

A Pseudo-Likelihood Approach to Analysis of Nested Case-Control Studies

Sven Ove Samuelsen

Department of Mathematics, University of Oslo
P.O. Box 1053 Blindern, 0316 Oslo, Norway

email: osamuels@math.uio.no

Abstract

In nested case-control studies the controls are sampled from the risk set at the failure times of the case. The analytical basis for such studies has been limited to semi-parametric estimators under proportional hazard models. In this paper it is observed that conditional inclusion probabilities of ever being included in the nested case-control study can be obtained, here conditioning is on the information needed to carry out a nested case-control study. The inclusion probabilities are used in pseudo-likelihoods by weighting by their inverse. This makes it possible to fit parametric regression models. Also a new semi-parametric estimator under the proportional hazard model is obtained. The methods are illustrated by simulation experiments and by application to a dataset.

1 Introduction

The two major types of epidemiological study designs are cohort studies and case-control studies. Breslow & Day, (1980, 1987) discuss the advantages and problems associated with both types of studies. With the intent to retain the virtues of both type of studies so-called case-control within cohort studies have been constructed.

The most commonly applied of these study designs is probably the nested case-control design suggested by Thomas (1977). In this design controls are selected from the population at risk at the event times. Subsequent samples of controls are independent.

⁰*Key words:* Epidemiology; Inclusion probabilities; Nested case-control studies; Parametric regression; Partial likelihood; Pseudo-likelihood; Semi-parametric regression; Survival analysis.

⁰*Abbreviated title:* Pseudo-likelihood for nested case-control.

Thomas suggested, under a proportional hazard model, an estimator of relative risks maximizing a Cox-type likelihood which is in fact a partial likelihood (Oakes, 1982). The asymptotic properties of the estimator have been formally derived (Goldstein & Langholz, 1992, Borgan, Langholz & Goldstein, 1995) using counting process / martingale theory (Andersen et al., 1993). Recent papers also extend the nested case-control design to counter matching, i.e. stratified sampling of the controls (Langholz & Borgan, 1995), and show how base-line hazards and local standardized mortality ratios (Borgan & Langholz, 1993) as well as excess risk models (Borgan & Langholz, 1995) may be fitted.

Another type of case-control study within a cohort study is the case-cohort design of Prentice (1986) in which a subcohort is randomly chosen from the complete cohort at the outset of the study. Covariate information is obtained for the cases at their failure times and on a follow-up basis for the subcohort. Under a proportional hazard model Prentice showed how to obtain estimates of relative risks also by maximizing a Cox-type likelihood. This “likelihood” is however not a partial likelihood because each case is compared to the subcohort, thus the control sets at different event times are dependent. Prentice used the term pseudo-likelihood for the function that was maximized.

The pseudo-likelihood idea has been further discussed in relation to case-control studies within a cohort framework. Kalbfleisch & Lawless (1988) showed how to fit parametric and semi-parametric regression models when cases and controls are sampled with specified probabilities. For the parametric case the estimators are obtained by maximizing the pseudo-log-likelihood $\sum l_i/p_i$ where l_i is the log-likelihood contribution and p_i the probability of sampling individual i . The sum is taken over sampled cases and controls. In the semi-parametric case it was pointed out how one may extend the pseudo-likelihood of Prentice. In the parametric setting Kalbfleisch & Lawless also suggested the asymptotic properties of their pseudo-likelihood estimator when individuals are sampled independently of each other. Flanders & Greenland (1991) extended this to the simple random sampling of cases and controls. Similar issues have been discussed in Samuelsen (1989).

The point of this paper is to note how pseudo-likelihoods similar to those of Kalbfleisch & Lawless can be constructed for the nested case-control design. The essential idea here is that the conditional probability that an individual will ever be chosen as a control in a nested case-control study can be obtained. These inclusion probabilities have a simple representation resembling Kaplan-Meier estimates. The conditioning here is on who is in the risk set at failure times and on which individuals become cases. In order to perform nested case-control studies this skeleton of the cohort information is required. Also noting that conditional on this skeleton cases are included with probability 1 the pseudo-likelihood idea can be applied directly.

The outline of the paper is as follows. In the next section notation and models for the complete cohort study are introduced. Furthermore the sampling to the nested case-control data is described and the (conditional) probabilities for ever being included in the nested case-control study are derived. It is also described how pseudo-likelihoods corresponding to conventional likelihoods and to the Cox-partial likelihood are constructed. In Section 3 the asymptotic variance of the normalized pseudo-score and the asymptotic value of the information matrix is derived and hence the asymptotic distribution of the pseudo-likelihood estimators are motivated. Estimators for variance of the pseudo-likelihood estimators are also given. In Section 4 a small simulation study is presented. In Section 5 the methods is employed on data from a cohort study. The paper ends with a discussion on some extensions and modifications.

2 Mathematical description of the cohort study and nested case-control study

2.1 The cohort study

We assume that the cohort study consists of n individuals and that individual no. i enters the study at age b_i , and is followed up until time c_i . The individuals are assumed not to have the disease under investigation when they enter the study. If the disease is developed at age T_i and $b_i < T_i \leq c_i$ this is recorded in the cohort study. Also if $T_i > c_i$ then the T_i is right censored at c_i . Associated with individual no. i is a p -vector of possibly time-dependent covariates $Z_i(t) = (Z_{1i}(t), \dots, Z_{pi}(t))$. Models for the T_i are typically formulated through the hazard functions $\lambda_i(t)$ dependent on the covariates $Z_i(t)$. We will in particular discuss the proportional hazard model

$$\lambda_i(t) = \exp(\beta' Z_i(t)) \lambda_0(t) \quad (2.1)$$

where $\beta' = (\beta_1, \dots, \beta_p)$ is a regression parameter and $\lambda_0(t)$ some common baseline hazard function. Another model of interest is the accelerated hazard model for which

$$\lambda_i(t) = \exp(\beta' Z_i(t)) \lambda_0(t \exp(\beta' Z_i(t))) \quad (2.2)$$

where again β is a regression parameter and $\lambda_0(t)$ is a common hazard function. A further choice might be additive models $\lambda_i(t) = \beta_0(t) + \sum_{j=1}^p \beta_j(t) Z_{ji}(t)$ where $\beta(t)' = (\beta_0(t), \dots, \beta_p(t))$ may be functions (Aalen, 1980, 1989). The point of mentioning two models in addition to the proportional hazard model is to emphasize that (2.1) may very well give a poor fit and an inappropriate interpretation of particular data. Methods that are not restricted to the model (2.1) is thus of considerable importance.

The model (2.1) is frequently estimated by maximizing the Cox partial likelihood

$$L_C(\beta) = \prod_{i=1}^n \left[\frac{\exp(\beta' Z_i(X_i))}{\sum_{j \in \mathcal{R}_i} \exp(\beta' Z_j(X_i))} \right]^{D_i} = \prod_{i=1}^n \left[\frac{\exp(\beta' Z_i(X_i))}{S^{(0)}(\beta, X_i)} \right]^{D_i} \quad (2.3)$$

where $X_i = \min(T_i, c_i)$, $D_i = I(X_i = T_i)$, $\mathcal{R}_i = \{j : X_j \geq X_i > b_j\}$ and

$$S^{(0)}(\beta, t) = \sum_{j: b_j < t \leq X_j} \exp(\beta' Z_j(t)).$$

If we specify the hazard rate as $\lambda_0(t) = \lambda_0(t; \theta)$ for some parameter θ , e.g. as a Weibull hazard or a piecewise constant hazard, the maximum likelihood estimates (MLE) of (β, θ) under models (2.1) and (2.2) are obtained by maximizing the log-likelihood (minus a constant depending on the censoring)

$$l(\beta, \theta) = \sum_{i=1}^n [D_i \log(\lambda_i(X_i)) - \int_{b_i}^{X_i} \lambda_i(s) ds]. \quad (2.4)$$

Of course the log-likelihood (2.4) is valid for all other parametric regression models defined by $\lambda_i(t) = \lambda_i(t; \theta)$. For instance one may specify the additive models mentioned above parametric letting $\beta_j(t) = \beta_j(t, \theta_j)$.

2.2 The nested case-control study

For the description of the sampling of the controls we follow Goldstein & Langholz (1992). Within their framework the cohort is followed up prospectively with respect to membership in the cohort and with respect to occurrence of disease. Thus the risk sets \mathcal{R}_i for all X_i are known, in particular when $D_i = 1$. At the times X_i we thus know $Y_i = \#\mathcal{R}_i$, i.e. the number of individuals at risk in the cohort. For each of the X_i with $D_i = 1$ there will now be sampled without replacement a set $\widetilde{\mathcal{R}}_{i0} = \{j_{i1}, \dots, j_{im}\}$ of individuals (where $m(< Y_i)$) from $\mathcal{R}_i \setminus \{i\}$. The sets $\widetilde{\mathcal{R}}_{i0}$ are assumed to be independent (for the i with $D_i = 1$). Letting $\widetilde{\mathcal{R}}_i = \widetilde{\mathcal{R}}_{i0} \cup \{i\}$ it is assumed that the covariate values $Z_j(X_i)$ are obtained when $j \in \widetilde{\mathcal{R}}_i$. An estimate of the regression parameter in proportional hazard model (2.1) is now obtained by maximizing the following partial likelihood, suggested by Thomas (1977),

$$\widetilde{L}_C(\beta) = \prod_{i=1}^n \left[\frac{\exp(\beta' Z_i(X_i))}{\sum_{j \in \widetilde{\mathcal{R}}_i} \exp(\beta' Z_j(X_i))} \right]^{D_i} \quad (2.5)$$

i.e. the same expression as (2.3) only with the summation in the denominator being over $\widetilde{\mathcal{R}}_i$ rather than over \mathcal{R}_i . This is in fact a partial likelihood. Relevant literature is mentioned in the Introduction. Note also the formal equality of $\widetilde{L}_C(\beta)$ with the conditional likelihood for matched case-control studies (Breslow & Day, 1980).

