Using martingale residuals to assess goodness-of-fit for sampled risk set data

Ørnulf Borgan and **Bryan Langholz**

University of Oslo and University of Southern California

Department of Mathematics
University of Oslo
P.O. Box 1053 Blindern
N-0316 Oslo, Norway
e-mail: `borgan@math.uio.no`

Department of Preventive Medicine
University of Southern California
1540 Alcazar Street CHP-220
Los Angeles, Ca 90033, U.S.A.
e-mail: `langholz@usc.edu`

Abstract: Standard use of Cox's regression model and other relative risk regression models for censored survival data requires collection of covariate information on all individuals under study even when only a small fraction of them die or get diseased. For such situations risk set sampling designs offer useful alternatives. For cohort data, methods based on martingale residuals are useful for assessing the fit of a model. Here we introduce grouped martingale residual processes for sampled risk set data, and show that plots of these processes provide a useful tool for checking model-fit. Further we study the large sample properties of the grouped martingale residual processes, and use these to derive a formal goodness-of-fit test to go along with the plots. The methods are illustrated using data on lung cancer deaths in a cohort of uranium miners.

1

1. Introduction

Cox regression is central to modern survival analysis, and it is the method of choice when one wants to assess the influence of risk factors and other covariates on mortality or morbidity. A number of methods, both graphical methods and formal tests, have been proposed to assess the goodness-of-fit of Cox's model; see e.g. the recent textbooks by Hosmer and Lemeshow [10], Klein and Moeschberger [12], and Therneau and Grambsch [22].

One important tool for checking the fit of Cox's regression model is the martingale residuals introduced by Barlow and Prentice [4]. Therneau, Grambsch and Fleming [23] proposed to use a smoothed plot of the these residuals versus a covariate as a means to detect its correct functional form, while Grambsch, Therneau and Fleming [7] suggested a similar, improved plot; see [22, Section 5.7] for a review and further discussion. Another approach was taken by Aalen [2]. In the context of his additive model [1], he proposed to plot martingale residual processes, aggregated over groups of individuals, versus time as an omnibus procedure to check the fit of a model. Aalen's idea was implemented for Cox's regression by Grønnesby and Borgan [8], who also derived a formal goodness-of-fit test to go along with the graphical procedure.

The commonly used methods for inference in Cox's regression model, including the methods for goodness-of-fit, require collection of covariate information on all individuals under study. This may be very expensive in large epidemiologic cohort studies of a rare disease. Risk set sampling designs, where covariate information is collected for all failing individuals (cases), but only for a sample of the non-failing ones (controls) then offer useful alternatives which may drastically reduce the resources that need to be allocated to a study for data collection and checking.

In the present paper we will use the counting process framework of Borgan, Goldstein and Langholz [5] to generalize the martingale residual processes to sampled risk set data. In this context it does not seem feasible to obtain graphical procedures analogous to the smoothed martingale residual plot [23] or the related plot of Grambsch *et al.* [7]. However, we may still generalize the grouped martingale residual processes plots of Grønnesby and Borgan [8] and the accompanying goodness-of-fit test. In doing this we will not restrict ourselves to Cox's regression model, but consider a general class of relative risk regression models.

The outline of the paper is as follows. In section 2 we introduce the class of relative risk regression models, describe the type of failure time data considered for the cohort, and review how the cohort data may be formu-

lated by means of counting processes. Then we outline how the martingale residuals and grouped martingale residual processes follow naturally from the counting process formulation. Section 3 is devoted to risk set sampling. We first introduce a general framework for risk set sampling [5], describe how it specializes for simple random and counter-matched sampling, and review methods for inference for sampled risk set data. Then we outline how sampled risk set data can be described by processes counting jointly the occurrence of failures and the sampling of controls, and we use this counting process formulation to generalize the grouped martingale residual processes and accompanying goodness-of-fit test of Grønnesby and Borgan [8] to sampled risk set data. An illustration for a study of lung cancer death in a cohort of uranium miners is provided in section 4, while proofs are collected in section 5. In section 6 we briefly explain how the results extend to matched risk set sampling designs, while some concluding comments are given in the final section 7. Throughout the paper we will without further references use standard results for counting processes [3, 6].

2. Cohort data

We consider a cohort of n individuals, and denote by $\alpha(t; \mathbf{z}_i)$ the hazard rate at time t for an individual i with vector of covariates $\mathbf{z}_i(t) = (z_{i1}(t), \ldots, z_{ip}(t))^\mathsf{T}$. Here the time-variable t may be age, time since employment, or some other time-scale relevant to the problem at hand, where we throughout assume that $t \in (0, \tau]$ for a given terminal study time τ. A covariate may be time-fixed or time-dependent; in the latter case its value at time t is assumed to be known "just before" time t, i.e., the covariate is assumed to be predictable. We assume that the covariates of individual i are related to its hazard rate by the relative risk regression model

$$\alpha(t; \mathbf{z}_i) = c(\boldsymbol{\beta}_0, \mathbf{z}_i(t)) \, \alpha_0(t). \tag{2.1}$$

Here $c(\boldsymbol{\beta}_0, \mathbf{z}_i(t))$ is a relative risk function, $\boldsymbol{\beta}_0 = (\beta_{01}, \ldots, \beta_{0p})^\mathsf{T}$ is a vector of regression coefficients describing the effect of the covariates, while the baseline hazard rate $\alpha_0(t)$ is left unspecified. Throughout we use $\boldsymbol{\beta}_0$ to denote the vector of true regression coefficients, while we use $\boldsymbol{\beta}$ as an argument in the partial likelihood and similar quantities. We normalize the relative risk function by assuming $c(\boldsymbol{\beta}_0, \mathbf{0}) = 1$. Thus $\alpha_0(t)$ corresponds to the hazard rate of an individual with all covariates identically equal to zero. For the exponential relative risk function $c(\boldsymbol{\beta}_0, \mathbf{z}_i(t)) = \exp(\boldsymbol{\beta}_0^\mathsf{T} \mathbf{z}_i(t))$, formula (2.1) gives the usual Cox regression model. Other possibilities include the linear

relative risk function $c(\boldsymbol{\beta}_0, \mathbf{z}_i(t)) = 1 + \boldsymbol{\beta}_0^\mathsf{T}\mathbf{z}_i(t)$ and the excess relative risk model $c(\boldsymbol{\beta}_0, \mathbf{z}_i(t)) = \prod_{j=1}^{p}(1 + \beta_{0j}\, z_{ij}(t))$.

The individuals in the cohort may be followed over different periods of time, i.e., our observations may be subject to left-truncation and right censoring. It is a fundamental assumption throughout that the left truncation and right censoring are independent in the sense that the additional knowledge of which individuals have entered the study or have been censored before any time t do not carry information on the risks of failure at t; see [3, Sections III.2-3] and [11, Sections 1.3 and 6.2] for a general discussion on the concept of independent censoring.

We let $t_1 < t_2 < \cdots$ be the times when failures are observed and, assuming that there are no tied failures, denote by i_j the individual who fails at t_j. The risk set $\mathcal{R}_j$ is the collection of all individuals who are under observation "just before" time t_j. In particular the case i_j is a member of $\mathcal{R}_j$. Then the vector of regression parameters in (2.1) is estimated by $\widehat{\boldsymbol{\beta}}$, the value of $\boldsymbol{\beta}$ maximizing Cox's partial likelihood, while the cumulative baseline hazard rate $A_0(t) = \int_0^t \alpha_0(u)\mathrm{d}u$ is estimated by the Breslow estimator

$$\widehat{A}_0(t) = \sum_{t_j \leq t} \frac{1}{\sum_{l \in \mathcal{R}_j} c(\widehat{\boldsymbol{\beta}}, \mathbf{z}_l(t_j))},$$

e.g. [3, Section VII.2].

In order to define the martingale residuals, we first need to review some basic facts on counting processes, (cumulative) intensity processes and martingales. To this end, introduce the processes

$$N_i(t) = \sum_{t_j \leq t} I\{i_j = i\}; \qquad i = 1, 2, \ldots, n; \tag{2.2}$$

counting the number of observed events for individual i in $(0, t]$ (which is 0 or 1 for survival data). The intensity processes λ_i of the counting process N_i is given heuristically by $\lambda_i(t)\mathrm{d}t = P(\mathrm{d}N_i(t) = 1 \,|\, \mathcal{H}_{t-})$, where $\mathrm{d}N_i(t)$ is the increment of N_i over the small time interval $[t, t + \mathrm{d}t)$, and $\mathcal{H}_{t-}$ denotes all information available to the researcher "just before" time t. Then by (2.1) and the independent censoring assumption,

$$\lambda_i(t) = Y_i(t)\, \alpha(t; \mathbf{z}_i) = Y_i(t)\, c(\boldsymbol{\beta}_0, \mathbf{z}_i(t))\, \alpha_0(t), \tag{2.3}$$

with $Y_i(t)$ a left-continuous at risk indicator for individual i. Thus $\mathcal{R}(t) = \{i \,|\, Y_i(t) = 1\}$ is the risk set at time t, and $n(t) = |\mathcal{R}(t)|$ is the number at risk "just before" time t. Note that $\mathcal{R}_j = \mathcal{R}(t_j)$.

Corresponding to λ_i, we define the cumulative intensity process

$$\Lambda_i(t) = \int_0^t \lambda_i(u)\,\mathrm{d}u = \int_0^t Y_i(u)\,c(\boldsymbol{\beta}_0, \mathbf{z}_i(u))\,\alpha_0(u)\,\mathrm{d}u. \qquad (2.4)$$

By standard results on counting processes, it then follows that $M_i(t) = N_i(t) - \Lambda_i(t)$; $i = 1, 2, \ldots, n$; are local square integrable martingales. If we insert the maximum partial likelihood estimator $\widehat{\boldsymbol{\beta}}$ for $\boldsymbol{\beta}_0$ and the increment $\mathrm{d}\widehat{A}_0(u)$ of the Breslow estimator for $\alpha_0(u)\mathrm{d}u$ in (2.4), we get the estimated cumulative intensity processes

$$\widehat{\Lambda}_i(t) = \int_0^t Y_i(u)\,c(\widehat{\boldsymbol{\beta}}, \mathbf{z}_i(u))\,\mathrm{d}\widehat{A}_0(u) = \sum_{t_j \leq t} \frac{Y_i(t_j)\,c(\widehat{\boldsymbol{\beta}}, \mathbf{z}_i(t_j))}{\sum_{l \in \mathcal{R}_j} c(\widehat{\boldsymbol{\beta}}, \mathbf{z}_l(t_j))},$$

and the *martingale residual processes* $\widehat{M}_i(t) = N_i(t) - \widehat{\Lambda}_i(t)$. Evaluating the martingale residual processes at $t = \infty$ we arrive at the *martingale residuals* $\widehat{M}_i = \widehat{M}_i(\infty)$ first considered by Barlow and Prentice [4].

