* (6.15).

PROOF. First note that by (3.4), (3.5) and (4.1) we may write

$$\hat{\mathbf{B}}(t; \beta) = \int_0^t \sum_{\mathbf{r} \in \mathscr{P}} \pi_u(\mathbf{r}) \mathbf{E}_{\mathbf{r}}(\beta, u) \{S_{\mathbf{r}}^{(0)}(\beta, u)\}^{-1} \, dN_{\mathbf{r}}(u). \tag{6.18}$$

Then, by (3.1) and (3.4),

$$\begin{aligned} &\sup_{t \in [0, \tau]} \left\| \hat{\mathbf{B}}(t; \beta^*) - \mathbf{B}(t; \beta_0) \right\| \\ &\quad \le \sup_{t \in [0, \tau]} \left\| \hat{\mathbf{B}}(t; \beta^*) - \hat{\mathbf{B}}(t; \beta_0) \right\| \\ &\qquad + \sup_{t \in [0, \tau]} \left\| \int_0^t \sum_{\mathbf{r} \in \mathscr{P}} \pi_u(\mathbf{r}) \mathbf{E}_{\mathbf{r}}(\beta_0, u) \{S_{\mathbf{r}}^{(0)}(\beta_0, u)\}^{-1} \, dM_{\mathbf{r}}(u) \right\| \\ &\qquad + \int_0^\tau \left\| \sum_{\mathbf{r} \in \mathscr{P}} \pi_u(\mathbf{r}) \mathbf{E}_{\mathbf{r}}(\beta_0, u) - e(\beta_0, u) \right\| \alpha_0(u) \, du. \end{aligned} \tag{6.19}$$

We will show that each of these three terms converge to zero in probability.

Denote the jth component of $\hat{\mathbf{B}}(t; \beta)$ by $\hat{B}^j(t; \beta)$. Then, by (6.6), (6.7), (6.18) and a Taylor series expansion, we have, when $\beta^* \in \mathscr{B}_0$,

$$\begin{aligned} &\left| \hat{B}^j(t; \beta^*) - \hat{B}^j(t; \beta_0) \right| \\ &\quad \le 2p \| \beta^* - \beta_0 \| \int_0^\tau \sum_{\mathbf{r} \in \mathscr{P}} \pi_u(\mathbf{r}) X_{\mathbf{r}}(u)^2 \{S_{\mathbf{r}}^{(0)}(\check{\beta}, u)\}^{-1} \, dN_{\mathbf{r}}(u), \end{aligned}$$

with $\check{\beta}$ on the line segment joining β^* and β_0. Therefore (6.5), Condition 4 and an application of Lenglart's inequality similar to (6.14) give that the leading term on the right-hand side of (6.19) tends to zero in probability as $\beta^* \to_{\mathbb{P}} \beta_0$.

The predictable variation process, evaluated at $t = \tau$, of the stochastic integral in the second term of (6.19) is bounded by $1/n$ times a constant times $\int_0^\tau (n/n(t)) D_n(t) \alpha_0(t)\, dt$. That this term converges to zero in probability therefore follows by Lenglart's inequality and Conditions 4 and 5. Finally, the third term on the right-hand side of (6.19) tends to zero in probability by dominated convergence (Proposition 1), invoking (6.6) and Conditions 4 and 6. □

We may then prove the following result about the asymptotic joint distribution of the estimator (4.4) and $\hat{\beta}$.

THEOREM 3. *Assume Conditions* 1–6, *and let* $\phi(\beta_0, t)$ *and* $\mathbf{B}(t; \beta_0)$ *be defined by* (6.16) *and* (6.17). *Then*

$$W(\cdot) = \sqrt{n}\left(\hat{A}_0(\cdot; \hat{\beta}) - A_0(\cdot)\right) + \sqrt{n}\left(\hat{\beta} - \beta_0\right)^T B(\cdot; \beta_0)$$

and $\sqrt{n}(\hat{\beta} - \beta_0)$ *are asymptotically independent, and* W *converges weakly to a mean-zero Gaussian martingale with variance function*

$$\omega^2(t; \beta_0) = \int_0^t \phi(\beta_0, u) \alpha_0(u)\, du, \tag{6.20}$$

which may be estimated uniformly consistently by $n\hat{\omega}^2(t; \hat{\beta})$ [*cf.* (4.7)].

PROOF. Except for the last result about uniform consistency of $n\hat{\omega}^2(t; \hat{\beta})$, the proof is similar to the one of Theorem VII.2.3 in Andersen, Borgan, Gill and Keiding (1993). For the first part we therefore only need to prove that $\sqrt{n}\,\hat{W}$, with $\hat{W}$ defined by (4.5), converges weakly to a Gaussian martingale with covariance function ω^2 given by (6.20). However, this follows by the martingale central limit theorem since

$$\langle \sqrt{n}\,\hat{W} \rangle(t) = n \int_0^t \sum_{\mathbf{r} \in \mathscr{P}} \pi_u(\mathbf{r})^2 \{S_{\mathbf{r}}^{(0)}(\beta_0, u)\}^{-1} \alpha_0(u)\, du, \tag{6.21}$$

tends in probability to $\omega^2(t)$ for all $t \in [0, \tau]$ by dominated convergence (Proposition 1) invoking (6.5) and Conditions 4, 5 and 6. Moreover, for the Lindeberg condition a Chebychev-type inequality gives, for any $\varepsilon > 0$,

$$n \int_0^\tau \sum_{\mathbf{r} \in \mathscr{P}} \pi_u(\mathbf{r})^2 \{S_{\mathbf{r}}^{(0)}(\beta_0, u)\}^{-1} I\left\{\sqrt{n}\, \pi_u(\mathbf{r}) \left(S_{\mathbf{r}}^{(0)}(\beta_0, u)\right)^{-1} > \varepsilon\right\} \alpha_0(u)\, du$$

$$\leq \frac{n^{3/2}}{\varepsilon} \int_0^\tau \sum_{\mathbf{r} \in \mathscr{P}} \pi_u(\mathbf{r})^3 \{S_{\mathbf{r}}^{(0)}(\beta_0, u)\}^{-2} \alpha_0(u)\, du,$$

which converges to zero in probability by (6.5) and Conditions 4 and 5.