2.3 The inclusion probabilities in a nested case-control study

As the risk sets \mathcal{R}_i and the D_i need to be known in order to carry out the nested case-control study we also know

$$\mathcal{F} = \{(b_i, X_i, D_i); i = 1, \dots, n\}.$$

Conditionally on this “skeleton” of the cohort \mathcal{F} the probability that individual j is ever being selected as a control in the nested case-control study is given by

$$p_{0j} = 1 - \prod_{b_j < X_i < X_j} \left[1 - \frac{m}{Y_i - 1} D_i\right]. \quad (2.6)$$

This follows because the probability of being selected at time X_i is $(mD_i)/(Y_i - 1)$ if $b_j < X_i < X_j$ and because the sampled control sets $\widetilde{\mathcal{R}}_{i0}$ are independent. Equation (2.6) is the key result of this paper.

In order to make use of this result we make the following

- **Assumption:** The complete covariate histories are obtained for both cases and controls.

Thus we need to know the *functions* $Z_j(t)$ for $b_j < t \leq X_j$ if individual j is either a case or a control. This is trivially fulfilled if the covariates are constants. In general the assumption appears not to be very restrictive. Valid methods of inference can be constructed if the assumption is weakened to requiring that the covariate histories are obtained only for the controls, but it is hard to imagine that this can be of practical importance. We will in the following make use of the probabilities, conditional on \mathcal{F} ,

$$p_j = \begin{cases} 1 & D_j = 1 \\ p_{0j} & D_j = 0 \end{cases}$$

of being included in the nested case-control study. We will also let V_{j0} be the indicator that individual j is ever selected as a control, and $V_j = \max(D_j, V_{j0})$ the indicator that individual j is ever either a case or a control. The set of all controls and cases ever sampled in the nested case-control study is denoted $\widetilde{\mathcal{R}}$, i.e. $\widetilde{\mathcal{R}} = \cup \widetilde{\mathcal{R}}_i$ where the union is taken over all i such that $D_i = 1$.

2.4 The pseudo-likelihood approach

Similarly to Kalbfleisch & Lawless (1988) we now suggest to fit models that result in a log-likelihood like (2.4), e.g. models (2.1) and (2.2) with a parametric specified $\lambda_0(t) = \lambda_0(t; \theta)$, by maximizing

$$l_{(p)} = \sum_{i \in \widetilde{\mathcal{R}}} \frac{1}{p_i} l_i = \sum_{i=1}^n \frac{V_i}{p_i} l_i \quad (2.7)$$

where

$$l_i = D_i \log(\lambda_i(X_i)) - \int_{b_i}^{X_i} \lambda_i(s) ds$$

i.e. the log-likelihood contribution of individual i in the cohort.

Following Kalbfleisch & Lawless we use the term pseudo-log-likelihood for (2.7). We will also call derivatives of pseudo-log-likelihoods for pseudo-score functions and second derivatives for pseudo-information matrices.

For the proportional hazard model we further suggest that regression parameters be estimated by maximizing the following pseudo-partial-likelihood

$$\tilde{L}_{(p)}(\beta) = \prod_{i=1}^n \left[\frac{\exp(\beta' Z_i(X_i))}{\tilde{S}^{(0)}(\beta, X_i)} \right]^{D_i} \quad (2.8)$$

where

$$\tilde{S}^{(0)}(\beta, X_i) = \sum_{j \in \mathcal{R}_i} \frac{V_j}{p_j} \exp(\beta' Z_j(X_i)) = \sum_{j \in \tilde{\mathcal{R}}_i} \frac{1}{p_j} \exp(\beta' Z_j(X_i)).$$

Similar estimators were suggested in Prentice (1986), Prentice & Self (1988) and Kalbfleisch & Lawless (1988) for the case-cohort design. The estimators of Prentice and of Self & Prentice are somewhat simpler since the weighting by the inverse of the inclusion probabilities can be omitted. The pseudo-score function for Prentice's estimator has expectation zero just like the score-function from the usual estimator for nested case-control studies obtained from (2.5).

Kalbfleisch & Lawless consider strata in which all individuals have the same probability of being selected, whereas in our situation these probabilities will generally differ from individual to individual. In our situation the p_j are functions of \mathcal{F} and hence random variables. Although Kalbfleisch & Lawless consider given probabilities of being selected in the different strata the membership in for instance the stratum of cases is random. Thus the probability that an individual is being selected is a random variable also in the setting of Kalbfleisch & Lawless. This problem is not explicitly discussed by those authors and a formal justification of several of their results requires arguments similar to those that will be presented in the next section. In the setting of Kalbfleisch & Lawless the indicators of being selected are independent. This is in contrast with our situation and will make our calculation of the covariance matrices of the score functions more difficult also from a large sample perspective.

3 Outline of asymptotic properties

3.1 The inclusion probabilities

Let

$$\mathcal{G} = \{[(b_i, X_i, D_i, Z_i(t)); b_i < t \leq X_i], i = 1, \dots, n\}$$

i.e. \mathcal{G} denotes the complete information in the cohort study. The skeleton of the cohort \mathcal{F} is contained in \mathcal{G} , but all information regarding the sampling is given by \mathcal{F} . We now have

$$E\left[\frac{V_{0i}}{p_{0i}}|\mathcal{G}\right] = E\left[\frac{V_{0i}}{p_{0i}}|\mathcal{F}\right] = 1 \text{ and similarly } E\left[\frac{V_i}{p_i}|\mathcal{G}\right] = E\left[\frac{V_i}{p_i}|\mathcal{F}\right] = 1.$$

Define $Y(t)$ as the number of individuals at risk and $\mathcal{R}(t)$ the risk set at time $t-$. When the cohort size n is large the probability of being included in the nested case-control study as a control for a particular individual i can then be approximated by

$$p_{0i} \approx 1 - \exp\left(-m \int_{b_i}^{X_i} \lambda(t) dt\right)$$

under the assumptions that $\inf_t Y(t)$ goes to infinity and that for some function $\lambda(t)$,

$$\frac{1}{Y(t)} \sum_{i \in \mathcal{R}(t)} \lambda_i(t) \rightarrow \lambda(t)$$

uniformly in probability as $n \rightarrow \infty$. This is so because under the stated assumptions

$$\prod_{X_i < t} \left[1 - \frac{m}{Y_i - 1} D_i\right] \rightarrow \exp\left(-m \int_{X_k < t} \lambda(s) ds\right)$$

uniformly in probability. This limit is obtained along the lines of the proof of consistency for the Kaplan-Meier estimator (Andersen et al., 1993).

We will furthermore need the conditional (on \mathcal{G} and on \mathcal{F}) probabilities p_{0ij} where $i \neq j$ that both individual i and individual j are selected as controls, or equivalently, the conditional probabilities q_{0ij} that neither i nor j are selected as controls. Letting $V_{0i}(t)$ be the indicator that individual i is selected as a control at time t we thus need

$$q_{0ij} = E[(1 - V_{0i})(1 - V_{0j})|\mathcal{F}] = E\left[\prod_{t,s} (1 - V_{0i}(t))(1 - V_{0j}(s))\right]|\mathcal{F}$$

Now, the control samples at different times are independent. Hence, if $t \neq s$ then $V_{0i}(t)$ and $V_{0j}(s)$ are independent. Also if $t \leq b_j$ or $t \geq X_j$ only individual i may be sampled as a control and similarly for individual j . Thus

$$q_{0ij} = \prod E[1 - V_{0i}(t)|\mathcal{F}]E[1 - V_{0j}(s)|\mathcal{F}] \prod E[(1 - V_{0i}(t))(1 - V_{0j}(t))|\mathcal{F}]$$

where the first product is taken over those (t, s) such that $b_i < t < X_i$, $b_j < s < X_j$ and $t \neq s$ and the second over those t in $\langle m_{1ij}, m_{2ij} \rangle$ where $m_{1ij} = \max(b_i, b_j)$ and $m_{2ij} = \min(X_i, X_j)$. If $m_{2ij} \leq m_{1ij}$ then the second product is replaced by 1. Now if $m_{1ij} < X_k < m_{2ij}$ we have

$$E[(1 - V_{0i}(X_k))(1 - V_{0j}(X_k))|\mathcal{F}] = 1 - 2\frac{m}{Y_k - 1} + \frac{m}{Y_k - 1} \frac{m - 1}{Y_k - 2}$$

for k with $D_k = 1$. Thus with

$$\rho_{ij} = \frac{q_{0ij}}{(1 - p_{0i})(1 - p_{0j})} - 1,$$

we obtain $\rho_{ij} = 0$ if $m_{2ij} \leq m_{1ij}$ and

$$\rho_{ij} = \prod [1 - 2\frac{mD_k}{Y_k - 1} + \frac{m(m - 1)D_k}{(Y_k - 1)(Y_k - 2)}] / [1 - \frac{mD_k}{Y_k - 1}]^2 - 1 \quad (3.1)$$

otherwise. Again the product is taken over k such that $m_{1ij} < X_k < m_{2ij}$. The interpretation of ρ_{ij} is given by

$$Cov(V_{0i}, V_{0j}|\mathcal{G}) = Cov(1 - V_{0i}, 1 - V_{0j}|\mathcal{F}) = \rho_{ij}(1 - p_{0i})(1 - p_{0j}) \quad (3.2)$$

Equation (3.1) and (3.2) both imply that $\rho_{ij} \leq 0$. In Appendix A it is shown that asymptotically we have $\rho_{ij} = -(m/n) \int \lambda(t)/y(t)dt + O_p(n^{-2}) = O_p(n^{-1})$ if in addition to the assumptions used for the asymptotic considerations of p_{0i} we also have $Y(t)/n \rightarrow y(t)$ uniformly in probability where for some $\epsilon > 0$ it hold that $\inf_t y(t) > \epsilon$. This result will imply that though the V_{0i} are asymptotically uncorrelated the dependence may contribute also in the large sample variance of the pseudo-likelihood estimators.