Following Aalen [2], Grønnesby and Borgan [8] considered the *grouped* martingale residual processes, obtained by aggregating the individual martingale residual processes $\widehat{M}_i(t)$ over groups of individuals. Specifically, assume that we have some grouping of the individuals, typically based on the values of one or two covariates, and denote the groups by $J = 1, \ldots, G$. We will allow the grouping of the individuals to depend on time. Thus an individual may move from one group to another as time passes, as will often be the case when the grouping is performed on the basis of one or more time-dependent covariates. It is a prerequisite, however, that the information used for grouping at time t is available "just before" time t, i.e., the grouping must be based on the "history" $\mathcal{H}_{t-}$. Then, if we denote by $\mathcal{J}(u)$ the set of all individuals who belong to group J at time u, the group J martingale residual process takes the form

$$\widehat{M}_J(t) = \int_0^t \sum_{i \in \mathcal{J}(u)} \mathrm{d}\widehat{M}_i(u) = N_J(t) - \sum_{t_j \leq t} \frac{\sum_{i \in \mathcal{R}_j \cap \mathcal{J}(t_j)} c(\widehat{\boldsymbol{\beta}}, \mathbf{z}_i(t_j))}{\sum_{l \in \mathcal{R}_j} c(\widehat{\boldsymbol{\beta}}, \mathbf{z}_l(t_j))}. \qquad (2.5)$$

Here $N_J(t) = \int_0^t \sum_{i \in \mathcal{J}(u)} \mathrm{d}N_i(u)$ is the observed number of failures in group J in $(0, t]$, while the last term on the right-hand side of (2.5) is an estimate of the expected number of failures in that group if the relative risk regression model (2.1) holds true. In section 4 we illustrate how a plot of the grouped martingale residual processes provides a useful tool for checking the fit of the model.

For the special case of an exponential relative risk function, Grønnesby and Borgan [8] studied the large sample properties of the grouped martingale residual processes. The corresponding results for a general relative risk function, may be obtained as a special case of the results for sampled risk set data given in section 3.5 below. There we also derive a formal goodness-of-fit test based on the grouped martingale residual processes.

3. Risk set sampling designs

In sections 3.4 and 3.5 below we will see how martingale residuals may be defined for risk set sampling designs. Before we do that, however, we will review the framework for risk set sampling of Borgan *et al.* [5] and generalize some of their results to the situation with a general relative risk function.

3.1. A model for risk set sampling

For risk set sampling one selects, whenever a failure occurs, a (typically small) number of controls for the failing individual. The set consisting of these controls together with the failing individual (the case) is called a *sampled risk set*. In order to describe in general terms how the sampling of controls is performed, we need to introduce the "cohort and sampling history" $\mathcal{F}_{t-}$, which contains information about events in the cohort (i.e. $\mathcal{H}_{t-}$) as well as on the sampling of controls, up to, but not including, time t. Based on the parts of this history that are available to the researcher, one decides on a sampling strategy for the controls. Such a strategy may be described in probabilistic terms as follows. Let $\mathcal{P}$ be the power set of $\{1, 2, \ldots, n\}$, i.e. the set of all subsets of $\{1, 2, \ldots, n\}$, and let $\mathcal{P}_i = \{\mathbf{r} : \mathbf{r} \in \mathcal{P}, i \in \mathbf{r}\}$. Then, given $\mathcal{F}_{t-}$, if an individual i fails at time t, we select the set $\mathbf{r} \in \mathcal{P}_i$ as our sampled risk set with (known) probability $\pi_t(\mathbf{r} \,|\, i)$. Thus, if $Y_i(t) = 1$, then $\pi_t(\mathbf{r} \,|\, i)$ is a probability distribution over sets $\mathbf{r} \in \mathcal{P}_i$. For notational convenience we let $\pi_t(\mathbf{r} \,|\, i) = 0$ whenever $Y_i(t) = 0$.

It turns out to be useful to have a factorization of the sampling probabilities $\pi_t(\mathbf{r} \,|\, i)$. To this end we introduce

$$\pi_t(\mathbf{r}) = n(t)^{-1} \sum_{l \in \mathbf{r}} \pi_t(\mathbf{r} \,|\, l), \qquad (3.1)$$

and note that

$$\sum_{\mathbf{r} \in \mathcal{P}} \pi_t(\mathbf{r}) = n(t)^{-1} \sum_{l=1}^{n} \sum_{\mathbf{r} \in \mathcal{P}_l} \pi_t(\mathbf{r} \,|\, l) = n(t)^{-1} \sum_{l=1}^{n} Y_l(t) = 1.$$

Thus $\pi_t(\mathbf{r})$ is a probability distribution over sets $\mathbf{r} \in \mathcal{P}$. We also introduce

$$w_i(t, \mathbf{r}) = \frac{\pi_t(\mathbf{r} \mid i)}{\pi_t(\mathbf{r})}, \tag{3.2}$$

and get the factorization

$$\pi_t(\mathbf{r} \mid i) = w_i(t, \mathbf{r}) \, \pi_t(\mathbf{r}). \tag{3.3}$$

Note that the above framework allows the sampling probabilities to depend in an arbitrary way on events in the past, i.e., on events that are contained in $\mathcal{F}_{t-}$. The sampling probabilities may, however, not depend on events in the future. For example, one may not exclude as a potential control for a current case an individual that subsequently fails. Also note that the selection of controls is done independently at the different failure times, so that subjects may serve as controls for multiple cases, and cases may serve as controls for other cases that failed when the case was at risk. A basic assumption throughout is that not only the truncation and censoring, but also the sampling of controls, are independent in the sense that the additional knowledge of which individuals have entered the study, have been censored or have been selected as controls before any time t do not carry information on the risks of failure at t.

3.2. Two common sampling designs

The most common risk set sampling design is simple random sampling; the classical *nested case-control design* [24]. For this design, if individual i fails at time t, one selects $m - 1$ controls by simple random sampling from the $n(t) - 1$ non-failing individuals at risk. In probabilistic terms the design is given by

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

for any set $\mathbf{r} \in \mathcal{P}_i$. Here the factorization (3.3) applies with

$$\pi_t(\mathbf{r}) = \binom{n(t)}{m}^{-1} I\{|\mathbf{r}| = m, \mathbf{r} \subset \mathcal{R}(t)\}; \quad \mathbf{r} \in \mathcal{P};$$

$$w_i(t, \mathbf{r}) = \frac{n(t)}{m} I\{i \in \mathbf{r}\}. \tag{3.4}$$

To select a simple random sample, the only piece of information needed from $\mathcal{F}_{t-}$ is the at risk status of the individuals. Often, however, some additional information is available for all cohort members, e.g., a surrogate

measure of the exposure of main interest may be available for everyone. Langholz and Borgan [13] have developed an "exposure" stratified design which makes it possible to incorporate such information into the sampling process in order to obtain a more informative sample of controls. For this design, called *counter-matching*, one applies the additional piece of information from $\mathcal{F}_{t-}$ to classify each individual at risk into one of say, S, strata. We denote by $\mathcal{R}_s(t)$ the subset of the risk set $\mathcal{R}(t)$ which belongs to stratum s, and let $n_s(t) = |\mathcal{R}_s(t)|$ be the number at risk in this stratum just before time t. If individual i fails at t, we want to sample our controls such that the sampled risk set contains a prespecified number m_s of individuals from each stratum s; $s = 1, \ldots, S$. This is obtained as follows. Assume that the failing individual i belongs to stratum $s(i)$. Then for $s \neq s(i)$ one samples randomly without replacement m_s controls from $\mathcal{R}_s(t)$. From the case's stratum $s(i)$ only $m_{s(i)} - 1$ controls are sampled. The failing individual i is, however, included in the sampled risk set so this contains a total of m_s from each stratum. Even though it is not made explicit in the notation, we note that the classification into strata may be time-dependent. A crucial assumption, however, is that the information on which the stratification is based has to be known "just before" time t. In probabilistic terms, counter-matched sampling may be described as follows. For any set $\mathbf{r} \in \mathcal{P}_i$ which is a subset of $\mathcal{R}(t)$ and satisfies $|\mathbf{r} \cap \mathcal{R}_s(t)| = m_s$ for $s = 1, \ldots, S$, we have

$$\pi_t(\mathbf{r} \,|\, i) = \left\{ \binom{n_{s(i)}(t) - 1}{m_{s(i)} - 1} \prod_{s \neq s(i)} \binom{n_s(t)}{m_s} \right\}^{-1}.$$

For counter-matched sampling the factorization (3.3) applies with

$$\pi_t(\mathbf{r}) = \left\{ \prod_{s=1}^{S} \binom{n_s(t)}{m_s} \right\}^{-1} I(|\mathbf{r} \cap \mathcal{R}_s(t)| = m_s; \, s = 1, \ldots, S); \quad \mathbf{r} \in \mathcal{P};$$

$$w_i(t, \mathbf{r}) = \frac{n_{s(i)}(t)}{m_{s(i)}} I\{i \in \mathbf{r}\}.$$

Other sampling designs for the controls are discussed in [5] and [14]. Note that also the full cohort study is a special case of our general framework in which the full risk set is sampled with probability one, i.e., $\pi_t(\mathbf{r} \,|\, i) = I\{\mathbf{r} = \mathcal{R}(t)\}$ for all $i \in \mathcal{R}(t)$, and $\pi_t(\mathbf{r} \,|\, i) = 0$ otherwise.

3.3. Inference for sampled risk set data

As in section 2 we denote by $t_1 < t_2 < \cdots$ the times when failures are observed, and let i_j be the individual who fails at t_j. As described above, the sampled risk set $\widetilde{\mathcal{R}}_j$ is selected according to a sampling distribution $\pi_{t_j}(\mathbf{r} \,|\, i_j)$ specified by the researcher, and it consists of the case i_j and its controls. Covariate information is collected on the cases and their controls, but are not needed for the other individuals in the cohort. It was shown by Borgan *et al.* [5] that from sampled risk set data one may estimate the vector of regression parameters in (2.1) by $\widehat{\boldsymbol{\beta}}$, the value of $\boldsymbol{\beta}$ maximizing the partial likelihood

$$L(\boldsymbol{\beta}) = \prod_{t_j} \frac{c(\boldsymbol{\beta}, \mathbf{z}_{i_j}(t_j)) w_{i_j}(t_j, \widetilde{\mathcal{R}}_j)}{\sum_{l \in \widetilde{\mathcal{R}}_j} c(\boldsymbol{\beta}, \mathbf{z}_l(t_j)) w_l(t_j, \widetilde{\mathcal{R}}_j)}. \tag{3.5}$$

We note that (3.5) is similar to the full cohort partial likelihood. In fact, the full cohort partial likelihood is the special case of (3.5) in which the entire risk set is sampled with probability one and all weights are unity. Note that for simple random sampling, the weights (3.4) are the same for all individuals and hence cancel from (3.5) giving partial likelihood of Oakes [19].