Finally, to show that $n\hat{\omega}^2(t;\hat{\beta})$ is a uniformly consistent estimator for (6.20), note that $n\hat{\omega}^2(\cdot;\beta_0)$ is the optional variation process of the local square integrable martingale $\sqrt{n}\,\hat{W}$. Now, by Rebolledo's theorem [Andersen, Borgan, Gill and Keiding (1993), Theorem II.5.1], this optional variation process tends uniformly in probability to the same limit as the predictable variation process of the martingale. Thus $n\hat{\omega}^2(\cdot;\beta_0)$ is a uniformly consistent estimator for (6.20). That this remains true when β_0 is replaced by $\hat{\beta}$ follows by (6.5) and Conditions 4 and 5, using an argument similar to the one used to handle the first term on the right-hand side of (6.19). □

In Examples 2–5 the sampling distribution, conditional on the sampled risk set being chosen from a "block" of a partition of the support of $\pi_t(\mathbf{r})$, is uniform. The blocks of the partitions for these examples are specified as all subsets of the risk set of a particular composition or particular size. Here it is convenient to verify Conditions 2–6 conditionally in order to get the required unconditional results.

The general formulation of this "partition framework" is as follows. For a cohort of size n, let $\mathscr{P}(t)$ be all subsets of $\mathscr{R}(t) = \{i\colon Y_i(t) = 1\}$ which are assigned positive probability by π_t. Let I be an arbitrary index set which does not depend on n, and suppose for all t we have a partition of $\mathscr{P}(t)$ as a disjoint union $\bigcup_{\alpha\in I}\mathscr{S}_\alpha(t)$. Note that we can always factor the probabilities $\pi_t(\mathbf{r})$ (which depend on n) as

$$\pi_t(\mathbf{r}) = \sum_{\alpha\in I}\nu_{\alpha,n}\pi_{t,\alpha}(\mathbf{r}) \quad\text{and}\quad \nu_{\alpha,n}(t) = \sum_{\mathbf{r}\in\mathscr{S}_\alpha(t)}\pi_t(\mathbf{r}), \tag{6.22}$$

where $\pi_{t,\alpha}(\mathbf{r})$ is $\pi_t(\mathbf{r})$ conditioned on $\mathbf{r}\in\mathscr{S}_\alpha(t)$, and $\nu_{\alpha,n}(t)$ gives the probability that a set sampled from the distribution $\pi_t(\cdot)$ falls in $\mathscr{S}_\alpha(t)$. Recalling the factorization (4.1) and the definitions (3.4), (3.5), (6.1) and (6.3), we introduce

$$\begin{aligned}
\pi_{t,\alpha}(\mathbf{r}|i) &= \pi_{t,\alpha}(\mathbf{r})w_i(t,\mathbf{r}),\\
S_{\mathbf{r},\alpha}^{(\gamma)}(\beta,t) &= \sum_{i\in\mathbf{r}} Y_i(t)\mathbf{Z}_i(t)^{\otimes\gamma}\exp(\beta^T\mathbf{Z}_i(t))\pi_{t,\alpha}(\mathbf{r}|i),\\
Q_{\alpha,n}^{(\rho,\gamma)}(\beta_0,t) &= \frac{1}{n}\sum_{\mathbf{r}\in\mathscr{P}}\mathbf{E}_{\mathbf{r}}(\beta_0,t)^{\otimes\rho}S_{\mathbf{r},\alpha}^{(\gamma)}(\beta_0,t)
\end{aligned}\tag{6.23}$$

and

$$D_{\alpha,n}(t) = \sum_{\mathbf{r}\in\mathscr{P}}\pi_{t,\alpha}(\mathbf{r})\exp(b_0X_{\mathbf{r}}(t)). \tag{6.24}$$

Then we have

$$Q^{(\rho,\gamma)}(\beta_0,t) = \sum_{\alpha\in I}\nu_{\alpha,n}(t)Q_{\alpha,n}^{(\rho,\gamma)}(\beta_0,t) \quad\text{and}\quad D_n(t) = \sum_{\alpha\in I}\nu_{\alpha,n}(t)D_{\alpha,n}(t).$$

Similar expressions are valid for the left-hand sides of (6.15) and (6.16).

By the dominated convergence theorem, invoking inequalities similar to (6.5)–(6.8), we then get the following.

PROPOSITION 3. *Assume* (6.2) *and Condition* 5. *Suppose that, as* $n \to \infty$, *for all* $\alpha \in I$ *and* $t \in [0, \tau]$, $\nu_{\alpha,n}(t) \to \nu_\alpha(t)$, $D_{\alpha,n}(t) \to_{\mathbb{P}} D_{(\alpha)}(t)$ *and* $D_n(t) \to_{\mathbb{P}} D(t)$, *where* $D(t) = \sum_{\alpha \in I} \nu_\alpha(t) D_{(\alpha)}(t)$. *Further assume that there exist functions* $q_\alpha^{(\rho,\gamma)}(\beta_0, t)$, $e_\alpha(\beta_0, t)$ *and* $\phi_\alpha(\beta_0, t)$ *such that, for all* α, t *as* $n \to \infty$, *the following hold*:

$$Q_\alpha^{(\rho,\gamma)}(\beta_0, t) \to_{\mathbb{P}} q_\alpha^{(\rho,\gamma)}(\beta_0, t);$$

$$\sum_{\mathbf{r} \in \mathscr{P}} \pi_{t,\alpha}(\mathbf{r}) \mathbf{E}_{\mathbf{r}}(\beta_0, t) \to_{\mathbb{P}} e_\alpha(\beta_0, t);$$

$$n \sum_{\mathbf{r} \in \mathscr{P}} \pi_{t,\alpha}(\mathbf{r})^2 \left\{ S_{\mathbf{r},\alpha}^{(0)}(\beta_0, t) \right\}^{-1} \to_{\mathbb{P}} \phi_\alpha(\beta_0, t).$$

Then Conditions 2 *and* 6 *are satisfied with*

$$q^{(\rho,\gamma)}(\beta_0, t) = \sum_{\alpha \in I} \nu_\alpha(t) q_\alpha^{(\rho,\gamma)}(\beta_0, t),$$

$$e(\beta_0, t) = \sum_{\alpha \in I} \nu_\alpha(t) e_\alpha(\beta_0, t),$$

$$\phi(\beta_0, t) = \sum_{\alpha \in I} \nu_\alpha(t) \phi_\alpha(\beta_0, t).$$