3.2 The pseudo-likelihood estimator

In this subsection asymptotic properties of the pseudo-likelihood estimator (PSMLE) are addressed. We will obtain a formal requirement for consistency when assuming that the complete log-likelihood is concave. For asymptotic normality of the PSMLE we will rely on heuristics. Expressions for the covariance matrix of the pseudo-score and the expectation of the pseudo-information matrix are derived. Then assuming that the pseudo-score function divided by \sqrt{n} has a limiting normal distribution and that the usual Taylor-argument hold the approximate normal distribution of the PSMLE is determined by these two matrices and the true parameter value. The expressions for the covariance matrix of the pseudo-score and the expectation of the

pseudo-information matrix have natural estimators based on the data from the nested case-control study. From these one immediately obtains estimators of the covariance matrix of the PSMLE.

Let $\gamma = (\beta, \theta)$ summarize the $q(> p)$ dimensional parameters of the model and $\tilde{\gamma}' = (\tilde{\beta}', \tilde{\theta}')$ the PSMLE of γ . We now have

$$El_{(p)}(\gamma) = E \sum_{i=1}^n l_i(\gamma) E\left[\frac{V_i}{p_i} | \mathcal{G}\right] = E \sum_{i=1}^n l_i(\gamma) E\left[\frac{V_i}{p_i} | \mathcal{F}\right] = E \sum_{i=1}^n l_i(\gamma) = El(\gamma). \quad (3.3)$$

We assume that the complete log-likelihood $l(\gamma)$ is concave, that $l(\gamma)/n \rightarrow l_0(\gamma)$ in probability for each γ and that the true parameter value γ_0 is the unique maximum of $l_0(\gamma)$. Using the argument of Andersen & Gill (1982) it turns out that the MLE $\hat{\gamma}$ is consistent. But since $l(\gamma)$ is concave it also holds true that the pseudo-likelihood $l_{(p)}(\gamma)$ is concave. This is so because the pseudo-likelihood will coincide with a complete likelihood when observing $1/p_i$ individuals with the same covariates $Z_i(\cdot)$, being observed for the same period $< b_i, X_i]$ and having the same indicator D_i as individual i (among those with $V_i = 1$). If now for all γ

$$\frac{1}{n}(l_{(p)}(\gamma) - l(\gamma)) = \frac{1}{n} \sum_{i=1}^n \left(\frac{V_i}{p_i} - 1\right) l_i(\gamma) \xrightarrow{P} 0 \quad (3.4)$$

then we also have pointwise convergence of $l_{(p)}(\gamma)/n \rightarrow l_0(\gamma)$ in probability and a second application of the argument of Andersen & Gill gives consistency of the PSMLE $\tilde{\gamma}$. From (3.3) we see that (3.4) hold if

$$\text{Var}\left(\frac{1}{n}(l_{(p)}(\gamma) - l(\gamma))\right) \rightarrow 0.$$

Since $\text{Cov}(V_i, V_j | \mathcal{G}) \leq 0$ we also have, similarly to (3.7) below, that

$$\text{Var}\left(\frac{1}{n}(l_{(p)}(\gamma) - l(\gamma))\right) < H_n(\gamma) = \frac{1}{n^2} E \sum_{i=1}^n l_i^2(\gamma) \frac{1 - p_i}{p_i}$$

and thus $H_n(\gamma) \rightarrow 0$ is sufficient to ensure consistency. It is however far from a necessary condition, indeed (3.4) alone without assuming a concave log-likelihood provide a strong heuristic argument for consistency of $\tilde{\gamma}$.

Turning towards the covariance matrix of the PSMLE let $u(\gamma)$ be the score-function and $I(\gamma)$ be the observed information of $l(\gamma)$. Similarly let $u_{(p)}(\gamma)$ and $I_{(p)}(\gamma)$ be the pseudo-score-function and the pseudo-information. We will also need the contribution $u_i(\gamma)$ to the score function from individual no. i . Now $u(\gamma) = \sum_{i=1}^n u_i(\gamma)$ and $u_{(p)}(\gamma) = \sum_{i=1}^n V_i u_i(\gamma)/p_i$ and similar equations hold true for I and $I_{(p)}$. At the true parameter

value γ_0 we write $l = l(\gamma_0)$, $u = u(\gamma_0)$, $I = I(\gamma_0)$, $u_{(p)} = u_{(p)}(\gamma_0)$, $I_{(p)} = I_{(p)}(\gamma_0)$ and $u_i = u_i(\gamma_0)$.

Similarly to (3.3) we have that

$$Eu_{(p)}(\gamma) = Eu(\gamma) \text{ and } EI_{(p)}(\gamma) = EI(\gamma) \quad (3.5)$$

and also that $u_{(p)}$ and $u_{(p)} - u$ are uncorrelated, thus

$$VAR(u_{(p)}) = VAR(u) + VAR(u_{(p)} - u) \quad (3.6)$$

where $VAR(X)$ denotes the covariance matrix of X . The last term in (3.6) can be written, with $v^{\otimes 2} = vv^t$,

$$\Delta_n = VAR(u_{(p)} - u) = E\left[\sum_{i=1}^n u_i^{\otimes 2} Var\left(\frac{V_i}{p_i} | \mathcal{G}\right) + \sum_{i \neq j} u_i u_j^t Cov\left(\frac{V_i}{p_i}, \frac{V_j}{p_j} | \mathcal{G}\right)\right],$$

thus

$$\Delta_n = E\left[\sum_{i=1}^n u_i^{\otimes 2} \frac{1-p_i}{p_i} + \sum_{i \neq j} u_i u_j^t \frac{(1-p_i)(1-p_j)\rho_{ij}}{p_i p_j}\right]. \quad (3.7)$$

Also letting $\Sigma_n = VAR(u) = EI = EI_{(p)}$ we conjecture that $\tilde{\gamma}$ has an approximate normal distribution with expectation γ_0 and covariance matrix $\Sigma_n^{-1} + \Sigma_n^{-1}\Delta_n\Sigma_n^{-1}$. Letting $n \rightarrow \infty$ we will under weak assumptions have that $n^{-1}\Sigma_n \rightarrow \Sigma$ and $n^{-1}\Delta_n \rightarrow \Delta$ for semi-positive matrices Σ and Δ . We will assume that Σ and Σ_n are positive definite. Note here that the second term in $n^{-1}\Delta_n$ generally will not vanish since $\rho_{ij} = O_p(n^{-1})$, thus the dependence between the sampling indicators may matter also when the sample is large.

From the nested case-control study we may estimate the covariance matrix of $\tilde{\gamma}$ by estimates $\tilde{\Sigma}_n = I_{(p)}(\tilde{\gamma})$ of Σ_n and

$$\tilde{\Delta}_n = \sum_{i=1}^n V_i u_i(\tilde{\gamma})^{\otimes 2} \frac{1-p_i}{p_i^2} + \sum_{i \neq j} \frac{V_i V_j}{p_{0ij}} \frac{(1-p_i)(1-p_j)\rho_{ij}}{p_i p_j} u_i(\tilde{\gamma}) u_j(\tilde{\gamma})^t \quad (3.8)$$

of Δ_n . Here $p_{0ij} = E[V_{0i}V_{0j} | \mathcal{F}] = \rho_{ij}(1-p_{0i})(1-p_{0j}) + p_{0i}p_{0j}$. Since $\rho_{ij} = O_p(n^{-1})$ we can substitute p_{0ij} by $p_i p_j$ if n is large. Note that in Δ_n only controls contribute to the sums and it thus does not matter if we write p_i or p_{0i} .

In order to formalize the above argument it appears natural to follow the approach of Self & Prentice (1988). Let $F(x; \Gamma)$ denote the cumulative of the q -dimensional normal distribution with expectation zero and covariance matrix Γ . It is known that $\text{pr}(u/\sqrt{n} \leq x) \rightarrow F(x; \Sigma)$. Assume that also, when \mathcal{G}_n is the complete cohort information with n individuals,

$$\text{pr}\left(\frac{1}{\sqrt{n}}(u_{(p)} - u) \leq y | \mathcal{G}_n\right) \rightarrow F(y; \Delta).$$

Then

$$\begin{aligned} \text{pr}\left(\frac{1}{\sqrt{n}}u \leq x, \frac{1}{\sqrt{n}}(u_{(p)} - u) \leq y\right) &= EI\left\{\frac{1}{\sqrt{n}}u \leq x\right\} \text{pr}\left(\frac{1}{\sqrt{n}}(u_{(p)} - u) \leq y | \mathcal{G}_n\right) \\ &\rightarrow F(x; \Sigma)F(y; \Delta) \end{aligned} \quad (3.9)$$

by the dominated convergence theorem. Thus u/\sqrt{n} and $(u_{(p)} - u)/\sqrt{n}$ each converge to normal distributions and they are asymptotically independent. The problem here is proving the convergence of $(u_{(p)} - u)/\sqrt{n}$. In the situation of Self & Prentice (1988) the variables corresponding to the V_i were indicators of being sampled by simple random sampling. They thus could apply a finite population large sample result of Hajek(1961) discussed by e.g. Cochran (1977) and Lehman (1975). In the present situation the sampling distribution is rather nonstandard and it is not much hope of finding a result directly from the literature. However, conditionally on the cohort information \mathcal{G} , the sampling distribution of the V_{0i} is fairly simple and one would believe that it is possible to furnish a proof of such a result.