The maximum partial likelihood estimator $\widehat{\boldsymbol{\beta}}$ enjoys similar large sample properties as ordinary maximum likelihood estimators. Specifically $\widehat{\boldsymbol{\beta}}$ is approximately multinormally distributed around the true parameter vector $\boldsymbol{\beta}_0$ with a covariance matrix that may be estimated as $\mathcal{I}(\widehat{\boldsymbol{\beta}})^{-1}$, the inverse of the expected information matrix

$$\mathcal{I}(\widehat{\boldsymbol{\beta}}) = \sum_{t_j} \left\{ \frac{\mathbf{S}_{\widetilde{\mathcal{R}}_j}^{(2)}(\widehat{\boldsymbol{\beta}}, t_j)}{S_{\widetilde{\mathcal{R}}_j}^{(0)}(\widehat{\boldsymbol{\beta}}, t_j)} - \left(\frac{\mathbf{S}_{\widetilde{\mathcal{R}}_j}^{(1)}(\widehat{\boldsymbol{\beta}}, t_j)}{S_{\widetilde{\mathcal{R}}_j}^{(0)}(\widehat{\boldsymbol{\beta}}, t_j)} \right)^{\otimes 2} \right\}. \tag{3.6}$$

Here

$$S_{\widetilde{\mathcal{R}}_j}^{(0)}(\widehat{\boldsymbol{\beta}}, t_j) = \sum_{l \in \widetilde{\mathcal{R}}_j} c(\widehat{\boldsymbol{\beta}}, \mathbf{z}_l(t_j))\, w_l(t_j, \widetilde{\mathcal{R}}_j), \tag{3.7}$$

$$\mathbf{S}_{\widetilde{\mathcal{R}}_j}^{(1)}(\widehat{\boldsymbol{\beta}}, t_j) = \sum_{l \in \widetilde{\mathcal{R}}_j} \dot{\mathbf{c}}(\widehat{\boldsymbol{\beta}}, \mathbf{z}_l(t_j))\, w_l(t_j, \widetilde{\mathcal{R}}_j), \tag{3.8}$$

$$\mathbf{S}_{\widetilde{\mathcal{R}}_j}^{(2)}(\widehat{\boldsymbol{\beta}}, t_j) = \sum_{l \in \widetilde{\mathcal{R}}_j} \frac{\dot{\mathbf{c}}(\widehat{\boldsymbol{\beta}}, \mathbf{z}_l(t_j))^{\otimes 2}}{c(\widehat{\boldsymbol{\beta}}, \mathbf{z}_l(t_j))}\, w_l(t_j, \widetilde{\mathcal{R}}_j),$$

where $\dot{\mathbf{c}}(\boldsymbol{\beta}, \mathbf{z}_i(t)) = \partial c(\boldsymbol{\beta}, \mathbf{z}_i(t))/\partial \boldsymbol{\beta}$, and $\mathbf{v}^{\otimes 2}$ of a column vector $\mathbf{v}$ equals the matrix $\mathbf{v}\mathbf{v}^\mathsf{T}$. The main steps in the proofs of these properties for the situation with a general relative risk function are given in section 5.1. For the special case of Cox's regression model, detailed proofs are provided by [5].

3.4. Counting process formulation and martingale residuals

To derive the partial likelihood (3.5) and study the asymptotic properties of the maximum partial likelihood estimator, Borgan *et al.* [5] expressed the sampled risk set data by means of the processes

$$N_{(i,\mathbf{r})}(t) = \sum_{j \geq 1} I\{t_j \leq t, (i_j, \widetilde{\mathcal{R}}_j) = (i, \mathbf{r})\} \tag{3.9}$$

counting the observed number of failures for individual i in $(0, t]$ with associated sampled risk set $\mathbf{r}$. These counting processes are also key for deriving the martingale residual processes for sampled risk set data.

From the counting processes $N_{(i,\mathbf{r})}(t)$ we may aggregate over sets $\mathbf{r} \in \mathcal{P}_i$ to recover the counting process (2.2) registering the observed failures for the ith individual, i.e., $N_i(t) = \sum_{\mathbf{r} \in \mathcal{P}_i} N_{(i,\mathbf{r})}(t)$. In a similar manner we may for a set $\mathbf{r} \in \mathcal{P}$ aggregate over individuals $i \in \mathbf{r}$ to obtain the process

$$N_{\mathbf{r}}(t) = \sum_{i \in \mathbf{r}} N_{(i,\mathbf{r})}(t) = \sum_{j \geq 1} I\{t_j \leq t, \widetilde{\mathcal{R}}_j = \mathbf{r}\} \tag{3.10}$$

counting the number of times in $(0, t]$ the sampled risk set equals the set $\mathbf{r}$.

The assumption that not only truncation and censoring, but also the sampling of controls, are independent ensures that the intensity processes of the counting processes N_i are given by (2.3), not only w.r.t. the "cohort history" $\mathcal{H}_{t-}$, but also w.r.t. the "cohort and sampling history" $\mathcal{F}_{t-}$. From this and (3.3) it follows that the intensity process $\lambda_{(i,\mathbf{r})}(t)$ of the counting processes (3.9) takes the form

$$\lambda_{(i,\mathbf{r})}(t) = \lambda_i(t)\pi_t(\mathbf{r}\,|\,i) = Y_i(t)c(\boldsymbol{\beta}_0, \mathbf{z}_i(t))w_i(t, \mathbf{r})\pi_t(\mathbf{r})\alpha_0(t). \tag{3.11}$$

Therefore by general results for counting processes

$$M_{(i,\mathbf{r})}(t) = N_{(i,\mathbf{r})}(t) - \Lambda_{(i,\mathbf{r})}(t) \tag{3.12}$$

with

$$\Lambda_{(i,\mathbf{r})}(t) = \int_0^t \lambda_{(i,\mathbf{r})}(u)\mathrm{d}u = \int_0^t Y_i(u)c(\boldsymbol{\beta}_0, \mathbf{z}_i(u))w_i(u, \mathbf{r})\pi_u(\mathbf{r})\alpha_0(u)\mathrm{d}u \tag{3.13}$$

are local square integrable martingales. As for cohort data, we will insert estimates for $\boldsymbol{\beta}_0$ and $\alpha_0(u)\mathrm{d}u$ in (3.13) to obtain an estimated cumulative intensity process $\widehat{\Lambda}_{(i,\mathbf{r})}(t)$. For $\boldsymbol{\beta}_0$ we insert the maximum partial likelihood estimator $\widehat{\boldsymbol{\beta}}$, and for $\alpha_0(u)\mathrm{d}u$ we insert $\mathrm{d}\widehat{A}_{0\mathbf{r}}(t)$, where

$$\widehat{A}_{0\mathbf{r}}(t) = \sum_{t_j \leq t, \widetilde{\mathcal{R}}_j = \mathbf{r}} \frac{1}{\sum_{l \in \mathbf{r}} c(\widehat{\boldsymbol{\beta}}, \mathbf{z}_l(t_j)) w_l(t_j, \mathbf{r}) \pi_{t_j}(\mathbf{r})}. \tag{3.14}$$

Thus we get the estimated cumulative intensity process

$$\widehat{\Lambda}_{(i,\mathbf{r})}(t) = \int_0^t Y_i(u) c(\widehat{\boldsymbol{\beta}}, \mathbf{z}_i(u)) w_i(u, \mathbf{r}) \pi_u(\mathbf{r}) \mathrm{d}\widehat{A}_{0\mathbf{r}}(u)$$

and the corresponding *martingale residual processes*

$$\widehat{M}_{(i,\mathbf{r})}(t) = N_{(i,\mathbf{r})}(t) - \widehat{\Lambda}_{(i,\mathbf{r})}(t). \tag{3.15}$$

The martingale residual processes (3.15) are of little use in their own right, in fact most of them will be identically equal to zero. But they provide the building blocks for the grouped martingale residual processes for sampled risk set data.

3.5. Grouped martingale residual processes and a chi-squared goodness-of-fit test

As in section 2, we assume that we have a grouping of the individuals into G groups, and denote by $\mathcal{J}(u)$ the set of all individuals who belong to group J at time u; $J = 1, \ldots, G$. Then the group J martingale residual process for sampled risk set data corresponding to (2.5) is given by

$$\widehat{M}_J(t) = \int_0^t \sum_{i \in \mathcal{J}(u)} \sum_{\mathbf{r} \in \mathcal{P}_i} \mathrm{d}\widehat{M}_{(i,\mathbf{r})}(u) = \int_0^t \sum_{i \in \mathcal{J}(u)} \mathrm{d}N_i(u) - \int_0^t \sum_{\mathbf{r} \in \mathcal{P}} \sum_{i \in \mathbf{r} \cap \mathcal{J}(u)} \mathrm{d}\widehat{\Lambda}_{(i,\mathbf{r})}(u),$$

which may be rewritten as

$$\widehat{M}_J(t) = N_J(t) - \sum_{t_j \leq t} \frac{\sum_{i \in \widetilde{\mathcal{R}}_j \cap \mathcal{J}(t_j)} c(\widehat{\boldsymbol{\beta}}, \mathbf{z}_i(t_j)) w_i(t_j, \widetilde{\mathcal{R}}_j)}{\sum_{l \in \widetilde{\mathcal{R}}_j} c(\widehat{\boldsymbol{\beta}}, \mathbf{z}_l(t_j)) w_l(t_j, \widetilde{\mathcal{R}}_j)} \tag{3.16}$$

with $N_J(t) = \int_0^t \sum_{i \in \mathcal{J}(u)} \mathrm{d}N_i(u)$. As for cohort data, these grouped martingale residual processes may be interpreted as observed minus expected number of events in the given groups.