Finally,

$$\boldsymbol{\Sigma} = \sum_{\alpha \in I} \boldsymbol{\Sigma}_\alpha \quad \textit{where } \boldsymbol{\Sigma}_\alpha = \int_0^\tau \nu_\alpha(t) \left[q_\alpha^{(0,2)}(\beta_0, t) - q_\alpha^{(2,0)}(\beta_0, t) \right] \alpha_0(t)\, dt,$$

so that $\boldsymbol{\Sigma}$ *is positive definite if there exists an* α *with* $\boldsymbol{\Sigma}_\alpha$ *positive definite.*

7. Asymptotic results for specific sampling schemes. In this section, we will show that Theorems 1–3 may be invoked for the four sampling schemes presented in Examples 2–5. We will be working with an iid model where it is assumed that the at-risk indicators, covariate processes and classification variables $(Y_i(t), \mathbf{Z}_i(t), C_i(t))$, $i = 1, 2, \dots, n$, are independent copies of $(Y(t), \mathbf{Z}(t), C(t))$, with $Y(\cdot)$, $\mathbf{Z}(\cdot)$ and $C(\cdot)$ adapted and left-continuous with right-hand limits on $[0, \tau]$. Recall that the classification variables $C_i(t)$, introduced in Section 5, are elements of a (small) finite set $\mathscr{C} \subset \mathbb{R}$ that gives the "type" of the individuals at time t. The assumption that the $C_i(\cdot)$ are independent covers cases where classification is based on the absolute exposure status of the individuals, yet it does not cover classification based on relative exposure where the "type" of an individual depends on values observed on other cohort members. Below, we let $p(t) = \mathbb{P}(Y(t) = 1)$, and for a set $\mathbf{r} = \{i_1, i_2, \dots, i_m\}$ we define $\mathbf{C}_{\mathbf{r}}(t) = (C_{i_1}(t), \dots, C_{i_m}(t))$ and $Y_{\mathbf{r}}(t) = \prod_{i \in \mathbf{r}} Y_i(t)$.

We work in the framework of the partition structure considered in Proposition 3. We suppose here that, for each t and α, there is an $m = m(\alpha)$ such that $\pi_{t,\alpha}(\mathbf{r})$, given in (6.22), is uniform over some collection $\mathscr{S}_\alpha(t)$ of sets of size m in a regular way that we now describe: Let $\mathbf{c} = (c_1, \dots, c_m)$ denote an m-vector of functions c_k, $1 \le k \le m$, each adapted, left-continuous with right-hand limits on $[0,\tau]$, taking values in $\mathscr{C}$. For a permutation σ, we define $c_\sigma(t) = (c_{\sigma(1)}(t), \dots, c_{\sigma(m)}(t))$. We say that $\mathscr{A}$, a finite collection of such $\mathbf{c}$, is symmetric if $\mathbf{c} \in \mathscr{A}$ if and only if $\mathbf{c}_\sigma \in \mathscr{A}$ for all σ. Although in our examples the functions c_k are constant, the added generality allows, for example, the study of counter-matching type designs where m_l, $l \in \mathscr{C}$, the number of individuals of type l in the sampled risk set, is allowed to depend on $\mathscr{F}_{t-}$. We say that $\pi_{t,\alpha}(\mathbf{r})$, or alternatively $\pi_t(\mathbf{r})$, is *regular uniform* if, for all $\alpha \in I$ and $t \in [0,\tau]$, there exists $m = m(\alpha)$ such that $\pi_{t,\alpha}(\mathbf{r})$ is uniform over sets $\mathscr{S}_\alpha(t) \subset \{\mathbf{r} \subset \mathscr{R}(t): |\mathbf{r}| = m\}$ such that $I(\mathbf{r} \in \mathscr{S}_\alpha(t)) = I_{\{|\mathbf{r}|=m\}} Y_\mathbf{r}(t) \Sigma_{\mathbf{c} \in \mathscr{A}} I(\mathbf{C}_\mathbf{r}(t) = \mathbf{c}(t))$, with $\mathscr{A}$ symmetric, and $\inf_{t \in [0,\tau]} \mathbb{P}(\mathbf{u} \in \mathscr{S}_\alpha(t)) > 0$, where here and in the following $\mathbf{u} = \{1, 2, \dots, m\}$.

We now indicate how Examples 2–5 fit into the above partition framework. In Examples 2 and 3 the partition consists of a single block of sets of given size m. In Example 2, $\alpha = m$ and the sampling distribution is uniform over $\mathscr{S}_m(t)$, the (single block) of all subsets of $\mathscr{R}(t)$ of size m; that is, $I(\mathbf{r} \in \mathscr{S}_m(t)) = I_{\{|\mathbf{r}|=m\}} Y_\mathbf{r}$. In counter-matching, Example 3, I consists of the single element $\alpha = (m_1, m_2, \dots, m_L)$ and the sampling distribution is uniform over $\mathscr{S}_\alpha(t)$, the (single block) of all subsets of $\mathscr{R}(t)$ of size $m = \Sigma_{l \in \mathscr{C}} m_l$ with m_l individuals of type l. Here we take $\mathscr{A}$ to be the collection of all constant vectors that have m_l entries of l, for $l \in \mathscr{C}$. In quota sampling, Example 4, we partition sampled risk sets according to size, so $\alpha \in I = \{(\alpha_0, \alpha_1): \alpha_0 = m - m_1,\ \alpha_1 = m_1,\ m = 1, 2, \dots\}$ and $\pi_{t,\alpha}(\mathbf{r})$ is uniform over all such subsets of size m of $\mathscr{R}(t)$ with m_1 target individuals; that is, quota sampling reduces to counter-matching with $\mathscr{C} = \{0, 1\}$ given the size m. In Example 5, the number of counter-matched subjects $\overline{m}$ and the number of additionally sampled controls $\tilde{m}$ are fixed, with $m = \overline{m} + \tilde{m}$ the total number in the sampled risk set. Let $\alpha \in I = \{(\alpha_1, \dots, \alpha_L): \overline{m}_l \le \alpha_l,\ \Sigma_{l=1}^L \alpha_l = m\}$. Then, conditional on the composition of the sampled risk given by α, the sampling distribution is uniform over $\mathscr{S}_\alpha(t)$, all subsets of the risk set with α_l individuals of type l; that is, this scheme reduces to counter-matching given the composition α.

Our main result in this section shows how Theorems 1–3 may be invoked in the regular uniform case.