3.3 The pseudo partial likelihood estimator

We now turn to semi-parametric estimators in the proportional hazard model. Let, with $v^{\otimes 0} = 1, v^{\otimes 1} = v$ and $v^{\otimes 2} = vv^t$,

$$S^{(r)}(\beta, t) = \sum_{j \in \mathcal{R}(t)} Z_j^{\otimes r} \exp(\beta' Z_j) \text{ and } \tilde{S}^{(r)}(\beta, t) = \sum_{j \in \mathcal{R}(t)} \frac{V_j}{p_j} Z_j^{\otimes r} \exp(\beta' Z_j).$$

The score-function and the information matrix of the Cox-likelihood (2.3) then equals

$$U_C(\beta) = \sum_{i=1}^n \left[Z_i - \frac{S^{(1)}(\beta, X_i)}{S^{(0)}(\beta, X_i)} \right] D_i \text{ and } I_C(\beta) = \sum_{i=1}^n \left[\frac{S^{(2)}(\beta, X_i)}{S^{(0)}(\beta, X_i)} - \left(\frac{S^{(1)}(\beta, X_i)}{S^{(0)}(\beta, X_i)} \right)^{\otimes 2} \right] D_i$$

while the pseudo-score $\tilde{U}_{(p)}(\beta)$ and pseudo-information $\tilde{I}_{(p)}(\beta)$ corresponding to (2.8) are given by substituting $S^{(r)}(\beta, t)$ in the equations above by $\tilde{S}^{(r)}(\beta, t)$.

In this case we now longer have $E\tilde{U}_{(p)}(\beta) = EU(\beta)$ or $E\tilde{I}_{(p)}(\beta) = EI(\beta)$. Rather the equalities that correspond to (3.5) are

$$E\tilde{S}^{(r)}(\beta, t) = ES^{(r)}(\beta, t)$$

Pointwise convergence in probability of $n^{-1}(\tilde{S}^{(r)}(\beta, t) - S^{(r)}(\beta, t))$ to zero can be ensured by assumptions corresponding to those e.g. leading to (3.4). We will assume the stronger

$$\sup_t \left| \frac{1}{n} (\tilde{S}^{(r)}(\beta, t) - S^{(r)}(\beta, t)) \right| \rightarrow 0$$

in probability. As the standard conditions of convergence of the Cox-estimator require that $n^{-1}S^{(0)}(\beta, t)$ converge to a function bounded away from zero we then have

$$\frac{1}{n} (\ln(\tilde{L}_{(p)}(\beta)) - \ln(L_C(\beta))) = \frac{1}{n} \sum_{i=1}^n \ln\left(\frac{S^{(0)}(\beta, X_i)}{\tilde{S}^{(0)}(\beta, X_i)}\right) D_i \rightarrow 0 \quad (3.10)$$

in probability. The Cox-likelihood (2.3) is log-concave and the pseudo-partial likelihood (2.8) retains this property by the corresponding argument for concave pseudo-log-likelihoods. The pointwise convergence (3.10) implies that the maximizer of (2.8) $\tilde{\beta}$ is consistent by the same argument giving consistency of the usual Cox-estimator (Andersen & Gill, 1982).

Define $Y_j(t) = I\{b_j < t \leq X_j\}$. Now letting

$$W_j^0(\beta) = \sum_{i=1}^n \left[Z_j(X_i) - \frac{S^{(1)}(\beta, X_i)}{S^{(0)}(\beta, X_i)} \right] Y_j(X_i) \exp(\beta' Z_j(X_i)) \frac{D_i}{\tilde{S}^{(0)}(\beta, X_i)}$$

we find, with U_C and $\tilde{U}_{(p)}$ being the score and the pseudo-score and W_j^0 being $W_j^0(\beta)$ at the true parameter value that

$$\tilde{U}_{(p)} - U_C = \sum_{j=1}^n \left(1 - \frac{V_j}{p_j}\right) W_j^0.$$

It can be shown that with

$$W_j = \int \left[Z_j(t) - \frac{S^{(1)}(\beta, t)}{S^{(0)}(\beta, t)} \right] Y_j(t) \exp(\beta' Z_j(t)) \lambda_0(t) dt,$$

then $W_j - W_j^0 \rightarrow 0$ in probability. Under regularity assumptions it also hold that

$$\frac{1}{\sqrt{n}} [(\tilde{U}_{(p)} - U_C) - \sum_{j=1}^n \left(1 - \frac{V_j}{p_j}\right) W_j] = \frac{1}{\sqrt{n}} \sum_{j=1}^n \left(1 - \frac{V_j}{p_j}\right) (W_j^0 - W_j) \rightarrow 0$$

in probability. From this it follows that $n^{-\frac{1}{2}}\tilde{U}_{(p)}$ and

$$\frac{1}{\sqrt{n}} U_C + \frac{1}{\sqrt{n}} \sum_{j=1}^r \left(1 - \frac{V_j}{p_j}\right) W_j \quad (3.11)$$

has the same limiting distribution. Note that the W_j do not depend on the sampling, but only on the cohort information \mathcal{G} . The two terms in (3.11) both have expectation zero. Furthermore they are uncorrelated likewise u and $u_{(p)} - u$ in the parametric

setting. We have that the covariance matrix of the score U_C equals $\Sigma_n^C = EI_C(\beta_0)$ and that the covariance matrix of the second term in (3.11) is given by

$$\Delta_n^C = VAR\left(\sum_{j=1}^n \left(1 - \frac{V_j}{p_j}\right) W_j\right) = E\left[\sum_{i=1}^n W_i^{\otimes 2} \frac{1-p_i}{p_i} + \sum_{i \neq j} W_i W_j^t \frac{(1-p_i)(1-p_j)\rho_{ij}}{p_i p_j}\right].$$

By the assumptions of Andersen & Gill (1982) we have that $n^{-1}\Sigma_n^C \rightarrow \Sigma_C$ for some positive definite matrix Σ_C . It will also hold true that $n^{-1}\tilde{I}_{(p)}$ will converge to Σ_C . Furthermore $n^{-1}\Delta_n^C$ will converge to a semi-positive definite matrix Δ_C . Assuming that also

$$\frac{1}{\sqrt{n}} \sum_{j=1}^n \left(1 - \frac{V_j}{p_j}\right) W_j$$

is converging in distribution to normal distribution with expectation zero and covariance matrix Δ_C an argument corresponding to (3.9) shows that $n^{-1/2}\tilde{U}_{(p)}$ is also approximately normal with expectation zero and with covariance matrix $\Sigma_C + \Delta_C$. The usual Taylor expansion argument then gives that $\sqrt{n}(\tilde{\beta} - \beta)$ is approximately normal with expectation zero and covariance matrix $\Sigma_C^{-1} + \Sigma_C^{-1}\Delta_C\Sigma_C^{-1}$. Estimators of the matrices Σ_C and Δ_C are respectively $n^{-1}\tilde{I}_{(p)}(\tilde{\beta})$ and

$$n^{-1}\tilde{\Delta}_n^C = n^{-1}\left[\sum_{i=1}^n V_i(\tilde{W}_i^0)^{\otimes 2} \frac{1-p_i}{p_i^2} + \sum_{i \neq j} \frac{V_i V_j}{p_{0ij}} \frac{(1-p_i)(1-p_j)\rho_{ij}}{p_i p_j} \tilde{W}_i^0 (\tilde{W}_j^0)^t\right] \quad (3.12)$$

where $\tilde{W}_i^0 = W_i^0(\tilde{\beta})$. In the last equation we may, as for (3.8), substitute p_{0ij} by $p_{0i}p_{0j}$.

3.4 Estimators given by explicit expressions

By estimators given by explicit expressions we have in mind for instance the Kaplan-Meier estimator $\hat{F}(t) = 1 - \prod_{X_i \leq t} [1 - D_i/Y(X_i)]$ of a common distribution function $F(t)$, the Nelson-Aalen estimator $\hat{\Lambda}(t) = \sum_{X_i \leq t} D_i/Y(X_i)$ of the corresponding cumulative hazard function $\Lambda(t) = -\ln(1 - F(t))$ and the Breslow-estimator $\hat{\Lambda}_0(t) = \sum_{X_i \leq t} D_i/S^{(0)}(\hat{\beta}, X_i)$ of a cumulative baseline hazard under the proportional hazard model. Here $\hat{\beta}$ is the Cox-estimator obtained by maximizing (2.3).

Letting $\tilde{Y}(t) = \sum_{i=1}^n V_i Y_i(t)/p_i$ and substituting $Y(X_i)$ by $\tilde{Y}(X_i)$ in the Kaplan-Meier and Nelson-Aalen estimators we get estimators $\tilde{F}(t)$ and $\tilde{\Lambda}(t)$ of $F(t)$ and $\Lambda(t)$ using only data from the nested case-control study. Similarly substituting $S^{(0)}(\hat{\beta}, X_i)$ by $\tilde{S}^{(0)}(\tilde{\beta}, X_i)$ in $\hat{\Lambda}_0(t)$ we also arrive at an analogue $\tilde{\Lambda}_0(t)$ from the nested case-control study.

We refrain from going through the details in conjecturing the asymptotic properties of these estimators. This however can be done by arguments similar to those in the parametric and semi-parametric situations presented above. The conjectures, however, are stated below.

Let under an i.i.d. model $\lambda(t)$ equal the common hazard rate and let $y(t)$ be probability of being in the risk set at time t . Define $v_i(t) = \int_0^t Y_i(s)y(s)^{-1}\lambda(s)ds$ and

$$\Delta_n^v(t) = \sum_{i=1}^n v_i(t)^2 \frac{1-p_i}{p_i} + \sum_{i \neq j} v_i(t)v_j(t) \frac{(1-p_i)(1-p_j)\rho_{ij}}{p_i p_j}.$$

Assuming that $n^{-1}\Delta_n^v(t)$ is converging in probability to $\Delta^v(t)$ then $\sqrt{n}(\tilde{\Lambda}(t) - \Lambda(t))$ is approximately normal with expectation zero and variance $h(t) = \int_0^t y(s)^{-1}\lambda(s)ds + \Delta^v(t)$. Similarly $\sqrt{n}(\tilde{F}(t) - F(t))$ is also normal, having expectation zero and variance $(1 - F(t))^2 h(t)$. One may also derive corresponding expressions for the covariance functions of $\sqrt{n}(\tilde{\Lambda}(\cdot) - \Lambda(\cdot))$ and $\sqrt{n}(\tilde{F}(\cdot) - F(\cdot))$. Estimates of $\Delta^v(t)$ and hence the variance of $\tilde{F}(t)$ and $\tilde{\Lambda}(t)$ are obtained as equations similar to (3.8) and (3.12).