In section 5.2 we note that, if we could have used the true value $\boldsymbol{\beta}_0$ instead of its estimate $\widehat{\boldsymbol{\beta}}$ in (3.16), then the grouped martingale residual processes would have been martingales. However, since the regression coefficients have to be estimated, the grouped martingale residual processes are only approximately martingales. In section 5.2 we also show that, properly normalized, the vector of grouped martingale residual processes $(\widehat{M}_1, \ldots, \widehat{M}_G)^{\mathsf{T}}$ converges weakly to a mean zero multivariate Gaussian process. Further the covariance between $\widehat{M}_I(s)$ and $\widehat{M}_J(t)$ can be estimated by

$$\widehat{\sigma}_{IJ}(s,t) = \widehat{\phi}_{IJ}(0, s \wedge t, \widehat{\boldsymbol{\beta}}) - \widehat{\boldsymbol{\psi}}_I(0, s, \widehat{\boldsymbol{\beta}})^{\mathsf{T}} \, \mathcal{I}(\widehat{\boldsymbol{\beta}})^{-1} \, \widehat{\boldsymbol{\psi}}_J(0, t, \widehat{\boldsymbol{\beta}}), \qquad (3.17)$$

where

$$\widehat{\phi}_{IJ}(s_1, s_2, \widehat{\boldsymbol{\beta}}) = \sum_{s_1 < t_j \leq s_2} \frac{S^{(0)}_{\widetilde{\mathcal{R}}_j I}(\widehat{\boldsymbol{\beta}}, t_j)}{S^{(0)}_{\widetilde{\mathcal{R}}_j}(\widehat{\boldsymbol{\beta}}, t_j)} \left\{ \delta_{IJ} - \frac{S^{(0)}_{\widetilde{\mathcal{R}}_j J}(\widehat{\boldsymbol{\beta}}, t_j)}{S^{(0)}_{\widetilde{\mathcal{R}}_j}(\widehat{\boldsymbol{\beta}}, t_j)} \right\} \qquad (3.18)$$

with δ_{IJ} a Kronecker delta, and

$$\widehat{\boldsymbol{\psi}}_J(s_1, s_2, \widehat{\boldsymbol{\beta}}) = \sum_{s_1 < t_j \leq s_2} \left\{ \frac{\mathbf{S}^{(1)}_{\widetilde{\mathcal{R}}_j J}(\widehat{\boldsymbol{\beta}}, t_j)}{S^{(0)}_{\widetilde{\mathcal{R}}_j}(\widehat{\boldsymbol{\beta}}, t_j)} - \frac{S^{(0)}_{\widetilde{\mathcal{R}}_j J}(\widehat{\boldsymbol{\beta}}, t_j) \, \mathbf{S}^{(1)}_{\widetilde{\mathcal{R}}_j}(\widehat{\boldsymbol{\beta}}, t_j)}{S^{(0)}_{\widetilde{\mathcal{R}}_j}(\widehat{\boldsymbol{\beta}}, t_j)^2} \right\}. \qquad (3.19)$$

Here $S^{(0)}_{\widetilde{\mathcal{R}}_j J}(\widehat{\boldsymbol{\beta}}, t_j)$ and $\mathbf{S}^{(1)}_{\widetilde{\mathcal{R}}_j J}(\widehat{\boldsymbol{\beta}}, t_j)$ are given by expressions similar to (3.7) and (3.8), but with the summation restricted to individuals $l \in \widetilde{\mathcal{R}}_j \cap \mathcal{J}(t_j)$.

As will be illustrated in section 4, a plot of the grouped martingale residual processes is a useful tool for assessing the fit of the relative risk regression model (2.1). In addition the grouped martingale residual processes may be used to derive formal goodness-of fit tests. In section 7 we briefly discuss different possible goodness-of-fit tests. Here we restrict our attention to a simple chi-squared test based on a comparison of observed and expected number of events in the G groups in K disjoint time intervals. To this end let $0 = a_0 < a_1 < \cdots < a_{K-1} < a_K = \tau$ be a partitioning of the study time interval, and introduce (for $H = 1, 2, \ldots, K$ and $J = 1, 2, \ldots, G$)

$$\widehat{M}_{HJ} = \widehat{M}_J(a_H) - \widehat{M}_J(a_{H-1}) = O_{HJ} - E_{HJ}. \qquad (3.20)$$

Here $O_{HJ} = N_J(a_H) - N_J(a_{H-1})$ is the observed number of events in group J in time interval H, while

$$E_{HJ} = \sum_{a_{H-1} < t_j \leq a_H} \frac{\sum_{i \in \widetilde{\mathcal{R}}_j \cap \mathcal{J}(t_j)} c(\widehat{\boldsymbol{\beta}}, \mathbf{z}_i(t_j)) \, w_i(t_j, \widetilde{\mathcal{R}}_j)}{\sum_{l \in \widetilde{\mathcal{R}}_j} c(\widehat{\boldsymbol{\beta}}, \mathbf{z}_l(t_j)) \, w_l(t_j, \widetilde{\mathcal{R}}_j)}$$

is the corresponding expected number under model (2.1). The martingale residual processes (3.16) sum to zero at any given time t. To derive a chi-squared goodness-of-fit test, we therefore disregard the contribution from one of the groups, say the first group, and consider the $K(G-1)$-vector $\widehat{\mathbf{M}}$ with elements $\widehat{M}_{HJ}$ for $H = 1, 2, \ldots, K$; $J = 2, 3, \ldots, G$. By the large sample distributional results for the grouped martingale residual processes summarized in connection with (3.17), it follows that $\widehat{\mathbf{M}}$ is approximately mean zero multinormally distributed in large samples when model (2.1) holds true. Its covariance matrix may be estimated by the matrix $\widehat{\boldsymbol{\Sigma}} = \{\widehat{\sigma}_{LI,HJ}\}$ with elements

$$
\begin{aligned}
\widehat{\sigma}_{LI,HJ} &= \widehat{\mathrm{Cov}}(\widehat{M}_{LI}, \widehat{M}_{HJ}) \\
&= \delta_{LH}\, \widehat{\phi}_{IJ}(a_{H-1}, a_H, \widehat{\boldsymbol{\beta}}) - \widehat{\boldsymbol{\psi}}_I(a_{L-1}, a_L, \widehat{\boldsymbol{\beta}})^{\mathsf{T}} \mathcal{I}(\widehat{\boldsymbol{\beta}})^{-1} \widehat{\boldsymbol{\psi}}_J(a_{H-1}, a_H, \widehat{\boldsymbol{\beta}});
\end{aligned}
$$

$H, L = 1, 2, \ldots, K$; $J, I = 2, 3, \ldots, G$; where δ_{LH} is a Kronecker delta. Therefore a goodness-of-fit test may be based on the statistic $\chi^2 = \widehat{\mathbf{M}}^{\mathsf{T}} \widehat{\boldsymbol{\Sigma}}^{-1} \widehat{\mathbf{M}}$, which is approximately chi-squared distributed with $K(G-1)$ degrees of freedom in large samples when model (2.1) holds true.

Large sample results for the grouped martingale residual processes and the goodness-of-fit test for full cohort data, are the special cases of the above results in which the sampled risk set equals the full risk set with probability one and all weights are unity. In particular for cohort data with exponential relative risk function and only one time interval (i.e. $K = 1$), the test statistic χ^2 specializes to the goodness-of-fit statistic of Grønnesby and Borgan [8].

May and Hosmer [18] showed how the test of Grønnesby and Borgan can be obtained as the score test for the addition of categorical grouping variables. A similar result holds here as well. More specifically, consider the extension of model (2.1) where an individual i who belongs to group J at time $t \in (a_{H-1}, a_H]$ has a hazard rate of the form

$$
\alpha(t; \mathbf{z}_i) = c(\boldsymbol{\beta}_0, \mathbf{z}_i(t))\, e^{\gamma_{HJ}}\, \alpha_0(t); \tag{3.21}
$$

$J = 2, 3, \ldots, G$. Then by some straightforward, but tedious algebra along the lines of [18, Appendix A] one may show that the goodness-of-fit statistic χ^2 is algebraically equivalent to the score test for the hypothesis that all the additional $K(G-1)$ parameters γ_{HJ} in (3.21) are equal to zero.

4. An illustration

To illustrate the use of the grouped martingale residual processes and the accompanying goodness-of-fit test, we will use data on lung cancer death

among a cohort of uranium miners from the Colorado Plateau. The cohort was assembled to study the effects of radon exposure and smoking on lung cancer risk and has been described in detail in earlier publications; e.g. [9]. The cohort consists of 3,347 Caucasian male miners recruited between 1950 and 1960 and was traced for mortality outcomes through December 31, 1982, by which time 258 lung cancer deaths were observed. Exposure data include radon exposure, in working level months (WLM), and smoking histories, in number of packs of cigarettes (20 cigarettes per pack) smoked per day. We consider age as the basic time scale and summarize radon and smoking data into cumulative exposures lagged by two years. Thus we consider the covariates $\mathbf{z}(t) = (z_{i1}(t), z_{i2}(t))^\mathsf{T}$, where $z_{i1}(t)$ is cumulative radon exposure measured in working level months (WLM) up to two years prior to age t, and $z_{i2}(t)$ is cumulative smoking in number of packs smoked up to two years prior to t. Although covariate information is available on all cohort subjects, in order to illustrate the methods we selected simple random and counter-matched samples with three controls per case. These data sets are denoted 1:3 simple random and counter-matched samples, respectively. The 23 tied failure times were broken randomly so that there was only one case per risk set. Counter-matching was based on radon exposure grouped into four strata according to the quartiles of the cumulative radon exposure for the cases [14, Section 5], and one control was sampled at random from each stratum except the one of the case.

As has been the case in previous analyzes of these data (cf. [14] and its references), the excess relative risk model was used. Thus the hazard rate for miner i is assumed to take the form

$$\alpha(t; \mathbf{z}_i) = [1 + \beta_{01}\, z_{i1}(t)]\, [1 + \beta_{02}\, z_{i2}(t)]\, \alpha_0(t). \qquad (4.1)$$

For the 1:3 simple random data, the estimated radon excess relative risk (with standard error) is $\widehat{\beta}_1 = 0.556$ (0.215) per 100 WLMs cumulative radon exposure, while the smoking excess relative risk is $\widehat{\beta}_2 = 0.276$ (0.093) per 1000 packs of cigarettes smoked. For the 1:3 counter-matched data, the estimates become $\widehat{\beta}_1 = 0.420$ (0.137) and $\widehat{\beta}_2 = 0.205$ (0.068).