THEOREM 4. *Assume the above iid model, Condition* 1 *and* $\tau < \infty$. *Further, suppose that* $\pi_t(\mathbf{r})$ *is regular uniform,* $\nu_{\alpha,n}(t) \to \nu_\alpha(t)$ *as* $n \to \infty$ [*cf.* (6.22)] *and* $\inf_{0 \le t \le \tau} \alpha_0(t) > 0$. *Either let* $M = \{m = m(\alpha): \exists\ t,\ \nu_{\alpha,n}(t) \neq 0\}$ *be independent of n and finite and, for each* $m \in M$, *let* $\int_0^\tau \{\mathbb{E} \exp(2 b_0 \|\mathbf{Z}(t)\|)\}^m \alpha_0(t)\, dt < \infty$ *or assume the covariate process* $\mathbf{Z}$ *is bounded. Further, suppose there exists* $\overline{w}_i(t,\mathbf{r})$, *with* $\overline{w}_i(t,\mathbf{r}) \ge \delta$ *for all i for some* $\delta > 0$, *and* $\max_i |w_i(t,\mathbf{r})/n(t) - \overline{w}_i(t,\mathbf{r})| \to_{\mathbb{P}} 0$ *as* $n \to \infty$. *Finally, let* $\mathbf{u} = \{1, 2, \dots, m\}$, *and assume that, for all* $\alpha \in I$, *the matrix*

$$\text{(7.1)} \quad \mathbf{V}_\alpha = \int_0^\tau \nu_\alpha(t) \mathbb{E}\Big\{\mathrm{Cov}\big(\hat{\mathbf{Z}}_\alpha(t) | \mathbf{Z}_\alpha(t)\big)\Big\} \alpha_0(t)\, dt \quad \textit{is positive definite},$$

where the distribution of $\mathbf{Z}_\alpha(t) = (\mathbf{Z}_{\alpha,1}(t), \mathbf{Z}_{\alpha,2}(t), \dots, \mathbf{Z}_{\alpha,m}(t)) \in \mathbb{R}^{p \times m}$ *is given by*

$$\mathbb{P}(\mathbf{Z}_\alpha(t) \in B) = \mathbb{P}\big((\mathbf{Z}_1(t), \mathbf{Z}_2(t), \dots, \mathbf{Z}_m(t)) \in B | \mathbf{u} \in \mathscr{S}_\alpha(t)\big),$$

and the distribution of $\hat{\mathbf{Z}}_\alpha(t)$ *is specified through the conditional law given* $\mathbf{Z}_\alpha(t)$,

$$\mathbb{P}\big(\hat{\mathbf{Z}}_\alpha(t) = \mathbf{Z}_{\alpha,j}(t)|\mathbf{Z}_\alpha(t)\big) = \frac{\exp\big(\beta_0^T\mathbf{Z}_{\alpha,j}(t)\big)\overline{w}_j(t,\mathbf{u})}{\sum_{i\in\mathbf{u}}\exp\big(\beta_0^T\mathbf{Z}_{\alpha,i}(t)\big)\overline{w}_i(t,\mathbf{u})}.$$

Then the hypotheses of Proposition 3 *and Theorems* 1–3 *are satisfied, and we may apply Proposition* 3 *with*

$$q_\alpha^{(\rho,\gamma)}(\beta_0,t) = p(t)\mathbb{E}\big\{\overline{\mathbf{E}}_\mathbf{u}(\beta_0,t)^{\otimes\rho}\overline{S}_\mathbf{u}^{(\gamma)}(\beta_0,t)|\mathbf{u}\in\mathscr{S}_\alpha(t)\big\},$$

$$e_\alpha(\beta_0,t) = \mathbb{E}\big\{\overline{\mathbf{E}}_\mathbf{u}(\beta_0,t)|\mathbf{u}\in\mathscr{S}_\alpha(t)\big\} = \mathbb{E}\Big\{\mathbb{E}\big(\hat{\mathbf{Z}}_\alpha(t)|\mathbf{Z}_\alpha(t)\big)\Big\}$$

and

$$\phi_\alpha(\beta_0,t) = p(t)^{-1}\mathbb{E}\Big\{\big[\overline{S}_\mathbf{u}^{(0)}(\beta_0,t)\big]^{-1}|\mathbf{u}\in\mathscr{S}_\alpha(t)\Big\},$$

where

$$\overline{S}_\mathbf{u}^{(\gamma)}(\beta_0,t) = \sum_{i\in\mathbf{u}} Y_i(t)\mathbf{Z}_i(t)^{\otimes\gamma}\exp\big(\beta_0^T\mathbf{Z}_i(t)\big)\overline{w}_i(t,\mathbf{u})$$

and

$$\overline{\mathbf{E}}_\mathbf{u}(\beta_0,t) = \frac{\overline{S}_\mathbf{u}^{(1)}(\beta_0,t)}{\overline{S}_\mathbf{u}^{(0)}(\beta_0,t)}.$$

In particular, the asymptotic covariance matrix is given by (6.25) *with terms*

$$\begin{aligned}\boldsymbol{\Sigma}_\alpha &= \int_0^\tau \nu_\alpha(t)p(t)\mathbb{E}\Big\{\sum_{i\in\mathbf{u}}\big[\mathbf{Z}_i(t)^{\otimes 2} - \overline{\mathbf{E}}_\mathbf{u}(\beta_0,t)^{\otimes 2}\big] \\ &\qquad\times\exp\big(\beta_0^T\mathbf{Z}_i(t)\big)\overline{w}_i(t,\mathbf{u})\Big|\mathbf{u}\in\mathscr{S}_\alpha(t)\Big\}\alpha_0(t)\,dt \qquad (7.2)\\ &= \mathbf{E}\Big\{\int_0^\tau \nu_\alpha(t)p(t)\mathrm{Cov}\big(\hat{\mathbf{Z}}_\alpha(t)|\mathbf{Z}_\alpha(t)\big)\sum_{i\in\mathbf{u}}\exp\big(\beta_0^T\mathbf{Z}_{\alpha,i}(t)\big)\overline{w}_i(t,\mathbf{u})\,\alpha_0(t)\,dt\Big\}.\end{aligned}$$

The proof of this theorem is given in the Appendix. Note, however, from (7.2) that $\boldsymbol{\Sigma}_\alpha$ is the expectation of the integral of the product of the partition probability, the at-risk probability, an estimate of the covariance of the covariate of the failure and an estimate of the average hazard of the cohort, the estimates based on the sampled risk set.