Actually the Kaplan-Meier and Nelson-Aalen estimators on the complete cohort is determined by the skeleton \mathcal{F} and these estimators can thus be calculated within the nested case-control study. The estimators $\tilde{S}(t)$ and $\tilde{\Lambda}(t)$ are then necessarily inferior. However, we may be interested in the incidence in some subgroup indexed by \mathcal{S} that is identified only on the cases and controls. This incidence may be estimated by $\tilde{S}(t)$ and $\tilde{\Lambda}(t)$ redefined taking the sums and products only over $i \in \mathcal{S}$. The same asymptotic properties follows only by redefining $y(t)$ as the probability of being under observation in the subgroup \mathcal{S} .

Under the proportional hazard model (2.1) let $\theta(t)$ be the asymptotic variance function of $\sqrt{n}(\hat{\Lambda}_0(t) - \Lambda_0(t))$. The explicit form of $\theta(t)$ can be found in e.g Andersen & Gill (1982) or Andersen et al. (1993). Let furthermore $s^{(0)}(\beta, t)$ the limit (in probability) of $n^{-1}S^{(0)}(\beta, t)$ and

$$w_i(t) = \int_0^t Y_i(s) \exp(\beta' Z_i) s^{(0)}(\beta, s)^{-1} \lambda_0(s) ds.$$

Define $\Delta^w(t)$ in the same way as $\Delta^v(t)$ above only substituting $v_i(t)$ by $w_i(t)$. Now $\sqrt{n}(\tilde{\Lambda}_0(t) - \Lambda_0(t))$ converge to a normal distribution with expectation zero and variance $\theta(t) + \Delta^w(t)$.

We mention that Borgan & Langholz (1995) present other estimators of $F(t)$, $\Lambda(t)$ and $\Lambda_0(t)$ in nested case-control studies. The relative merits of their estimators and those presented here is a worthwhile problem that have not yet been studied. Borgan & Langholz (1995) also discuss estimation of time-varying standardized mortality ratios. One may construct alternative estimators in the same spirit as $\tilde{\Lambda}(t)$ and $\tilde{\Lambda}_0(t)$.

4 A simulation study

The models that have been simulated are proportional hazard models with Weibull baseline hazards. This way it is possible to check how the pseudo-likelihood (2.7) may work with a parametric regression model and at the same time compare the performance of the usual partial likelihood estimator (2.5) with the pseudo-partial likelihood estimator (2.8). The issues that are considered are bias of the estimators and their variance estimators, normal approximation of the estimators and efficiency of the pseudo-likelihood estimators compared to the complete cohort estimators and to the partial likelihood estimator of nested case-control data. Efficiency is given both as asymptotic and empirical relative efficiency. Asymptotic relative efficiency (ARE) between two estimators is defined as the ratio between the average of the estimated variances of the estimators. Empirical relative efficiency (ERE) is similarly defined as the ratio between the empirical variances. It is also looked at coverage probabilities determined as the proportions of the simulations for which the true parameter value falls within the usual Wald 95 % confidence interval, i.e. the estimate ± 1.96 estimated standard deviations.

The models have one covariate Z_i sampled from a uniform $[0, 1]$ distribution. The model is thus given by $\lambda_i(t) = \exp(\beta Z_i + \gamma)\theta t^{\theta-1}$. The shape parameter θ is always put equal to 2, the regression parameter β is put equal to 1, the value zero is used for the intercept γ and the cohort size n is 1000. The simulation schemes differ with respect to the number of controls per case with values $m = 1$ or 3 and to the correspondence between the right censoring times C_i and the covariates Z_i . In Schemes 1 and 2 C_i and Z_i are independent whereas in Schemes 3 and 4 the C_i are proportional to Z_i . In both case the censoring times are uniformly distributed. The upper limit in the uniform distribution of C_i is chosen so that the expected number of cases is 125. The simulations are repeated 500 times.

The parameter and variance estimates were computed on the the simulated cohort in all simulations. This permits calculation of ARE and ERE relative to the cohort data. For the pseudo-likelihood estimator under the parametric model the efficiencies are relative to the parametric estimates, for both partial and pseudo-likelihood they are relative to the Cox-estimator. For ease of presentation other results on the cohort are omitted when they are not relevant to the behavior on the sample data.

The simulations were carried out in Splus. With the baseline hazard being the left unspecified estimates were obtained by the Cox-regression procedure in that program, directly on the cohort, by stratification on case-control set for the partial likelihood nested case-control estimator and by weighting by $1/p_i$ for the pseudo-likelihood estimator. With the Weibull baseline hazard a Splus function was written optimizing the parametric likelihood by the Newton-Raphson method. This function allowed for

weighting and it was thus possible to calculate the pseudo-likelihood estimates. To correct the covariance matrices calls to separately written C-routines calculating the matrices $\tilde{\Delta}_n$ and $\tilde{\Delta}_n^C$ were applied. The C-routines were necessary because for-loops in Splus proved to be impractically slow and a matrix approach besides being slower than the C-routines needed matrices having high dimensions. Appendix B contains some details.

In simulation scheme 1 the C_i are uniformly distributed on $[0, 0.5]$ and independent of Z_i . The number of controls per case m equals 1. The results from simulating this model are presented in Table 4.1.

Table 4.1. Averages, minimum and maximum of parameter estimates, average of variance estimates, empirical variances, coverage probabilities of 95% confidence intervals and asymptotic and empirical relative efficiencies from 500 simulations under Scheme 1 for pseudo-likelihood under parametric model and pseudo-likelihood and partial likelihood under semi-parametric model. ARE and ERE are relative to the cohort.

	Aver. est.	Min. est.	Max. est.	Aver. var.	Emp. var.	Cover. (%)	ARE	ERE
Parametric								
pseudo-likelihood,								
Shape par. θ	2.008	1.571	2.531	0.0267	0.0287	94.4	0.81	0.78
Intercept γ	0.013	-0.847	0.854	0.0919	0.1004	94.8	0.71	0.69
Regr. par. β	0.999	-0.108	2.143	0.1845	0.1871	95.0	0.55	0.55
Semi-parametric								
pseudo-likelihood								
Regr. par. β	1.002	-0.119	2.196	0.1842	0.1896	94.6	0.47	0.46
Semi-parametric								
partial likelihood								
Regr. par. β	0.999	-0.231	2.460	0.2167	0.2232	94.6	0.55	0.54

There is no bias to be found on the parameter estimates in this scheme. The variance estimates appears slightly smaller than the unbiased empirical variance estimates, but they are within what may be explained by random variation. The coverage probabilities are in good agreement with a 95 % confidence level. The normal approximation seemed to be acceptable for all estimators. The results on the parametric and semi-parametric regression parameter are quite similar, indeed the empirical correlation between these two estimates is as high as 0.998. The corresponding correlations with

the partial likelihood estimator are about 0.92. Actually the partial likelihood estimator appears to be more dispersed than the two pseudo-likelihood estimators. We find that both the ARE and the ERE between the semi-parametric pseudo-likelihood and the partial likelihood estimators are 0.85. On the regression parameter the efficiencies relative to the cohort were in good agreement on the case-control rule of thumb $m/(m+1) = 0.5$, but the other parameters seem to be estimated more efficiently.

Scheme 2 is the same as Model 1 for the cohort but in this case the number of controls m equals 3. Results for this model is given in Table 4.2.

Table 4.2. Averages, minimum and maximum of parameter estimates, average of variance estimates, empirical variances, coverage probabilities of 95% confidence intervals and asymptotic and empirical relative efficiencies from 500 simulations under Scheme 2 for pseudo-likelihood under parametric model and pseudo-likelihood and partial likelihood under semi-parametric model. ARE and ERE are relative to the cohort.

	Aver. est.	Min. est.	Max. est.	Aver. var.	Emp. var.	Cover. (%)	ARE	ERE
Parametric								
pseudo-likelihood,								
Shape par. θ	2.024	1.553	2.629	0.0232	0.0248	94.6	0.94	0.93
Intercept γ	0.034	-0.753	0.830	0.0724	0.0783	93.6	0.90	0.91
Regr. par. β	0.989	-0.060	1.929	0.1211	0.1162	95.6	0.83	0.88
Semi-parametric								
pseudo-likelihood								
Regr. par. β	0.988	-0.051	1.935	0.1212	0.1164	95.4	0.83	0.88
Semi-parametric								
partial likelihood								
Regr. par. β	0.998	-0.148	2.095	0.1364	0.1315	96.2	0.74	0.78

In Scheme 2 it appears that there is some bias on the shape parameter and intercept estimates. However, this bias is present also on the cohort. Again there are no indications that the variance estimates are biased, the coverage percentages of the confidence intervals are in fair agreement with the 95% confidence level and the distributions of the estimates seems to be well approximated by normal distributions. The efficiencies relative to the cohort increases as would be expected, but the pseudo-likelihood estimates of the regression parameter are now better than what the $m/(m+1)$ rule would indicate. Again the semi-parametric pseudo-likelihood estimate

is efficient to the partial likelihood estimate (ARE/ERE=0.89).

In Scheme 3 the we have the relation $C_i = Z_i/2.24$ between the censoring times C_i and the covariate Z_i . The number of controls equals 1. Results for the model is given in Table 4.3.

Table 4.3. Averages, minimum and maximum of parameter estimates, average of variance estimates, empirical variances, coverage probabilities of 95% confidence intervals and asymptotic and empirical relative efficiencies from 500 simulations under Scheme 3 for pseudo-likelihood under parametric model and pseudo-likelihood and partial likelihood under semi-parametric model. ARE and ERE are relative to the cohort.