Figure 1 shows the grouped martingale residual processes (3.16) for both data sets when the individuals are aggregated over groups defined by cumulative radon exposure (group I: below 500 WLMs; group II: 500–1500 WLMs; group III: above 1500 WLMs), while Table 1 summarizes the observed and expected number of lung cancer deaths in the three radon exposure groups for ages below and above 60 years. From the plots and the table it is seen that more lung cancer deaths than expected occur in the high exposure group
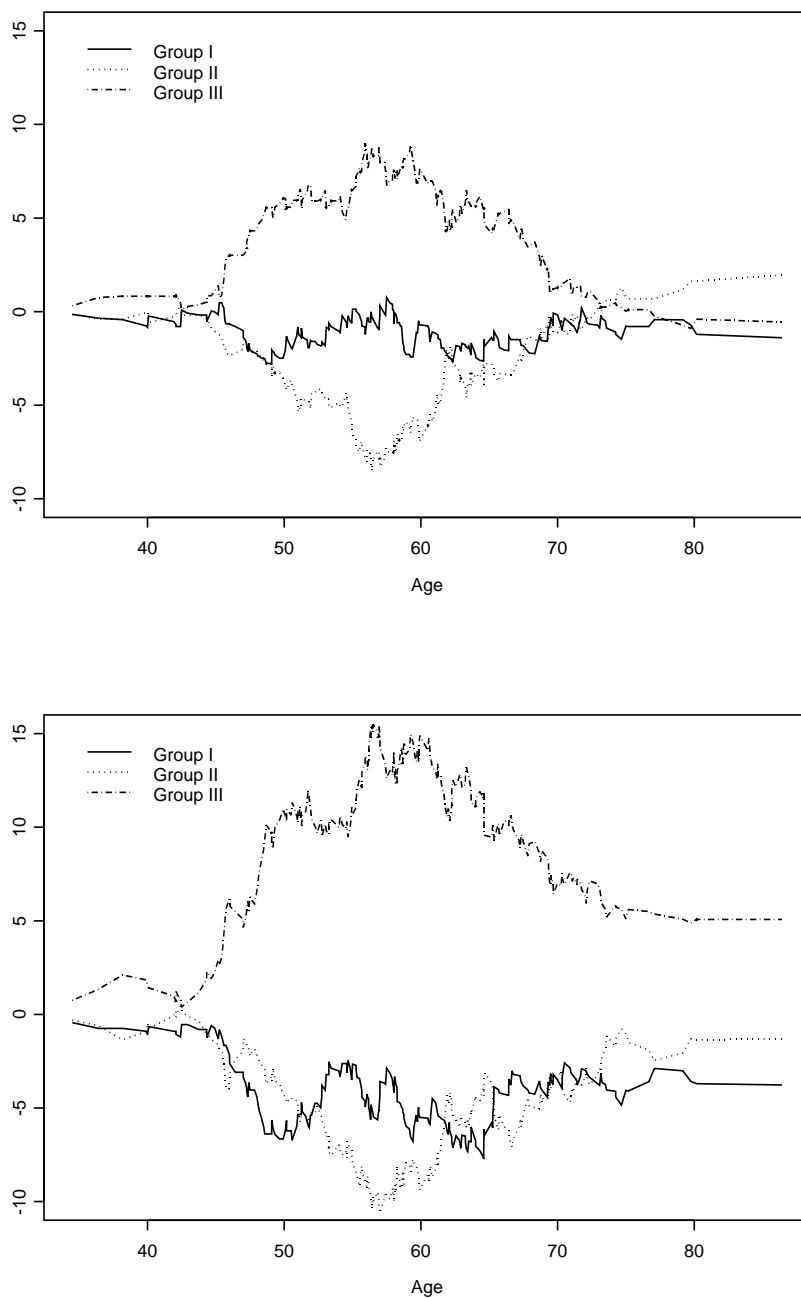
FIG 1. *Grouped martingale residual processes for the uranium miners based on a 1:3 simple random sample (upper panel) and a 1:3 counter-matched sample (lower panel). Grouping is done according to cumulative radon exposure: Group I: below 500 WLMs; group II: 500–1500 WLMs; group III: above 1500 WLMs.*

TABLE 1
Observed and expected number of lung cancer deaths.

Exposure group[a]	Observed numbers	Expected numbers	
		1:3 simple	1:3 counter-matched
Below 60 years of age			
Group I	30	30.7	35.5
Group II	39	45.9	48.4
Group III	81	73.4	66.1
Above 60 years of age			
Group I	27	27.7	25.3
Group II	45	36.1	36.9
Group III	36	44.2	45.8

a) Group I: below 500 WLMs; group II: 500–1500 WLMs;
group III: above 1500 WLMs.

(group III) below the age of 60 years, while fewer cases than expected occur above this age, the pattern being most pronounced for the counter-matched data. The chi-squared goodness-of-fit statistic with $2(3-1) = 4$ degrees of freedom based on the observed and expected numbers of Table 1 takes the values 10.5 and 14.2, respectively, for the 1:3 simple random sample and the 1:3 counter-matched sample, with corresponding P-values 0.032 and 0.007. Thus our analysis shows that the excess relative risk model (4.1), where the effect of radon depends linearly on cumulative exposure, is too simplistic.

The lack of fit is further illustrated in Table 2 for the 1:3 simple random sample. The table shows relative risks within radon exposure categories for individuals below and above 60 years of age, as well as the relative risks predicted by the excess relative risk model (4.1). Prior to age 60 years, lung cancer mortality rates increase faster than linear with radon exposure level while after age 60, the dose response is quite a bit slower than linear. There are a number of possible ways one might choose to accommodate this pattern of rates. One could simply accommodate the variation in a model that allows for changing shape of the dose response curve with age. But, since miners tended to experience the larger exposures at earlier ages, the observed change in exposure response curve shape with age may well be due to the time since exposure. Thus, a biologically appealing approach would be to summarize the exposure history in a way that accounts for the time since exposure (latency) and, perhaps, rate of exposure. In fact, it has been found that latency effects are a significant component in describing radiation exposure and lung cancer risk in the Colorado Plateau miners, e.g. [17, 15].

TABLE 2
Relative risks within categories of cumulative radon exposure by age at lung cancer death and relative risks predicted by the excess relative risk model (4.1). 1:3 simple random sampling.

Radon exposure category[a]	Mean exposure within category[b]	Relative risks for categorical model	Relative risks[c] predicted by model (4.1)
Below 60 years of age			
Group I	180	1	1
Group II	896	2.35	2.99
Group III	2885	10.91	8.51
Above 60 years of age			
Group I	187	1	1
Group II	923	3.89	3.01
Group III	3034	5.61	8.76

a) Group I: below 500 WLMs; group II: 500–1500 WLMs;
   group III: above 1500 WLMs.
b) Mean among controls.
c) Computed at category mean, normalized to mean of first category.

It is, however, beyond the scope of this paper to pursue such alternative models. Here we are content with the above illustration of how the grouped residual process plots play a useful role by identifying model lack of fit and by suggesting candidate changes that may yield a better fitting model.

5. Outline of proofs for sampled risk set data

In this section we give an outline of the proofs of the large sample properties of the maximum partial likelihood estimator $\widehat{\boldsymbol{\beta}}$ and the grouped martingale residual processes for sampled risk set data when we have a general relative risk function. Formal proofs may be written out along the lines of [5], that gives detailed proofs of the large sample properties of the maximum partial likelihood estimator for the special case of an exponential relative risk function.

5.1. Large sample properties of $\widehat{\boldsymbol{\beta}}$

The estimator $\widehat{\boldsymbol{\beta}}$ is the solution to $\mathbf{U}(\boldsymbol{\beta}) = \mathbf{0}$, where $\mathbf{U}(\boldsymbol{\beta}) = \partial \log L(\boldsymbol{\beta})/\partial \boldsymbol{\beta}$ is the vector of score functions, and $L(\boldsymbol{\beta})$ is the partial likelihood (3.5). Using counting process notation, the vector of score functions may be expressed

as

$$\mathbf{U}(\boldsymbol{\beta}) = \sum_{\mathbf{r}\in\mathcal{P}}\sum_{i\in\mathbf{r}}\int_0^\tau \left\{ \frac{\dot{\mathbf{c}}(\boldsymbol{\beta},\mathbf{z}_i(u))}{c(\boldsymbol{\beta},\mathbf{z}_i(u))} - \frac{\mathbf{S}_{\mathbf{r}}^{(1)}(\boldsymbol{\beta},u)}{S_{\mathbf{r}}^{(0)}(\boldsymbol{\beta},u)} \right\} \mathrm{d}N_{(i,\mathbf{r})}(u),$$

where τ is the terminal study time, and

$$S_{\mathbf{r}}^{(0)}(\boldsymbol{\beta},u) = \sum_{l\in\mathbf{r}} Y_l(u)\,c(\boldsymbol{\beta},\mathbf{z}_l(u))\,w_l(u,\mathbf{r}), \tag{5.1}$$

$$\mathbf{S}_{\mathbf{r}}^{(1)}(\boldsymbol{\beta},u) = \sum_{l\in\mathbf{r}} Y_l(u)\,\dot{\mathbf{c}}(\boldsymbol{\beta},\mathbf{z}_l(u))\,w_l(u,\mathbf{r}). \tag{5.2}$$

Further the observed partial information matrix $\mathbf{I}(\boldsymbol{\beta}) = -\partial\mathbf{U}(\boldsymbol{\beta})/\partial\boldsymbol{\beta}^{\mathsf{T}}$ becomes

$$\mathbf{I}(\boldsymbol{\beta}) = \sum_{\mathbf{r}\in\mathcal{P}}\sum_{i\in\mathbf{r}}\int_0^\tau \frac{\partial}{\partial\boldsymbol{\beta}^{\mathsf{T}}}\left\{ \frac{\mathbf{S}_{\mathbf{r}}^{(1)}(\boldsymbol{\beta},u)}{S_{\mathbf{r}}^{(0)}(\boldsymbol{\beta},u)} - \frac{\dot{\mathbf{c}}(\boldsymbol{\beta},\mathbf{z}_i(u))}{c(\boldsymbol{\beta},\mathbf{z}_i(u))} \right\} \mathrm{d}N_{(i,\mathbf{r})}(u). \tag{5.3}$$

If we evaluate the score function at $\boldsymbol{\beta}_0$, we find by some straightforward algebra [using (3.12) and (3.13)]:

$$\mathbf{U}(\boldsymbol{\beta}_0) = \sum_{\mathbf{r}\in\mathcal{P}}\sum_{i\in\mathbf{r}}\int_0^\tau \left\{ \frac{\dot{\mathbf{c}}(\boldsymbol{\beta}_0,\mathbf{z}_i(u))}{c(\boldsymbol{\beta}_0,\mathbf{z}_i(u))} - \frac{\mathbf{S}_{\mathbf{r}}^{(1)}(\boldsymbol{\beta}_0,u)}{S_{\mathbf{r}}^{(0)}(\boldsymbol{\beta}_0,u)} \right\} \mathrm{d}M_{(i,\mathbf{r})}(u). \tag{5.4}$$