In Examples 8, 9 and 11, $\nu_{\alpha,n}\nu_\alpha(t) = 1$ for a particular α. Assuming normal covariates, the moment condition of Theorem 4 for these examples (and for more general cases where M is finite and does not depend on n) is satisfied by the following argument. First verify that if $Z \sim\mathscr{N}(\mu,\sigma^2)$, $\mathbf{E}\{\exp(\gamma|Z|)\} \le 2\exp(\frac{1}{2}\gamma^2\sigma^2 + \gamma|\mu|)$. Then, as $X_\mathbf{r}(t) \le \sum_{j\in\mathbf{r}}\|\mathbf{Z}_j(t)\|$, we have $\int_0^\tau\{\mathbb{E}\exp(2b_0\|\mathbf{Z}(t)\|)\}^m\alpha_0(t)\,dt < \infty$, for any $m = 1,2,\dots,$ whenever

$$Z(t) \sim\mathscr{N}\big(\mu(t),\sigma^2(t)\big) \quad\text{with}\quad \sup_{t\in[0,\tau]} |\mu(t)| + \sigma^2(t) < \infty,$$

and $A_0(\tau) < \infty$. In Example 10 we assume the covariates are bounded.

EXAMPLE 8 (Nested case-control sampling). The sampling scheme of Example 2 is uniform over $\mathscr{S}_m(t) = \{\mathbf{r} \subset \mathscr{R}(t): |\mathbf{r}| = m\}$. Theorem 4 applies with $w_i(t,\mathbf{r})/n(t) = \overline{w}_i(t,\mathbf{r}) = 1/m$.

EXAMPLE 9 (Counter-matching). The sampling scheme of Example 3 is uniform over $\mathscr{S}_\alpha(t) = \{\mathbf{r} \subset \mathscr{R}(t)$: $\mathbf{r} \cap \mathscr{R}_l(t)| = m_l;\ l \in \mathscr{C}\}$. Theorem 4 applies with $w_i(t,\mathbf{r})/n(t) = n_{C_i(t)}(t)/(n(t)m_{C_i(t)}) \to_{\mathbb{P}} p_{C_i(t)}(t)/m_{C_i(t)} = \overline{w}_i(t,\mathbf{r})$, where, here and in the following two examples, $p_l(t) = \mathbb{P}(C(t) = l | Y(t) = 1)$.

EXAMPLE 10 (Quota sampling). The sampling scheme of Example 4 conditioned on sampling $m = m(\alpha)$ is uniform over

$$\mathscr{S}_\alpha(t) = \{\mathbf{r} \subset \mathscr{R}(t): |\mathbf{r} \cap \mathscr{R}_1(t)| = m_1, |\mathbf{r} \cap \mathscr{R}_0(t)| = m - m_1\},$$

and

$$\nu_{\alpha,n}(t) = \frac{\binom{m-1}{m_1-1}\binom{n(t)-m}{n_1(t)-m_1}}{\binom{n(t)}{n_1(t)}} \to \nu_\alpha(t) = \binom{m-1}{m_1-1} p_0^{m_0}(t) p_l^{m_1}(t)$$

as $n \to \infty$. Theorem 4 applies with

$$\frac{w_i(t,\mathbf{r})}{n(t)} = \overline{w}_i(t,\mathbf{r}) = \frac{m_1 - C_i(t)}{m_1(m-1)} \quad \text{for } \mathbf{r} \in \mathscr{S}_\alpha(t).$$

EXAMPLE 11 (Counter-matching with additional randomly sampled controls). Let $I = \{(\alpha_1, \dots, \alpha_L): \overline{m}_l \le \alpha_l,\ \sum_l \alpha_l = \overline{m} + \tilde{m}\}$. For $\alpha \in I$, the sampling scheme of Example 5 conditioned on $\mathbf{r} \in \mathscr{S}_\alpha(t) = \{\mathbf{r} \subset \mathscr{R}(t)$: $|\mathbf{r} \cap \mathscr{R}_l(t)| = \alpha_l = \overline{m}_l + \tilde{m}_l;\ l \in \mathscr{C}\}$ is uniform over $\mathscr{S}_\alpha(t)$ with

$$\nu_{\alpha,n}(t) = \frac{\prod_l \binom{n_l(t) - \overline{m}_l}{\tilde{m}_l}}{\binom{n(t) - \overline{m}}{\tilde{m}}} \to \nu_\alpha(t) = \binom{\tilde{m}}{\tilde{m}_1, \dots, \tilde{m}_L} \prod_{l \in \mathscr{C}} p_l^{\tilde{m}_l}(t)$$

as $n \to \infty$. Theorem 4 applies with

$$\frac{w_i(t,\mathbf{r})}{n(t)} = \frac{n_{C_i(t)}(t)}{n(t)\left(\overline{m}_{C_i(t)} + \tilde{m}_{C_i(t)}\right)} \to_{\mathbb{P}} \frac{p_{C_i(t)}(t)}{\overline{m}_{C_i(t)} + \tilde{m}_{C_i(t)}} \quad \text{for } \mathbf{r} \in \mathscr{S}_\alpha(t).$$

8. Discussion. The general framework we have presented makes it possible to analyze a large class of sampling designs. Counter-matching, described using these methods in Langholz and Borgan (1995), and the three completely novel designs given in Examples 4–6 illustrate the potential usefulness of the methods. Many techniques available for the analysis of full-cohort data are accommodated with little change for sampled data. In this paper, we have given estimation methods for relative risks (using the partial likelihood) and baseline hazards. Survival probabilities and extensions to multistrata and multistate problems are easily accommodated applying the approaches given in Andersen, Borgan, Gill and Keiding [(1993), Sections VII.1 and VII.2] in a straightforward way. Estimation of relative mortality is developed in Borgan and Langholz (1993). Further, the marked point process framework can be generalized to accommodate other design problems. For instance, in Langholz and Borgan (1995), a simple generalization of the mark space described in Section 2 is used to derive a partial likelihood when failures are also to be sampled from the cohort.

For cohort data, the estimator for the vector of regression parameters based on Cox's partial likelihood and the Breslow estimator for the integrated baseline hazard are asymptotically efficient [e.g., Andersen, Borgan, Gill and Keiding (1993), Section VIII.4.3]. As mentioned in Section 5, this will not be the case in general for the estimators proposed in the present paper. In this relation the results of Robins, Rotnitzky and Zhao [(1994), Section 8.3] are worth mentioning. They study the situation with time-independent covariates when covariate information is missing at random. For this situation they show how one, in principle, may construct an efficient estimator (given as the solution of an integral equation) for the vector of regression parameters, and they claim that their results may be modified to cover the nonrandom missingness one encounters in nested case-control sampling. It would be interesting to have these details worked out and to see a comparison of the performance of our estimator based on the partial likelihood to the optimal estimator.