	Aver. est.	Min. est.	Max. est.	Aver. var.	Emp. var.	Cover. (%)	ARE	ERE
Parametric								
pseudo-likelihood,								
Shape par. θ	2.020	1.578	2.515	0.0309	0.0313	94.6	0.99	0.99
Intercept γ	0.050	-1.892	1.948	0.3944	0.3763	95.4	0.81	0.81
Regr. par. β	0.968	-0.871	3.327	0.4500	0.4380	96.2	0.71	0.71
Semi-parametric								
pseudo-likelihood								
Regr. par. β	0.953	-0.981	3.451	0.4677	0.4573	95.8	0.72	0.72
Semi-parametric								
partial likelihood								
Regr. par. β	1.001	-1.239	3.834	0.7160	0.7033	96.4	0.47	0.47

The variances of the estimators are larger under this simulation scheme. This is so because the variation in the covariates is small for large failure times. Thus although the average estimate values deviate more from the true parameter values there is no obvious bias. Again variance estimates correspond well to empirical estimates and the coverage is in accordance with the confidence level. The normal approximations of the estimators still seem roughly to be justified, but there are signs of heavy right tails on the regression parameter estimators. The parametric estimate is somewhat more efficient than the semi-parametric, a trend also found on the cohort, and this is likely explained by the strong dependence between censoring time and covariate. Regarding efficiencies there is hardly any loss on the shape parameter. Also the pseudo-likelihood regression parameter estimators do well compared to the case-control rule of thumb. The efficiency of the usual nested case-control partial likelihood estimator compared

to the pseudo-likelihood estimators is now as small as 0.65.

Scheme 4 equals Scheme 3 for the cohort, but the number of controls m is like in Scheme 2 equal to 3.

Table 4.4. Averages, minimum and maximum of parameter estimates, average of variance estimates, empirical variances, coverage probabilities of 95% confidence intervals and asymptotic and empirical relative efficiencies from 500 simulations under Scheme 4 for pseudo-likelihood under parametric model and pseudo-likelihood and partial likelihood under semi-parametric model. ARE and ERE are relative to the cohort.

	Aver. est.	Min. est.	Max. est.	Aver. var.	Emp. var.	Cover. (%)	ARE	ERE
Parametric								
pseudo-likelihood,								
Shape par. θ	2.005	1.580	2.754	0.0300	0.0349	91.4	1.00	1.00
Intercept γ	-0.015	-1.688	1.910	0.3339	0.3840	94.4	0.94	0.95
Regr. par. β	1.033	-0.641	2.801	0.3490	0.3807	94.4	0.91	0.90
Semi-parametric								
pseudo-likelihood								
Regr. par. β	1.051	-0.564	2.931	0.3668	0.4004	95.6	0.91	0.90
Semi-parametric								
partial likelihood								
Regr. par. β	1.062	-0.773	3.235	0.4498	0.4852	93.6	0.74	0.75

The averages of the semi-parametric estimators in Scheme 4 borders to a significant bias, however they are not much worse than the average semi-parametric cohort estimator of 1.044. Similarly the coverage percentage of 91.6 on the shape parameter is significantly to small, but the corresponding number on the cohort is 92.2. The averages of the variances are all smaller than the empirical variances, assuming that $499S^2/\sigma^2 \sim \chi_{499}^2$ where S^2 is an empirical variance and σ^2 the true variance they are on the borderline of the confidence interval based on the empirical variances. The empirical and asymptotic efficiencies, however, are in good agreement, and as would be expected bigger than those of Model 3. The partial likelihood estimator is still less efficient than the semi-parametric pseudo-likelihood estimator, now with a factor about 0.8.

The behavior of the pseudo-likelihood estimators have also been studied in several

other situations. For one it has been looked at how the estimators perform when there is no effect of the covariate, i.e. $\beta = 0$. In order to compare with the previous results 4 new schemes, comparable to Schemes 1-4 respectively were carried out. Thus in all 4 the cohort size was 1000 and the expected number of cases were 125. The new schemes thus differed from each other with respect to the number of controls and whether the censoring times and covariates were independent. They also differed from Scheme 1-4 with respect to the value of γ . In general the results in the new schemes were quite similar to those of the previous schemes with respect to bias of parameter and variance estimators and with respect to normal approximations. The efficiency results were also quite comparable, but the gains for the semi-parametric pseudo-likelihood estimators over the partial likelihood counterpart were somewhat smaller.

It has furthermore been studied how the cohort size affects the results. New schemes were set up with cohort sizes of 5000, but with the expected number of cases still at 125. The regression parameter had values 1 and 0, censoring was independent or proportional to the covariate and the intercepts were adjusted accordingly. In these schemes there were only one control per case. Again there was no indication that the pseudo-likelihoods failed to work and the efficiency results were similar to those already presented. When censoring was independent of covariates the efficiency improvement of the pseudo-likelihood estimator relative to the partial likelihood estimator was negligible, but when $Z_i \propto C_i$ and $\beta = 1$ the improvement was still about 30%.

A couple of schemes were set up to see how the estimators performed with a very small expected number of cases, chosen to 25. The cohort size was 5000 and the regression parameter was 1 in both schemes. In the first scheme there was one control per case and the covariates and censoring times were independent. In the second scheme there were 3 controls per case and censoring was proportional to covariates. There were no indications of bias, but the distributions of the parameters began to show skewness and some heavy tails. In both situations the pseudo-likelihood estimators were less efficient compared to the cohort than in the previously presented models, but the semi-parametric pseudo-likelihood still was better than its partial likelihood counterpart.

In a last set of the schemes the situation in Scheme 1 was repeated only increasing the cohort size to 5000. The purpose of these runs were to check the performance of the variance estimates of normalized differences, say $\sqrt{n}(\tilde{\theta} - \theta_0)$. It turned out that the empirical variances of these variance estimates were about one fifth of those when the cohort size was 1000. This is in accordance with the variances of these variance estimates decreasing at the usual rate $1/n$.

5 Application on a data set

Eskild et al. (1993) report a follow-up study of 1009 intravenous drug users (IVDU) in Oslo, Norway. The drug users were included in the study when taking a test of having been infected with the human immunodeficiency virus (HIV). The aim of the study was to compare mortality rates between HIV positive and HIV negative IVDUs and at the same time adjust for potentially confounding variables. The HIV tests were taken in the period 1985-1991, follow up is until death or the end of the study period January 1st 1991 and time is measured as time since inclusion in the study. Time of death was obtained from the National Cause of Death Registry for totally 87 individuals. The mean follow up time was 3 years (range 1-67 months). In this paper only the covariates HIV serostatus (positive / negative) at inclusion in the study, age at inclusion in the study and sex are used. 18% of the IVDUs were HIV-positive, 36% were women and the mean age was 27 years (range 14-50). Here the data are reanalyzed using the developed and the usual nested case-control methodology. Of course the complete follow up data are available and our aim is only to give an example on how the pseudo-likelihood approach work on a real data set.

The data are analyzed under the proportional hazard model given by $\lambda_i(t)/\lambda_0(t) = \exp(\beta_1 Z_{1i} + \beta_2 Z_{2i} + \beta_3 Z_{3i})$ where Z_{1i} is age in years divided by 10, in order to get estimates of the same order of magnitude, and Z_{2i} and Z_{3i} binary covariates indicating respectively HIV-infection and women. The baseline hazard function is specified as a Weibull hazard $\lambda_0(t) = \exp(\gamma)\theta t^{(\theta-1)}$, an exponential hazard ($\theta = 1$) and is left unspecified. The cohort estimators are obtained under all three specifications. Nested case-control sampling is performed 500 times with 1 and with 3 controls per case.

When the baseline hazard was Weibull or left unspecified the Splus routines developed for the simulation study were used. Under the exponential hazard data are fitted with the generalized linear model procedure (with a Poisson family specification) in Splus. This procedure allows for weighting. The variances for the pseudo-likelihood estimators were calculated by help of calls to the C-routines used in the simulations.

Results are given in Tables 5.1. In column 1 and 2 of Table 5.1 the parameter estimates and the standard error estimates for the cohort data (SE_0) are given. (Under the semi-parametric model the same results are reported both as pseudo-likelihood and partial likelihood). Columns 3 and 7 give the average parameter estimates over the control samples for respectively 1 and 3 matched controls. Column 4 and 8 gives the square root of the average estimated variances SE_m^2 and column 5 and 9 gives the empirical standard deviations of the estimates S_m^2 over the control samples. From this empirical standard deviations $SE_m^* = (SE_0^2 + S_m^2)^{1/2}$ are computed. As notation indicates one would expect $SE_m \approx SE_m^*$ if the variance-estimates are unbiased.

Table 5.1. Estimated coefficients and standard errors (SE_0) for cohort, averaged estimated coefficients and standard errors (SE_m), empirical standard deviations (S_m) over control samples and empirical standard errors ($SE_m^* = (SE_0^2 + S_m^2)^{1/2}$) with $m = 1$ and 3 controls.