Here the integrands are predictable processes. Thus the score function is a sum of (vector-valued) stochastic integrals when evaluated at the true value of the regression coefficients. If, on the right hand side of (5.4), we replace the upper limit of integration by t, we get a stochastic process. This stochastic process is a martingale with a predictable variation process that evaluated at τ becomes

$$\langle \mathbf{U}(\boldsymbol{\beta}_0)\rangle(\tau) = \sum_{\mathbf{r}\in\mathcal{P}}\sum_{i\in\mathbf{r}}\int_0^\tau \left\{ \frac{\dot{\mathbf{c}}(\boldsymbol{\beta}_0,\mathbf{z}_i(u))}{c(\boldsymbol{\beta}_0,\mathbf{z}_i(u))} - \frac{\mathbf{S}_{\mathbf{r}}^{(1)}(\boldsymbol{\beta},u)}{S_{\mathbf{r}}^{(0)}(\boldsymbol{\beta},u)} \right\}^{\otimes 2} \lambda_{(i,\mathbf{r})}(u)\mathrm{d}u.$$

Using (3.11), we get after some straightforward algebra that

$$\langle \mathbf{U}(\boldsymbol{\beta}_0)\rangle(\tau) = \sum_{\mathbf{r}\in\mathcal{P}}\int_0^\tau \mathbf{V}_{\mathbf{r}}(\boldsymbol{\beta}_0,u)\,S_{\mathbf{r}}^{(0)}(\boldsymbol{\beta},u)\,\pi_u(\mathbf{r})\,\alpha_0(u)\mathrm{d}u, \tag{5.5}$$

where

$$\mathbf{V}_{\mathbf{r}}(\boldsymbol{\beta},u) = \frac{\mathbf{S}_{\mathbf{r}}^{(2)}(\boldsymbol{\beta},u)}{S_{\mathbf{r}}^{(0)}(\boldsymbol{\beta},u)} - \left(\frac{\mathbf{S}_{\mathbf{r}}^{(1)}(\boldsymbol{\beta},u)}{S_{\mathbf{r}}^{(0)}(\boldsymbol{\beta},u)} \right)^{\otimes 2} \tag{5.6}$$

with $S_{\mathbf{r}}^{(0)}(\boldsymbol{\beta}, u)$ and $\mathbf{S}_{\mathbf{r}}^{(1)}(\boldsymbol{\beta}, u)$ given by (5.1) and (5.2), respectively, and

$$\mathbf{S}_{\mathbf{r}}^{(2)}(\boldsymbol{\beta}, u) = \sum_{l \in \mathbf{r}} Y_l(u) \frac{\dot{\mathbf{c}}(\boldsymbol{\beta}, \mathbf{z}_l(u))^{\otimes 2}}{c(\boldsymbol{\beta}, \mathbf{z}_l(u))} w_l(u, \mathbf{r}).$$

If we insert $dN_{(i,\mathbf{r})}(u) = \lambda_{(i,\mathbf{r})}(u)du + dM_{(i,\mathbf{r})}(u)$ [cf. (3.12)] and use (3.11) in (5.3), we find after some algebra that the observed information matrix evaluated at $\boldsymbol{\beta}_0$ may be decomposed as

$$\mathbf{I}(\boldsymbol{\beta}_0) = \langle \mathbf{U}(\boldsymbol{\beta}_0) \rangle (\tau) + \sum_{\mathbf{r} \in \mathcal{P}} \sum_{i \in \mathbf{r}} \int_0^\tau \frac{\partial}{\partial \boldsymbol{\beta}^{\mathsf{T}}} \left\{ \frac{\mathbf{S}_{\mathbf{r}}^{(1)}(\boldsymbol{\beta}_0, u)}{S_{\mathbf{r}}^{(0)}(\boldsymbol{\beta}_0, u)} - \frac{\dot{\mathbf{c}}(\boldsymbol{\beta}_0, \mathbf{z}_i(u))}{c(\boldsymbol{\beta}_0, \mathbf{z}_i(u))} \right\} dM_{(i,\mathbf{r})}(u).$$

Thus, at the true value of the regression coefficient, the observed information matrix equals the predictable variation process of the score function plus a stochastic integral.

By the martingale central limit theorem, we may now show, under suitable regularity conditions, that $n^{-1/2}\mathbf{U}(\boldsymbol{\beta}_0)$ converges weakly to a multinormal distribution with mean zero and a covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{\beta}}$ that is the limit in probability of $n^{-1} \langle \mathbf{U}(\boldsymbol{\beta}_0) \rangle (\tau)$. We may also show that both $n^{-1}\mathbf{I}(\boldsymbol{\beta}_0)$ and $n^{-1}\mathbf{I}(\widehat{\boldsymbol{\beta}})$ converge in probability to $\boldsymbol{\Sigma}_{\boldsymbol{\beta}}$. From these results the large sample properties of $\widehat{\boldsymbol{\beta}}$ follow in the usual way. The main steps in the derivations are as follows. Since $\widehat{\boldsymbol{\beta}}$ is the solution to the score equation $\mathbf{U}(\widehat{\boldsymbol{\beta}}) = \mathbf{0}$, a Taylor expansion of the score equation around $\boldsymbol{\beta}_0$ gives $\mathbf{0} = \mathbf{U}(\widehat{\boldsymbol{\beta}}) \approx \mathbf{U}(\boldsymbol{\beta}_0) - \mathbf{I}(\boldsymbol{\beta}_0)(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)$. From this we obtain

$$\sqrt{n} \left(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0 \right) \approx \left(n^{-1}\mathbf{I}(\boldsymbol{\beta}_0) \right)^{-1} n^{-1/2}\mathbf{U}(\boldsymbol{\beta}_0) \approx \boldsymbol{\Sigma}_{\boldsymbol{\beta}}^{-1} n^{-1/2}\mathbf{U}(\boldsymbol{\beta}_0), \qquad (5.7)$$

and it follows that $\sqrt{n}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)$ converges weakly to a multinormal distribution with mean zero and covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{\beta}}^{-1}\boldsymbol{\Sigma}_{\boldsymbol{\beta}}\boldsymbol{\Sigma}_{\boldsymbol{\beta}}^{-1} = \boldsymbol{\Sigma}_{\boldsymbol{\beta}}^{-1}$. Thus, in large samples, $\widehat{\boldsymbol{\beta}}$ is approximately multinormally distributed around $\boldsymbol{\beta}_0$ with covariance matrix $n^{-1}\boldsymbol{\Sigma}_{\boldsymbol{\beta}}^{-1}$.

In order to estimate the covariance matrix of $\widehat{\boldsymbol{\beta}}$ we may use $\mathbf{I}(\widehat{\boldsymbol{\beta}})^{-1}$, the inverse of the observed information, or we may use the inverse of the (estimated) expected information matrix. The (estimated) expected information is obtained from (5.5) by inserting $\widehat{\boldsymbol{\beta}}$ for $\boldsymbol{\beta}_0$ and the increment

$$d\widehat{A}_{0\mathbf{r}}(u) = \frac{dN_{\mathbf{r}}(u)}{S_{\mathbf{r}}^{(0)}(\widehat{\boldsymbol{\beta}}, u)\pi_u(\mathbf{r})}$$

of the Breslow type estimator (3.14) for $\alpha_0(u)\mathrm{d}u$ to get

$$\mathcal{I}(\widehat{\boldsymbol{\beta}}) = \sum_{\mathbf{r}\in\mathcal{P}} \int_0^\tau \mathbf{V}_{\mathbf{r}}(\widehat{\boldsymbol{\beta}}, u)\, \mathrm{d}N_{\mathbf{r}}(u), \tag{5.8}$$

where $\mathbf{V}_{\mathbf{r}}(\boldsymbol{\beta}, u)$ is given by (5.6). This justifies (3.6) of section 3.3. By (5.8) and (5.3) we note that while the expected information matrix depends only on quantities that are aggregates over each sampled risk set, the observed information matrix depends specifically on the covariates of the cases. Therefore the expected information matrix tends to be the most stable of the two, and it is the one we recommend. For Cox's regression model the observed and expected information matrices coincide.

5.2. Large sample properties of the grouped martingale residual processes

We will derive similar large sample properties for the grouped martingale residuals for sampled risk set data as those of Grønnesby and Borgan [8] for Cox regression with cohort data. To this end we first note that the grouped martingale residual processes (3.16) may be given as

$$\widehat{M}_J(t) = \sum_{\mathbf{r}\in\mathcal{P}} \int_0^t \sum_{i\in\mathbf{r}\cap\mathcal{J}(u)} \mathrm{d}N_{(i,\mathbf{r})}(u) - \sum_{\mathbf{r}\in\mathcal{P}} \int_0^t \frac{S_{\mathbf{r}J}^{(0)}(\widehat{\boldsymbol{\beta}}, u)}{S_{\mathbf{r}}^{(0)}(\widehat{\boldsymbol{\beta}}, u)}\, \mathrm{d}N_{\mathbf{r}}(u), \tag{5.9}$$

where $S_{\mathbf{r}}^{(0)}(\boldsymbol{\beta}, u)$ is given by (5.1) and

$$S_{\mathbf{r}J}^{(0)}(\boldsymbol{\beta}, u) = \sum_{l\in\mathbf{r}\cap\mathcal{J}(u)} Y_l(u) c(\boldsymbol{\beta}, \mathbf{z}_l(u)) w_l(u, \mathbf{r}).$$

We also note that the intensity process of the counting process $N_{\mathbf{r}}(t)$ given by (3.10) takes the form

$$\lambda_{\mathbf{r}}(t) = \sum_{i\in\mathbf{r}} \lambda_{(i,\mathbf{r})}(t) = S_{\mathbf{r}}^{(0)}(\boldsymbol{\beta}_0, t)\, \pi_t(\mathbf{r})\, \alpha_0(t), \tag{5.10}$$

where we have used (3.11) and (5.1) to get the last equality. We also introduce the martingales

$$M_{\mathbf{r}}(t) = \sum_{i\in\mathbf{r}} M_{(i,\mathbf{r})}(t) = N_{\mathbf{r}}(t) - \int_0^t \lambda_{\mathbf{r}}(u)\mathrm{d}u. \tag{5.11}$$

Then, using (3.12), (3.13), (5.1), and (5.9) – (5.11), we find after some straightforward algebra that the normalized grouped martingale residual processes may be decomposed as

$$n^{-1/2}\widehat{M}_J(t) = \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad (5.12)$$

$$X_J^*(t) - n^{-1/2}\sum_{\mathbf{r}\in\mathcal{P}}\int_0^t\left\{\frac{S_{\mathbf{r}J}^{(0)}(\widehat{\boldsymbol{\beta}},u)}{S_{\mathbf{r}}^{(0)}(\widehat{\boldsymbol{\beta}},u)} - \frac{S_{\mathbf{r}J}^{(0)}(\boldsymbol{\beta}_0,u)}{S_{\mathbf{r}}^{(0)}(\boldsymbol{\beta}_0,u)}\right\}\mathrm{d}N_{\mathbf{r}}(u),$$

where

$$X_J^*(t) = n^{-1/2}\sum_{\mathbf{r}\in\mathcal{P}}\sum_{i\in\mathbf{r}}\int_0^t\left\{\delta_{iJ}(u) - \frac{S_{\mathbf{r}J}^{(0)}(\boldsymbol{\beta}_0,u)}{S_{\mathbf{r}}^{(0)}(\boldsymbol{\beta}_0,u)}\right\}\mathrm{d}M_{(i,\mathbf{r})}(u)$$

with $\delta_{iJ}(u) = 1$ if $i \in \mathcal{J}(u)$, i.e. if individual i belongs to group J at time u, and $\delta_{iJ}(u) = 0$ otherwise. Note that $X_J^*(t)$ is a stochastic integral, and hence itself a martingale. Thus the grouped martingale residual processes would have been martingales if we could use the true value $\boldsymbol{\beta}_0$ instead of its estimate $\widehat{\boldsymbol{\beta}}$ in (5.9).