As a practical approach in a specific situation, we suggest the following procedure: (i) develop a "cost function" which captures the goals of the study and the costs associated with collecting the needed information; (ii) think of some sensible designs for the problem; and (iii) compare the performance of the designs to each other, either using large-sample variances or by computer simulation. The results of Sections 6 and 7 should aid in making large-sample comparisons. For instance, using the asymptotic variance formula (7.2) with the weights specified as in Examples 8 and 9, the counter-matching method of Example 3 was found to have much smaller asymptotic variance than nested case-control sampling in situations of practical importance [Langholz and Borgan (1995)]. Of course, neither candidate designs nor the methods of analysis should be restricted to those described here. However, the simplicity of the estimators and the strong theoretical basis for these methods of analysis make such designs desirable.

In earlier work, Goldstein and Langholz (1992) developed the asymptotic theory for nested case-control sampling based on a different model from that given here. In their model, just after a change in Y_i or N_i for some subject i in the cohort, a set of controls is randomly (and independently) sampled for each at-risk subject. Then, when a failure occurs, the sampled risk set would be already established. The counting processes then just count failure occurrences, as in the full-cohort framework of Andersen and Gill (1982), and the fictitious sampling is predictable under an obvious enlarged filtration. In the marked point process approach of the present paper, the counting processes count joint failure *and* sampled risk set occurrences. The probability laws for the sampling are predictable but the sampling itself is adapted (but not predictable) with respect to the filtration $(\mathscr{F}_t)$. The observed scores from both models are identical but the score *process* (3.9) is exactly a martingale while that of Goldstein and Langholz is a martingale plus an additional term. This second term is due to the additional variation generated by the multiplicity of (fictitiously) sampled risk sets and is asymptotically negligible. The approach given here not only vastly simplifies proofs, allows for a partial likelihood interpretation and leads, quite naturally, to the estimator of the cumulative baseline hazard but also reflects how nested case-control sampling is actually done.

APPENDIX

Proof of Theorem 4. The proof of Theorem 4 depends on three lemmas. Recall first that $m = m(\alpha)$ and $\mathbf{u} = \{1, 2, \dots, m\}$, and define

$$f_{\alpha,n}^{-1}(t) \equiv g_{\alpha,n}(t) \equiv \frac{|\mathscr{S}_\alpha(t)|}{\binom{n}{m}} \tag{A.1}$$

and

$$f_\alpha^{-1}(t) \equiv g_\alpha(t) \equiv \mathbb{P}(\mathbf{u} \in \mathscr{S}_\alpha(t)).$$

Recall also that $g_\alpha(t)$ is assumed to be uniformly bounded away from zero.

LEMMA 1. *Let $\pi_{t,\alpha}(\mathbf{r})$ be regular uniform. Then, as $n \to \infty$,*

$$\sup_{t\in[0,\tau]} |g_{\alpha,n}(t) - g_\alpha(t)| \to_{\mathbb{P}} 0 \quad \textit{and} \quad \sup_{t\in[0,\tau]} \left| \frac{n(t)}{n} - p(t) \right| \to_{\mathbb{P}} 0.$$

PROOF. Note that, by the symmetry of $\mathscr{A}$,

$$Y_{i_1,\dots,i_m} I(\mathbf{C}_{i_1,\dots,i_m} = \mathbf{c}) = \prod_{k=1}^m Y_{i_k} I(C_{i_k} = c_k)$$

is well defined for any indices $i_1, i_2, \dots, i_m$, not necessarily distinct. For $|\mathbf{r}| = m$ we have that

$$\binom{n}{m}^{-1} \sum_{\mathbf{r}} Y_{\mathbf{r}} I(\mathbf{C}_{\mathbf{r}} = \mathbf{c}) - n^{-m} \sum_{i_1,\dots,i_m} Y_{i_1,\dots,i_m} I(\mathbf{C}_{i_1,\dots,i_m} = \mathbf{c})$$

converges to zero almost surely in the supremum norm. Further,

$$n^{-m}\sum_{i_1,\dots,i_m} Y_{i_1,\dots,i_m} I(\mathbf{C}_{i_1,\dots,i_m}=\mathbf{c}) = \prod_{k=1}^{m}\frac{1}{n}\sum_{i_k=1}^{n} Y_{i_k} I(C_{i_k}=c_k).$$

As $Y_{i_k}I(C_{i_k}=c_k)$ for $1\le i_k\le n$ are iid elements in $D[0,\tau]$ (after reversing the time axis), the strong law of large numbers in $D[0,1]$ of Rao (1963) implies that this last quantity converges in supremum norm almost surely to $\prod_{k=1}^{m}\mathbb{E}Y_k I(C_k=c_k)=\mathbb{E}Y_{\mathbf{u}}I(\mathbf{C}_{\mathbf{u}}=c)$. Hence

$$g_{\alpha,n}(t)=\binom{n}{m}^{-1}\sum_{\mathbf{r}}Y_{\mathbf{r}}\sum_{\mathbf{c}\in\mathscr{A}}I(\mathbf{C}_{\mathbf{r}}=\mathbf{c})$$

converges in supremum norm almost surely to $\mathbb{E}Y_{\mathbf{u}}\sum_{\mathbf{c}\in\mathscr{A}}I(\mathbf{C}_{\mathbf{u}}=\mathbf{c})=\mathbb{E}I(\mathbf{u}\in\mathscr{S}_\alpha(t))=g_\alpha(t)$. The second claim of the lemma follows by applying Rao's theorem to $n(t)/n=n^{-1}\sum_{i=1}^{n}Y_i(t)$ and the observation that, for any $\alpha\in I$, $p(t)\ge g_\alpha(t)$ follows from $\{Y_1(t)=1\}\subset\{\mathbf{u}\in\mathscr{S}_\alpha(t)\}$, and hence $\inf_{t\in[0,\tau]}p(t)>0$. □

LEMMA 2. *For $\mathbf{r}\in\mathscr{P}(t)$, $|\mathbf{r}|=m$, let $F_t(\mathbf{r})$ be a symmetric function, not depending on n, of the variates for $j\in\mathbf{r}$ and assume $\mathbb{E}F_t^2(\mathbf{u})<\infty$. Let*

$$H_n(t)=\binom{n}{m}^{-1}\sum_{\mathbf{r}\in\mathscr{P}}F_t(\mathbf{r})I(\mathbf{r}\in\mathscr{S}_\alpha(t)),\qquad H(t)=\mathbb{E}[F_t(\mathbf{u})I(\mathbf{u}\in\mathscr{S}_\alpha(t))]$$

and

$$a_n=\binom{n}{m}^{-2}|\{\mathbf{r}\cap\mathbf{s}\neq\varnothing: |\mathbf{r}|=|\mathbf{s}|=m\}|.$$