	Cohort		NCC, m=1				NCC, m=3			
	coef	SE_0	coef	SE_1	S_1	SE_1^*	coef	SE_3	S_3	SE_3^*
Weibull baseline hazard										
θ	1.134	0.111	1.135	0.119	0.043	0.119	1.133	0.114	0.024	0.114
γ	-5.684	0.646	-5.810	0.937	0.757	0.996	-5.711	0.747	0.379	0.749
β_1	0.632	0.209	0.680	0.317	0.263	0.336	0.643	0.247	0.134	0.249
β_2	0.760	0.223	0.795	0.342	0.253	0.337	0.765	0.263	0.136	0.261
β_3	0.015	0.228	0.029	0.337	0.256	0.343	0.019	0.264	0.139	0.267
Exponential baseline hazard										
γ	-5.487	0.625	-5.604	0.911	0.728	0.959	-5.513	0.725	0.373	0.728
β_1	0.629	0.209	0.663	0.315	0.258	0.332	0.630	0.246	0.133	0.248
β_2	0.777	0.223	0.812	0.340	0.252	0.336	0.782	0.262	0.136	0.261
β_3	0.015	0.228	0.029	0.334	0.254	0.341	0.019	0.263	0.138	0.266
Unspecified baseline hazard, pseudo-likelihood										
β_1	0.625	0.211	0.674	0.316	0.260	0.335	0.638	0.248	0.132	0.249
β_2	0.767	0.223	0.803	0.341	0.252	0.337	0.772	0.263	0.136	0.262
β_3	0.006	0.228	0.020	0.336	0.254	0.341	0.010	0.264	0.138	0.267
Unspecified baseline hazard, partial likelihood										
β_1	0.625	0.211	0.652	0.337	0.252	0.329	0.626	0.256	0.138	0.252
β_2	0.767	0.223	0.827	0.380	0.297	0.371	0.789	0.280	0.159	0.274
β_3	0.006	0.228	0.128	0.359	0.272	0.355	0.068	0.275	0.158	0.278

Table 5.2 contain some information derived from Table 5.1 concerning the evidence of bias and the efficiency of the case-control estimators relative to the cohort. Columns 1 and 4 present statistics $t_m = \sqrt{500}(\hat{\eta} - \tilde{\eta}_\bullet)/S_m$ where $\hat{\eta}$ is a cohort estimate and $\tilde{\eta}_\bullet$ an average estimate over the control samples. Presumably t_m is standard normal distributed if there is no bias. Columns 2 and 5 gives asymptotical relative efficiencies $ARE_m = (SE_0/SE_m)^2$ whereas column 3 and 6 give what may be called empirical relative efficiencies $ERE_m = (SE_0/SE_m^*)^2$.

Table 5.2. Significance check of bias (t-values) and efficiencies relative to the cohort.

	m=1			m=3		
	t	ARE	ERE	t	ARE	ERE
Weibull baseline hazard						
θ	0.52	0.87	0.87	-0.60	0.96	0.96
γ	-3.71	0.48	0.42	-1.55	0.75	0.74
β_1	4.02	0.44	0.39	1.86	0.72	0.71
β_2	3.10	0.43	0.44	0.81	0.72	0.73
β_3	1.26	0.46	0.44	0.69	0.75	0.73
Exponential baseline hazard						
γ	-3.57	0.47	0.42	-1.55	0.74	0.74
β_1	3.82	0.44	0.40	1.80	0.72	0.71
β_2	3.04	0.43	0.44	0.73	0.72	0.73
β_3	1.24	0.46	0.45	0.63	0.75	0.73
Unspecified baseline hazard, pseudo-likelihood						
β_1	4.21	0.44	0.40	2.07	0.72	0.72
β_2	3.21	0.43	0.44	0.81	0.72	0.73
β_3	1.23	0.46	0.45	0.65	0.75	0.73
Unspecified baseline hazard, partial likelihood						
β_1	2.31	0.39	0.41	0.17	0.68	0.70
β_2	4.51	0.35	0.36	3.01	0.64	0.66
β_3	10.01	0.40	0.41	8.69	0.69	0.68

Table 5.2 indicates that the pseudo-likelihood adds bias to several parameters when there is one control. However comparing to the sampling induced variation this bias is hardly important. Actually the most severe bias is found on the partial likelihood estimates of β_3 . The agreement between the estimated and empirical standard errors estimates is excellent with 3 controls, but with 1 control there are fairly large discrepancies concerning the standard errors of γ and β_1 . The efficiencies of the regression parameters are in accordance with case-control rule of thumb $m/(m+1)$ for $m=3$, but slightly below for $m=1$. The efficiencies on the intercept are comparable to of those the regression parameter and this differ from the results in the simulation study. The efficiencies on the shape parameter are in accordance with those in the simulation study. Also in this case the semi-parametric pseudo-likelihood estimator appears more efficient than the partial likelihood estimator. With $m=1$ the AREs are 0.88, 0.80 and 0.88 for the covariates age, HIV and sex respectively. The corresponding EREs are 1.04, 0.82 and 0.93. With $m=3$ we get AREs of 0.94, 0.88 and 0.92 and EREs of 0.98, 0.91 and 0.92.

6 Discussion

The weighting method used in this paper has been applied within context of likelihoods and partial likelihoods. It should be possible to apply this technique also in most other situations where the estimators or estimating equations can be expressed as sums over the individuals. One interesting example may be non-parametric estimation of the excess risk / additive model of Aalen (1980, 1989). Already Borgan & Langholz (1995) have shown that this model may be fitted non-parametrically if the number of matched controls per case is at least as large the number of covariate functions. With the techniques used in this paper the matched controls are reused and the restriction of Borgan & Langholz should be redundant. Indeed Borgan & Langholz experimented with reusing controls in neighbouring case-control sets with promising results.

The semi-parametric pseudo-likelihood estimator have in both the simulation experiments and the data example been efficient to the usually applied partial likelihood estimator. It may worth investigating how many control samples need to be joined before the essential efficiency gain is achieved. Deriving distributional results of such estimators could be very valuable.

In the literature on case-control within cohort studies there has been some discussion on whether designs that reuse controls actually improve accuracy of estimators. Langholz & Thomas (1991) developed several ingenious designs with this purpose and found that little or nothing was gained compared to nested case-control. Their conclusion was that it is false that the nested case-control design does not make good use of the data. This may be true, however, it also holds that reusing controls may give important improvements.

In several papers (see e.g. Scott & Wild, 1986, 1991, Wild, 1991) it has been pointed out that weighting techniques similar to those applied in this paper may be inefficient, even seriously so, within unmatched case-control studies. The pseudo-likelihood estimators may be inefficient in nested case-control studies too. It is also an open question whether it always is better than usual partial likelihood estimator.

The original purpose of the partial likelihood nested case-control estimator was to fit proportional hazard models on large data sets. If in addition the data are very time-dependent it is likely still the best option available. In such case the pseudo-likelihood estimator is less attractive, because reusing controls will slow down the computations. Furthermore improvement of the estimators is easily achieved by increasing the number of matched controls.

A key requirement for constructing the pseudo-likelihood estimators was that covari-

ate histories may be obtained for both cases and controls. It is possible to relax this to require that for the controls covariate histories are obtained, but that for the cases only covariate values at the failure time is available. In such case one may fit the models by maximizing parametric pseudo-likelihoods

$$l_{(p0)} = \sum_{i=1}^n D_i \log(\lambda_i(X_i)) - \sum_{i=1}^n \frac{V_{0i}}{p_{0i}} \int_{l_i}^{X_i} \lambda_i(s) ds$$

Similarly semi-parametric pseudo-likelihood estimates are obtained by substituting $\tilde{S}^{(0)}(\beta, X_i)$ in (2.8) by $\tilde{S}_0^{(0)}(\beta, X_i) = \sum V_{0j} \exp(\beta' Z_j(X_i)) / p_{0j}$ where the sum is taken over $j \in \mathcal{R}_i$. Distributional properties of such estimators should be possible to address with the techniques of this paper.

This discussion has focused much on efficiency improvement. The main point of the paper, however, is that the pseudo-likelihood weighting approach relieves the nested case-control of restrictive model assumptions. Although some methods have been devised that go beyond fitting proportional hazard models, virtually any method can be modified by weighting techniques. This also shows that most any model that is identifiable with cohort data, continues to be so with nested case-control data.

Acknowledgements

I wish to thank Ørnulf Borgan for helpful discussions.

Appendix A: Asymptotic behaviour of ρ_{ij}

If not otherwise specified the integrals, sums and product in the following are taken over intervals $\langle m_{1ij}, m_{2ij} \rangle$ assuming $m_{1ij} < m_{2ij}$. To show that $\rho_{ij} = -(m/n) \int \lambda(t)/y(t) dt + O_p(n^{-2}) = O_p(n^{-1})$ first note that the k -th term in the product in (3.1) can, after some algebra, be written

$$1 - \frac{m(Y_k - 1 - m)}{(Y_k - 1)^2(Y_k - 2)} D_k - Q_k^0 D_k = 1 - \frac{mQ_k^1}{Y_k^2} D_k$$

where with $u = m/(Y_k - 1)$ and $v = (m - 1)/(Y_k - 2)$ we have

$$Q_k^0 = uv \left[\frac{1}{(1-u)^2} - 1 \right] - 2u \left[\frac{1}{(1-u)^2} - 1 - 2u \right]$$

It follows that both $\sup_k |Q_k^0 Y_k^2| \rightarrow 0$ and $\sup_k |Q_k^1 - 1| \rightarrow 0$ in probability as $n \rightarrow \infty$. We now get

$$\rho_{ij} + 1 = \prod_k \left[1 - m \frac{Q_k^1}{Y_k^2} D_k \right] \leq \exp \left(-m \sum_k \frac{Q_k^1}{Y_k^2} D_k \right) = \exp \left(-\frac{m}{n} \int \frac{\lambda(t)}{y(t)} dt + O_p(n^{-2}) \right).$$

Thus

$$\rho_{ij} \leq -\frac{m}{n} \int \frac{\lambda(t)}{y(t)} dt + O_p(n^{-2}) = O_p(n^{-1})$$

The stated result now follows if the opposite inequality is established. The main interest in the result though is that $\rho_{ij} = O_p(n^{-1})$ and we will prove this first. By help of the trivial inequality

$$1 - \frac{a}{x} \geq (1 - a)^{1/x} \quad \text{when } 0 \leq a \leq 1 \text{ and } x \geq 1$$

it now turns out, with $Y_0 = \inf_k Y_k$, that

$$\rho_{ij} + 1 \geq \prod_k [1 - m \frac{Q_k^1}{Y_k Y_0} D_k] \geq \prod_k [1 - m \frac{Q_k^1}{Y_k} D_k]^{1/Y_0} = \exp(-\frac{m}{Y_0} \int \lambda(t) dt + O_p(n^{-2})).$$

The second inequality here requires that $\sup_k m Q_k^1 / Y_k \leq 1$ which will hold with a probability tending to 1. It now follows that also

$$\rho_{ij} \geq -\frac{m}{Y_0} \int \lambda(t) dt + O_p(n^{-2}),$$

thus $\rho_{ij} = O_p(n^{-1})$.