We now take a closer look at the last term in (5.12). By a Taylor series expansion, one may show that this term is asymptotically equivalent to

$$-n^{-1}\sum_{\mathbf{r}\in\mathcal{P}}\int_0^t\frac{\partial}{\partial\boldsymbol{\beta}^\mathsf{T}}\left\{\frac{S_{\mathbf{r}J}^{(0)}(\boldsymbol{\beta}_0,u)}{S_{\mathbf{r}}^{(0)}(\boldsymbol{\beta}_0,u)}\right\}\mathrm{d}N_{\mathbf{r}}(u)\sqrt{n}\left(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0\right).$$

Now, using (5.10) and (5.11), the latter expression may be shown to be asymptotically equivalent to $-\boldsymbol{\psi}_J(0,t,\boldsymbol{\beta}_0)^\mathsf{T}\sqrt{n}\left(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0\right)$, where $\boldsymbol{\psi}_J(s_1,s_2,\boldsymbol{\beta}_0)$ is the uniform (in s_1 and s_2) limit in probability of

$$n^{-1}\sum_{\mathbf{r}\in\mathcal{P}}\int_{s_1}^{s_2}\left\{\frac{\mathbf{S}_{\mathbf{r}J}^{(1)}(\boldsymbol{\beta}_0,u)}{S_{\mathbf{r}}^{(0)}(\boldsymbol{\beta}_0,u)} - \frac{S_{\mathbf{r}J}^{(0)}(\boldsymbol{\beta}_0,u)\,\mathbf{S}_{\mathbf{r}}^{(1)}(\boldsymbol{\beta}_0,u)}{S_{\mathbf{r}}^{(0)}(\boldsymbol{\beta}_0,u)^2}\right\}S_{\mathbf{r}}^{(0)}(\boldsymbol{\beta}_0,u)\pi_u(\mathbf{r})\alpha_0(u)\mathrm{d}u,$$

$$(5.13)$$

and

$$\mathbf{S}_{\mathbf{r}J}^{(1)}(\boldsymbol{\beta},u) = \sum_{l\in\mathbf{r}\cap\mathcal{J}(u)}Y_l(u)\dot{\mathbf{c}}(\boldsymbol{\beta},\mathbf{z}_l(u))w_l(u,\mathbf{r}).$$

Further, using (5.4) and (5.7), one may show that $\sqrt{n}\left(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0\right)$ is asymptotically equivalent to $\boldsymbol{\Sigma}_{\boldsymbol{\beta}}^{-1}\mathbf{X}^{**}(\tau)$, where

$$\mathbf{X}^{**}(t) = n^{-1/2}\sum_{\mathbf{r}\in\mathcal{P}}\sum_{i\in\mathbf{r}}\int_0^t\left\{\frac{\dot{\mathbf{c}}(\boldsymbol{\beta}_0,\mathbf{z}_i(u))}{c(\boldsymbol{\beta}_0,\mathbf{z}_i(u))} - \frac{\mathbf{S}_{\mathbf{r}}^{(1)}(\boldsymbol{\beta}_0,u)}{S_{\mathbf{r}}^{(0)}(\boldsymbol{\beta}_0,u)}\right\}\mathrm{d}M_{(i,\mathbf{r})}(u)$$

Combining all this, we get from (5.12) that, for $J = 1, \ldots, G$, the normalized martingale residual processes $n^{-1/2}\, \widehat{M}_J(t)$ are asymptotically equivalent (as stochastic processes in t) to

$$X_J(t) = X_J^*(t) - \boldsymbol{\psi}_J(t, \boldsymbol{\beta}_0)^{\mathsf{T}} \boldsymbol{\Sigma}_{\boldsymbol{\beta}}^{-1}\, \mathbf{X}^{**}(\tau). \tag{5.14}$$

Now $X_J^*(t)$ and $\mathbf{X}^{**}(t)$ are linear combinations of stochastic integrals, and hence themselves martingales. For given groups $I, J = 1, \ldots, G$ we find after some algebra that the predictable (co)variation process of the first of these martingales takes the form

$$\langle X_I^*, X_J^* \rangle(t) = \tag{5.15}$$

$$n^{-1} \sum_{\mathbf{r} \in \mathcal{P}} \int_0^t \frac{S_{\mathbf{r}I}^{(0)}(\boldsymbol{\beta}_0, u)}{S_{\mathbf{r}}^{(0)}(\boldsymbol{\beta}_0, u)} \left\{ \delta_{IJ} - \frac{S_{\mathbf{r}J}^{(0)}(\boldsymbol{\beta}_0, u)}{S_{\mathbf{r}}^{(0)}(\boldsymbol{\beta}_0, u)} \right\} S_{\mathbf{r}}^{(0)}(\boldsymbol{\beta}_0, u) \pi_u(\mathbf{r}) \alpha_0(u) \mathrm{d}u,$$

with $\delta_{IJ} = 1$ if $I = J$, and $\delta_{IJ} = 0$ otherwise, while $\langle X_J^*, \mathbf{X}^{**} \rangle(t)$ equals (5.13) and $\langle \mathbf{X}^{**} \rangle(\tau) = \langle \mathbf{U}(\boldsymbol{\beta}_0) \rangle (\tau)$ is given by (5.5). If, on the right hand side of (5.15), we integrate over $(s_1, s_2]$ instead of $(0, t]$, one may show that the resulting integral converges uniformly (in s_1 and s_2) to a limit function $\phi_{IJ}(s_1, s_2, \boldsymbol{\beta}_0)$, say. By (5.14) and the above results we may now conclude, using the martingale central limit theorem, that the normalized vector of grouped martigale residual processes $n^{-1/2}(\widehat{M}_1, \ldots, \widehat{M}_G)^{\mathsf{T}}$ converges weakly to a mean zero Gaussian process $\mathbf{U} = (U_1, \ldots, U_G)^{\mathsf{T}}$. The (I, J)-th entry of the covariance matrix $\boldsymbol{\Sigma}(s, t) = \mathrm{E}\{\mathbf{U}(s)^{\mathsf{T}}\mathbf{U}(t)\}$ between $\mathbf{U}(s)$ and $\mathbf{U}(t)$ becomes

$$
\begin{aligned}
\sigma_{IJ}(s, t) &= \mathrm{Cov}(U_I(s), U_J(t)) \\
&= \phi_{IJ}(0, s \wedge t, \boldsymbol{\beta}_0) - \boldsymbol{\psi}_I(0, s, \boldsymbol{\beta}_0)^{\mathsf{T}} \boldsymbol{\Sigma}_{\boldsymbol{\beta}}^{-1} \boldsymbol{\psi}_J(0, t, \boldsymbol{\beta}_0),
\end{aligned} \tag{5.16}
$$

where $\boldsymbol{\psi}_J(s_1, s_2, \boldsymbol{\beta}_0)$ and $\phi_{IJ}(s_1, s_2, \boldsymbol{\beta}_0)$ are defined just above (5.13) and just below (5.15), respectively. For estimation of the covariances (5.16), we may estimate $\boldsymbol{\Sigma}_{\boldsymbol{\beta}}$ consistently by $1/n$ times the expected information matrix (5.8). Further, using (5.10) and (5.11), one may prove that $\phi_{IJ}(s_1, s_2, \boldsymbol{\beta}_0)$ can be estimated uniformly (in s_1 and s_2) consistently by $1/n$ times (3.18), while $\boldsymbol{\psi}_J(s_1, s_2, \boldsymbol{\beta}_0)$ can be estimated uniformly consistently by $1/n$ times (3.19). Combining this, it follows that the asymptotic covariances (5.16) may be estimated uniformly consistently by $n^{-1}\widehat{\sigma}_{IJ}(s, t)$, where $\widehat{\sigma}_{IJ}(s, t)$ is given by (3.17) in section 3.4.

6. Matched risk set sampling

In order to keep the presentation simple, we have so far considered the relative risk regression model (2.1), where the baseline hazard rate is assumed to be the same for all individuals in the cohort. Sometimes this may not be reasonable, e.g., to control for the effect of one or more confounding factors, one may want to adopt a stratified version of (2.1) where the baseline hazard differs between (possibly time-dependent) population strata generated by the confounders. The regression coefficients are, however, assumed the same across these strata. Thus the hazard rate of an individual i from population stratum h is assumed to take the form

$$\alpha(t; \mathbf{z}_i) = c(\boldsymbol{\beta}_0, \mathbf{z}_i(t)) \, \alpha_{0h}(t). \tag{6.1}$$

When the stratified proportional hazards model (6.1) applies, the sampling of controls should be restricted to those at risk in the same population stratum as the case. We say that the controls are *matched* by the stratification variable. In particular for simple random sampling, if an individual in population stratum h fails at time t, one selects at random $m - 1$ controls from the $n^{(h)}(t) - 1$ non-failing individuals at risk in this population stratum. Similarly one may combine matching and counter-matching by selecting the controls among those in the sampling strata used for counter-matching who belong to the population stratum of the case. Note the distinction between the population strata, which form the basis for stratification in (6.1), and the sampling strata used for the counter-matched sampling of the controls.