Then

$$\mathbb{E}\{H_n(t)-H(t)\}^2\le a_n\mathbb{E}F_t^2(\mathbf{u}) \tag{A.2}$$

and

$$\sum_{\mathbf{r}\in\mathscr{P}}F_t(\mathbf{r})\pi_{t,\alpha}(\mathbf{r})\to_{\mathbb{P}}\mathbb{E}\{F_t(\mathbf{u})|\mathbf{u}\in\mathscr{S}_\alpha(t)\}. \tag{A.3}$$

If, for $G_t(\mathbf{r})$, depending on n, there exists an A_n with $\mathbb{P}(A_n)\to 1$ and $\mathbb{E}|G_t(\mathbf{r})-F_t(\mathbf{r})|I_{A_n}=\mathbb{E}|G_t(\mathbf{u})-F_t(\mathbf{u})|I_{A_n}\to 0$, then

$$\sum_{r\in\mathscr{P}}G_t(\mathbf{r})\pi_{t,\alpha}(\mathbf{r})\to_{\mathbb{P}}\mathbb{E}\{F_t(\mathbf{u})|\mathbf{u}\in\mathscr{S}_\alpha(t)\}.$$

PROOF. Clearly $\mathbb{E}H_n(t)=H(t)$, and we have (A.2) by expanding the square, using that $\mathbf{r}\cap\mathbf{s}=\varnothing$ implies $F_t(\mathbf{r})I(\mathbf{r}\in\mathscr{S}_\alpha(t))$ and $F_t(\mathbf{s})I(\mathbf{s}\in\mathscr{S}_\alpha(t))$ are independent and then applying the Cauchy–Schwarz inequality. Using now that $a_n\to 0$ we have that $H_n(t)\to_{\mathbb{P}}H(t)$. Using (A.1) and Lemma 1, $\sum_{\mathbf{r}\in\mathscr{P}}F_t(\mathbf{r})\pi_{t,\alpha}(\mathbf{r})=f_{\alpha,n}(t)H_n(t)\to_{\mathbb{P}}f_\alpha(t)H(t)=\mathbb{E}\{F_t(\mathbf{u})|\mathbf{u}\in\mathscr{S}_\alpha(t)\}$, showing

(A.3). Finally, note

$$\left| \binom{n}{m}^{-1} \sum_{\mathbf{r} \in \mathscr{P}} G_t(\mathbf{r}) - \binom{n}{m}^{-1} \sum_{\mathbf{r} \in \mathscr{P}} F_t(\mathbf{r}) \right| I_{A_n^c} \to_{\mathbb{P}} 0,$$

while

$$\mathbb{E}\left| \binom{n}{m}^{-1} \sum_{\mathbf{r} \in \mathscr{P}} G_t(\mathbf{r}) - \binom{n}{m}^{-1} \sum_{\mathbf{r} \in \mathscr{P}} F_t(\mathbf{r}) \right| I_{A_n} \le \mathbb{E}|F_t(\mathbf{u}) - G_t(\mathbf{u})| I_{A_n},$$

which tends to zero by assumption. □

LEMMA 3. *Let $\alpha \in I$ and $m = m(\alpha)$. With $X_{\mathbf{r}}(t) = \max_{j \in \mathbf{r}} \|\mathbf{Z}_j(t)\|$, suppose there exists $b_0 > 3\|\beta_0\|$ such that $\int_0^\tau \mathbb{E}\{\exp(2 b_0 X_{\mathbf{u}}(t))\} \alpha_0(t)\, dt < \infty$. Then, as $n \to \infty$,*

$$\int_0^\tau D_{\alpha, n}(t) \alpha_0(t)\, dt \to_{\mathbb{P}} \int_0^\tau D_{(\alpha)}(t) \alpha_0(t)\, dt. \tag{A.4}$$

PROOF. Recalling (6.24), (A.1) and Lemma 2, $D_{\alpha, n}(t) = f_{\alpha, n}(t) H_n(t)$ and $D_{(\alpha)}(t) = f_\alpha(t) H(t)$ when $F_t(\mathbf{r}) = \exp(b_0 X_{\mathbf{r}}(t))$. First,

$$\left| \int_0^\tau f_{\alpha, n}(t) H_n(t) \alpha_0(t)\, dt - \int_0^\tau f_\alpha(t) H(t) \alpha_0(t)\, dt \right|$$
$$\le \int_0^\tau |f_{\alpha, n}(t) - f_\alpha(t)| H_n(t) \alpha_0(t)\, dt + \int_0^\tau |H_n(t) - H(t)| f_\alpha(t) \alpha_0(t)\, dt.$$

Then by the Cauchy–Schwarz inequality,

$$\left(\int_0^\tau | f_{\alpha, n}(t) - f_\alpha(t) | H_n(t) \alpha_0(t)\, dt \right)^2$$
$$\le \int_0^\tau (f_{\alpha, n}(t) - f_\alpha(t))^2 \alpha_0(t)\, dt \int_0^\tau H_n^2(t) \alpha_0(t)\, dt$$

and

$$\left(\int_0^\tau | H_n(t) - H(t) | f_\alpha(t) \alpha_0(t)\, dt \right)^2$$
$$\le \int_0^\tau (H_n(t) - H(t))^2 \alpha_0(t)\, dt \int_0^\tau f_\alpha^2(t) \alpha_0(t)\, dt.$$

As $\int_0^\tau f_\alpha^2(t) \alpha_0(t) < \infty$ it suffices to show that the following hold:

$$\int_0^\tau (f_{\alpha, n}(t) - f_\alpha(t))^2 \alpha_0(t)\, dt \to_{\mathbb{P}} 0,$$

$$\int_0^\tau H_n^2(t) \alpha_0(t)\, dt = o_p(1)$$

and

$$\int_0^\tau (H_n(t) - H(t))^2 \alpha_0(t)\, dt \to_{\mathbb{P}} 0.$$

Using (A.2) and $\int_0^\tau \mathbb{E} F_t^2(\mathbf{u})\alpha_0(t) < \infty$ we have the third and hence the second claim. Next, by Lemma 1, with $\delta = \inf_t g_\alpha(t)/2$,

$$\Omega_n = \left\{ |\mathscr{S}_\alpha(t)| \geq \delta \binom{n}{m} \text{ for all } t \in [0, \tau] \right\},$$