The result that $\rho_{ij} \geq -(m/n) \int \lambda(t)/y(t) dt + O_p(n^{-2})$ follow by the same approach. Let $t_0 = 0 < t_1 < \dots$ be a partitioning of the positive number and for $t_{j-1} < t \leq t_j$ define $Y_0(t)$ as the step function taking values $\inf Y_k$ where infimum is taken over k such that $X_k \in [t_{j-1}, t_j]$. We get

$$\rho_{ij} \geq -m \sum_j \int_{t_{j-1}}^{t_j} \frac{\lambda(t)}{Y_0(t)} dt + O_p(n^{-2}).$$

Noting that the partitioning can be made arbitrarily fine the proof is completed.

Appendix B: Some computer code for correcting covariance matrices

The hardest part of obtaining the matrices $\tilde{\Delta}_n$ and $\tilde{\Delta}_n^C$ lies in computing

$$\tilde{\Delta}_{1n} = \sum_{i \neq j} \frac{V_i V_j (1 - p_i)(1 - p_j) \rho_{ij}}{p_i p_j} u_i(\tilde{\gamma}) u_j(\tilde{\gamma})^t$$

and the counterpart expression in (3.12). The direct approach would be a double for-loop, but this is too slow in Splus. Here parts of the computer code that was used is given. The code assumes that there is right censoring, but no left-truncation, on the cohort. Then we have that $\rho_{ij} = \rho_{ik}$ if both $X_i \leq X_j$ and $X_i \leq X_k$. It is also assumed that the individuals are sorted on the right censoring times.

In the simulations and the data example such calculations have been carried out using the Splus-procedure `delta1proc` below. In this procedure `p` corresponds to p_i and `rho` to ρ_{ij} for the sampled controls with $p_i < 1$. The number of such controls are `n` and `q` are the number of parameters that are estimated. The argument `hmat` is the matrix of score-contributions for the `n` controls with the parametric specifications and the corresponding matrix of the \widetilde{W}_{0i} in the semi-parametric case. The Splus-variable to be assigned the call to `delta1proc` need to be specified as a `qxq`-matrix before the call.

```
delta1proc<- function(hmat, p, rho, n, q)
{
  if(!is.loaded("delta1"))
    dyn.load("delta1.o")
  delta1<-matrix(rep(0,q^2),nrow=q)
  storage.mode(delta1) <- "double"
  .C("delta1",
    as.double(hmat),
    as.double(p),
    as.double(rho),
    as.integer(n),
    as.integer(q),
    delta1 = delta1)$delta1
}
```

This Splus-function call the compiled C-routine `delta1.o`. The source code `delta1.c` is given below. Note that the estimates use the approximation $p_{oij} \approx p_{oi}p_{oj}$.

```
void delta1(double *h, double *p, double *rho, long *in, long *ip,
            double *odelta1)
{
  long i,j,k,l,minij;
  double v;
  for (k=0;k<(*ip);k=k+1){
  for (l=0;l<(*ip);l=l+1){
  v=0;
  for(i=0;i<(*in);i=i+1)
  {for(j=0;j<(*in);j=j+1)
  if((i-j)!=0){
  if(i<j) minij=i; else minij=j;
  v=v+
  h[k+(*ip)*i]*h[l+(*ip)*j]*(1-p[i])*(1-p[j])*rho[minij]/(p[i]*p[i]*p[j]*p[j]);
  }}
  odelta1[l+k*(*ip)]=v;}}
}
```

Alternatively $\widetilde{\Delta}_{1n}$ can be computed by a matrix-approach, thus avoiding the C-routine. This approach was somewhat slower, but should work well when only the covariance estimates

from one case-control data set is considered. The matrices are of dimension $n \times n$ where n is still the number of controls with $p_i < 1$. If this number is high Splus may run into memory problems. In the following Splus-code `renkont` is a vector indicating an individual with $p_i < 1$, thus `putv0` becomes the p_i for these controls and `rho0` the vector of ρ_{ij} where $X_i \leq X_j$.

```

putv0<-1/w[renkont]
rho0<-rho[renkont]
P0<-matrix(putv0,ncol=1)\%*\%matrix(putv0,nrow=1)
Rho0<-matrix(rep(rho0,antkont),ncol=antkont)
Rho0<-Rho0*(Rho0>t(Rho0))+t(Rho0)*(Rho0<=t(Rho0))
P0<-(matrix(1-putv0,ncol=1)\%*\%matrix(1-putv0,nrow=1))/P0
P0<-P0/(matrix(putv0,ncol=1)\%*\%matrix(putv0,nrow=1))
Q0<-P0*Rho0

```

Then `Rho0` contains the matrix of ρ_{ij} for the controls, whereas `P0` is the matrix with $(1 - p_{0i})(1 - p_{0j}) / (p_{0i}p_{0j})^2$ as the ij -th element. The above code result in use of the approximation $p_{0ij} \approx p_{0i}p_{0j}$. Inserting the following statement between the third and fourth line above the approximation is avoided.

```

P0<-P0+Rho0*(matrix(1-putv0,ncol=1)\%*\%matrix(1-putv0,nrow=1))

```

Whenever checked the effect of this correction was negligible. In either case one now obtains the ij -th element of $\tilde{\Delta}_{1n}$ by the commands

```

H<-matrix(scorekont[i,],ncol=1)\%*\%t(scorekont[j,])
H<-H-diag(diag(H))
delta1[i,j]<-delta1[j,i]<-sum(H*Q0)

```

where `scorekont` is the matrix of score-contributions for the n controls.

References

- [1] Aalen, OO (1980) A model for non-parametric regression analysis of counting processes. *Springer Lect. Notes Statist.* **2**, 1-25.
- [2] Aalen, OO (1989) A linear regression model for the analysis of life times. *Statist. Med.* **8**, 907-925.
- [3] Andersen, PK, Borgan, Ø, Gill, RD, and Keiding, N (1993). *Statistical Models Based on Counting Processes*. Springer Verlag, New York.
- [4] Andersen, PK and Gill, RD (1982) Coxs regression model for counting processes: a large sample study. *Ann. Statist.* **10**, 1100-1120.

- [5] Borgan, Ø (1984) Maximum likelihood estimation in parametric counting processes models, with application to to censored failure time data. *Scand. J. Statist.*, **11**, 1-16.
- [6] Borgan, Ø, Goldstein, L, and Langholz, B (1995). Methods for the analysis of sampled cohort data in the Cox proportional hazards model. *Ann. Statist.* (in press).
- [7] Borgan, Ø and Langholz, B (1993). Non-parametric estimation of relative mortality from nested case-control studies. *Biometrics* **49**, 593–602.
- [8] Borgan, Ø and Langholz, B (1995). Estimation of excess risk from case-control data using Aalens linear regression model. Statistical Research Report, no. 5, University of Oslo.
- [9] Breslow, NE and Day, NE (1980). *Statistical Methods in Cancer Research. Volume I – The Design and Analysis of Case-Control Studies, IARC Scientific Publications*, Vol. 32. International Agency for Research on Cancer, Lyon.
- [10] Breslow, NE and Day, NE (1987). *Statistical Methods in Cancer Research. Volume II – The Design and Analysis of Cohort Studies, IARC Scientific Publications*, Vol. 82. International Agency for Research on Cancer, Lyon.
- [11] Cochran, WG (1977) *Sampling Techniques*, 3rd ed. Wiley, New York.
- [12] Cox, DR and Oakes, D (1984) *Analysis of Survival Data*. Chapman and Hall.
- [13] Eskild, A, Magnus, P, Samuelsen, SO, Sohlberg, C and Kittelsen, P (1993) Differences in mortality rates between HIV positive and HIV negative intravenous drug users. *Int. J. Epid.* **22**, 315-320.
- [14] Flanders, WD and Greenland, S (1991) Analytic Methods for two-stage case-control studies and other stratified designs. *Statist. Med.* **10**, 739-747.
- [15] Goldstein, L and Langholz, B (1992) Asymptotic Theory of Nested Case-Control Sampling in the Cox-Regression Model. *Ann. Statist.* **20**, 1903-1928.
- [16] Hajek, J. (1961) Some extensions of the Wald-Wolfowitz-Noether theorem. *Ann. Math. Statist.* **32**, 506-523.
- [17] Kalbfleisch, JD and Lawless, JF (1988) Likelihood analysis of multi-state models for disease incidence and mortality. *Statist. Med.* **7**, 147-160.
- [18] Langholz, B and Thomas, DC (1991) Efficiency of cohort sampling designs: some surprising results. *Biometrics* **47**, 1563-1571.

- [19] Langholz, B and Borgan, Ø (1995). Counter-matching: A stratified nested case-control sampling method. *Biometrika* **82**, 69–79.
- [20] Lehman, E (1975) *Nonparametrics*. Holden-Day, Inc. San Francisco.
- [21] Oakes, D (1981). Survival times: Aspects of partial likelihood (with discussion). *Internat. Statist. Rev.* **49**, 235–264.
- [22] Prentice, RL (1986) A case-cohort design for epidemiological cohort studies and disease prevention trials. *Biometrika* **73**, 1-11.
- [23] Samuelsen, SO (1989) *Two incomplete data problems in life-history analysis: Double censoring and the Case-cohort design*. PhD thesis, University of Oslo.
- [24] Self, SG and Prentice, RL (1987) Asymptotic distribution theory and efficiency results for case-cohort studies. *Ann. Statist.* **16**, 64-81.
- [25] Scott, AJ and Wild, CJ (1986) Fitting logistic regression models under case-control or choice-based sampling. *J. Roy. Statist. Soc. B* **48**, 170-182.
- [26] Scott, AJ and Wild, CJ (1991) Fitting logistic regression models in stratified case-control studies. *Biometrics* **47**, 497-510.
- [27] Thomas, DC (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. A* **140**, 469–491.
- [28] Wild, CJ (1991) Fitting prospective regression models to case-control data. *Biometrika* **78**, 705-717.