   In general matched risk set sampling may be described as follows. Given $\mathcal{F}_{t-}$, if an individual i in population stratum h fails at time t, we select our sampled risk set according to a probability distribution $\pi_t(\mathbf{r} \,|\, i)$ over sets $\mathbf{r}$ that contain i and where *all individuals in* $\mathbf{r}$ *belong to population stratum h at time t.* (Note that the sampling distribution will depend on the population stratum h of the failing individual, even though this is not made explicit in the notation.) For such sampling distributions we have the factorization $\pi_t(\mathbf{r} \,|\, i) = w_i(t, \mathbf{r}) \, \pi_t(\mathbf{r})$, where $\pi_t(\mathbf{r})$ is given by (3.1) with $n(t)$ replaced by $n^{(h)}(t)$, and $w_i(t, \mathbf{r})$ is obtained from (3.2) as before. In particular for matched risk set sampling with simple random sampling of the controls, the weights are $w_i(t, \mathbf{r}) = [n^{(h)}(t)/m] I\{i \in \mathbf{r}\}$ for individuals in population stratum h, while for matched risk set sampling with counter-matched sampling of the controls the weights are $w_i(t, \mathbf{r}) = [n_{s(i)}^{(h)}(t)/m_{s(i)}] I\{i \in \mathbf{r}\}$. Here $s(i)$ denotes the sampling stratum of individual i, while $n_s^{(h)}(t)$ is the number of individuals at risk "just before" time t in population stratum h who belong to sampling stratum s.

The general theory of sections 3 and 5 goes through almost unchanged for matched risk set sampling. In particular the partial likelihood (3.5) and the formula (3.6) for the expected information matrix apply without modification provided one uses the appropriate weights as just described. Also the expressions (3.16) and (3.18) for the grouped martingale residuals and their estimated covariances, as well as the chi-squared goodness-of-fit test derived from these expressions, remain valid for matched risk set sampling.

In order to prove these extensions of the results of sections 3 and 5, we have to consider the processes $N^{(h)}_{(i,\mathbf{r})}(t)$, counting the observed number of failures for individual i with associated sampled risk set $\mathbf{r}$ *while being a member of population stratum h*, and their associated (cumulative) intensity processes and martingales. The proofs follow step by step the arguments of sections 3 and 5, and we omit the details.

7. Discussion

We have shown how plots of grouped martingale residual processes and the accompanying chi-squared goodness-of-fit test provide useful tools for checking the fit of relative risk regression models based on sampled risk set data. However, a number of questions remain to be better understood in relation with these methods.

To use the methods one has to define a (possibly time-dependent) grouping of the individuals, and it is then a question how this best can be done. If the grouping is based on current covariate values, one has to decide which covariates to use for the grouping and how the cut points should be chosen. Another option is to follow the approach of Grønnesby and Borgan [8] and group the individuals according to their values of the estimated relative risks $c(\widehat{\boldsymbol{\beta}}, \mathbf{z}_i(t))$. As these depend on $\widehat{\boldsymbol{\beta}}$, such a grouping will violate our assumption that the grouping at time t should only depend on information available "just before" time t. We conjecture, however, that the large sample distributions of the grouped martingale residual processes and the accompanying chi-squared goodness-of-fit test can still be used as approximations in large samples, but simulation studies are needed to investigate this further.

A useful feature of the grouped martingale residual process plots is that they show how deviations from the model may change over time. For instance, in the uranium miners example, we saw how the highest radon exposure group had more observed lung cancer deaths than expected for ages below 60 years and fewer thereafter. Such deviations give useful hints as to how the model may be modified to obtain a better fit. However, a better

understanding is needed on how various deviations from the relative risk regression model (2.1) will turn up in the plots.

For the special case of an exponential relative risk function, one may use standard software for Cox regression to maximize the partial likelihood (3.5), formally treating the label of the sampled risk sets as a stratification variable in the Cox regression and including the $\log w_l(t_j, \widetilde{\mathcal{R}}_j)$ as offsets in the model. The package Epicure fits a wide variety of relative risk functions $c(\boldsymbol{\beta}, \mathbf{z}_i(t))$ and was used to estimate the parameters for the uranium miners data in section 4. But available statistical packages are in general not able to perform all the computations needed for the grouped martingale residual process plots and accompanying chi-squared goodness-of-fit test, and the computations in section 4 were done in separate programs written by the authors for SAS and for S-Plus. However, for the special case of Cox's regression model, the extended model (3.21) becomes a Cox model as well, and our chi-squared goodness-of-fit test can be computed as the score test for the addition of categorical grouping variables using standard software for Cox regression.

Our chi-squared goodness-of-fit test is based on a comparison of observed and expected number of failures in cells obtained by partitioning the space of covariates and time. This is in line with the test suggested by Schoenfeld [21] for Cox's regression model with cohort data. In fact, apart from details in the estimation of covariances, Schoenfeld's test is the special case of ours in which the relative risk function is exponential and the entire risk set is sampled with probability one. In order to use our chi-squared test, one has to decide on a grouping according to both covariates and time. As an alternative one may group only according to covariates and consider the maximum value of the chi-squared statistic over a time interval $[\tau_1, \tau_2] \subset (0, \tau]$. P-values for such a supremum type test statistic should be obtainable by the simulation approach of Lin, Wei and Ying [16] based on the asymptotic representation (5.14) of the grouped martingale residual processes. It should even be possible to avoid the grouping according to covariates by using the individual martingale residual processes $\widehat{M}_{(i,\mathbf{r})}(t)$ [cf. (3.15)] to derive cumulative sum of martingale-based residuals along the line of Lin *et al.* [16].

Acknowledgements

and Statistics, University of Newcastle upon Tyne, in the spring of 2005. The departments are acknowledged for their hospitality and for providing the best working facilities, while the University of Oslo and Johan and Mimi Wessmann's foundation are acknowledged for their economic support which made this stay possible.

References

[1] AALEN, O. O. (1989). A linear regression model for the analysis of life times. *Statist. Med.* **8**, 907-925.

[2] AALEN, O. O. (1993). Further results on the non-parametric linear regression model in survival analysis. *Statist. Med.* **12**, 1569–1588.

[3] ANDERSEN, P. K., BORGAN, Ø., GILL, R. D., AND KEIDING, N. (1993). *Statistical models based on counting processes.* Springer Series in Statistics. Springer-Verlag, New York. MR1198884

[4] BARLOW, W. E. AND PRENTICE, R. L. (1988). Residuals for relative risk regression. *Biometrika* **75**, 65–74. MR932818

[5] BORGAN, Ø., GOLDSTEIN, L., AND LANGHOLZ, B. (1995). Methods for the analysis of sampled cohort data in the Cox proportional hazards model. *Ann. Statist.* **23**, 1749–1778. MR1370306

[6] FLEMING, T. R. AND HARRINGTON, D. P. (1991). *Counting processes and survival analysis.* Wiley Series in Probability and Mathematical Statistics: Applied Probability and Statistics. John Wiley & Sons Inc., New York. MR1100924

[7] GRAMBSCH, P. M., THERNEAU, T. M., AND FLEMING, T. R. (1995). Diagnostic plots to reveal functional form for covariates in multiplicative intensity models. *Biometrics* **51**, 1469–1482.

[8] GRØNNESBY, J. K. AND BORGAN, Ø. (1996). A method for checking regression models in survival analysis based on the risk score. *Lifetime Data Anal.* **2**, 315–328.

[9] HORNUNG, R. AND MEINHARDT, T. (1987). Quantitative risk assessment of lung cancer in U. S. uranium miners. *Health Physics* **52**, 417–30.

[10] HOSMER, JR., D. W. AND LEMESHOW, S. (1999). *Applied survival analysis.* Wiley Series in Probability and Statistics: Texts and References Section. John Wiley & Sons Inc., New York. MR1674644

[11] KALBFLEISCH, J. D. AND PRENTICE, R. L. (2002). *The statistical analysis of failure time data.* Wiley Series in Probability and Statistics. Wiley-Interscience [John Wiley & Sons], Hoboken, NJ. MR1924807

[12] KLEIN, J. P. AND MOESCHBERGER, M. L. (2003). *Survival Analysis. Techniques for Censored and Truncated Data. 2nd ed..* Springer Verlag,

New York.

[13] LANGHOLZ, B. AND BORGAN, Ø. (1995). Counter-matching: A stratified nested case-control sampling method. *Biometrika* **82**, 69–79.

[14] LANGHOLZ, B. AND GOLDSTEIN, L. (1996). Risk set sampling in epidemiologic cohort studies. *Statist. Sci.* **11**, 35–53.

[15] LANGHOLZ, B., THOMAS, D. C., XIANG, A., AND STRAM, D. (1999). Latency analysis in epidemiologic studies of occupational exposures: Application to the Colorado Plateau uranium miners cohort. *Amer. J. Indust. Med.* **35**, 246–256.

[16] LIN, D. Y., WEI, L. J., AND YING, Z. (1993). Checking the Cox model with cumulative sums of martingale-based residuals. *Biometrika* **80**, 557–572. MR1248021

[17] LUBIN, J., BOICE, J., EDLING, C., HORNUNG, R., HOWE, G., KUNZ, E., KUSIAK, R., MORRISON, H., RADFORD, E., SAMET. J., TIRMARCHE, M., WOODWARD. A., XIANG, Y., AND PIERCE, D. (1994). Radon and lung cancer risk: A joint analysis of 11 underground miners studies. NIH Publication 94-3644. U.S. Department of Health and Human Services, Public Health Service, National Institutes of Health. Bethesda, MD.

[18] MAY, S. AND HOSMER, D. W. (2004). A cautionary note on the use of the Grønnesby and Borgan goodness-of-fit test for the Cox proportional hazards model. *Lifetime Data Anal.* **10**, 283–291. MR2086961

[19] OAKES, D. (1981). Survival times: aspects of partial likelihood. *Internat. Statist. Rev.* **49**, 235–264. MR651473

[20] PRENTICE, R. L. AND SELF, S. G. (1983). Asymptotic distribution theory for Cox-type regression models with general relative risk form. *Ann. Statist.* **11**, 804–813. MR707931

[21] SCHOENFELD, D. (1980). Chi-squared goodness-of-fit tests for the proportional hazards regression model. *Biometrika* **67**, 145–153. MR570516

[22] THERNEAU, T. M. AND GRAMBSCH, P. M. (2000). *Modeling survival data: extending the Cox model.* Statistics for Biology and Health. Springer-Verlag, New York. MR1774977

[23] THERNEAU, T. M., GRAMBSCH, P. M., AND FLEMING, T. R. (1990). Martingale-based residuals for survival models. *Biometrika* **77**, 147–160. MR1049416

[24] 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.