$\mathbb{P}(\Omega_n) \to 1$. Hence $[f_{\alpha,n}(t) - f_\alpha(t)]^2 I_{\Omega_n}$ is bounded and converges to zero in probability, and the dominated convergence theorem completes the proof of (A.4). □

Proof of Theorem 4. We will first verify Condition 4. Condition 4 is vacuous if the covariate processes are bounded. Otherwise, for given $\alpha \in I$ and $m = m(\alpha)$, let $X_\mathbf{r}(t) = \max_{i \in \mathbf{r}} \|\mathbf{Z}_i(t)\| \leq \sum_{i \in \mathbf{r}} \|\mathbf{Z}_i(t)\|$; apply Lemma 2 with $F_t(\mathbf{r}) = \exp(b_0 X_\mathbf{r}(t))$ to show $D_{\alpha,n}(t) \to_{\mathbb{P}} D_{(\alpha)}(t)$; and next invoke Lemma 3 to assert (A.4). When M is finite and does not depend on n, summing over $m \in M$ yields Condition 4. Next, in order to show Condition 2, by Proposition 3 it only remains to verify that $Q_\alpha^{(\rho,\gamma)}(\beta_0, t)$ in (6.23) converges in probability to $q_\alpha^{(\rho,\gamma)}(\beta_0, t)$. For this, we apply Lemma 2 (componentwise if necessary) with

$$G_t(\mathbf{r}) = \mathbf{E}_\mathbf{r}(\beta_0, t)^{\otimes \rho} \sum_{i \in \mathbf{r}} Y_i(t) \mathbf{Z}_i(t)^{\otimes \gamma} \exp(\beta_0^T \mathbf{Z}_i(t)) \frac{w_i(t, \mathbf{r})}{n},$$

$$F_t(\mathbf{r}) = p(t) \overline{\mathbf{E}}_\mathbf{r}(\beta_0, t)^{\otimes \rho} \overline{S}_\mathbf{r}^{(\gamma)}(\beta_0, t)$$

and

$$A_n = \left\{ \max_i \left| \frac{w_i(t, \mathbf{r})}{n(t)} - \overline{w}_i(t, \mathbf{r}) \right| \leq \frac{\delta}{2} \right\}.$$

Condition 5 is implied by Lemma 1, and Condition 6 by Lemma 2. An argument as in Goldstein and Langholz [(1992), Lemma 4] shows that (7.1) implies that $\mathbf{\Sigma}_\alpha$ given by (7.2) is positive definite and so Condition 3 follows. □

Acknowledgments. This research was initiated when Ørnulf Borgan and Bryan Langholz were on sabbatical leave at the MRC Biostatistics Unit, Cambridge, England, during the academic year 1991–1992. We acknowledge the MRC Biostatistics Unit for its hospitality and for providing use with the best working facilities during this year.

REFERENCES

Andersen, P. K., Borgan, Ø., Gill, R. D. and Keiding, N. (1993). *Statistical Models Based on Counting Processes*. Springer, New York.

Andersen, P. K. and Gill, R. D. (1982). Cox's regression model for counting processes: a large sample study. *Ann. Statist.* **10** 1100–1120.

Arjas, E. (1989). Survival models and martingale dynamics (with discussion). *Scand. J. Statist.* **16** 177–225.

Borgan, Ø. and Langholz, B. (1993). Nonparametric estimation of relative mortality from nested case-control studies. *Biometrics* **49** 593–602.

Brémaud, P. (1981). *Point Processes and Queues: Martingale Dynamics*. Springer, New York.

BRESLOW, N. E. and DAY, N. E. (1987). *Statistical Methods in Cancer Research* **2**. *The Design and Analysis of Cohort Studies*. International Agency for Research on Cancer, Lyon.

COX, D. R. (1972). Regression models and life-tables (with discussion). *J. Roy. Statist. Soc. Ser. B* **34** 187–220.

FLEMING, T. R. and HARRINGTON, D. P. (1991). *Counting Processes and Survival Analysis*. Wiley, New York.

GOLDSTEIN, L. and LANGHOLZ, B. (1992). Asymptotic theory for nested case-control sampling in the Cox regression model. *Ann. Statist.* **20** 1903–1928.

HJORT, N. L. and POLLARD, D. (1993). Asymptotics of minimisers of convex processes. Statistical Research Report 5/93, Institute of Mathematics, Univ. Oslo.

KARR, A. F. (1991). *Point Processes and Their Statistical Inference*, 2nd ed. Dekker, New York.

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

LUBIN, J. H. and GAIL, M. H. (1984). Biased selection of controls for case-control analyses of cohort studies. *Biometrics* **40** 63–75.

OAKES, D. (1981). Survival times: aspects of partial likelihood (with discussion). *Internat. Statist. Rev.* **49** 235–264.

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

RAO, R. R. (1963). The law of large numbers for $D[0,1]$-valued random variables. *Theory Probab. Appl.* **8** 70–74.

ROBINS, J. M., GAIL, M. H. and LUBIN, J. H. (1986). More on "Biased selection of controls for case-control analysis of cohort studies." *Biometrics* **42** 293–299.

ROBINS, J. M., ROTNITZKY, A. and ZHAO, L. P. (1994). Estimation of regression coefficients when some regressors are not always observed. *J. Amer. Statist. Assoc.* **89** 846–866.

SCHUSTER, E. F. and SYPE, W. R. (1987). On the negative hypergeometric distribution. *Internat. J. Math. Ed. Sci. Tech.* **18** 453–459.

THOMAS, D. C. (1977). Addendum to "Methods of cohort analysis: appraisal by application to asbestos mining," by F. D. K. Liddell, J. C. McDonald and D. C. Thomas. *J. Roy. Statist. Soc. Ser. A* **140** 469–491.

Ø. BORGAN
DEPARTMENT OF MATHEMATICS
UNIVERSITY OF OSLO
P.O. BOX 1053 BLINDERN
N-0316 OSLO 3
NORWAY

L. GOLDSTEIN
DEPARTMENT OF MATHEMATICS
UNIVERSITY OF SOUTHERN CALIFORNIA
LOS ANGELES, CALIFORNIA 90089-1113

B. LANGHOLZ
DEPARTMENT OF PREVENTIVE MEDICINE
UNIVERSITY OF SOUTHERN CALIFORNIA
LOS ANGELES, CALIFORNIA 90033-1042