

# Estimation of covariate-dependent Markov transition probabilities from nested case-control data

Ørnulf Borgan

Department of Mathematics, University of Oslo

P.O. Box 1053 Blindern, N-0316 Oslo, Norway

e-mail: borgan@math.uio.no

## Abstract

Multi-state models are used to describe situations where individuals may move among a finite number of states defined by specific conditions of health, including death. The transition intensities of the models are described by proportional hazards models, and it is reviewed how estimation of the regression parameters and the baseline transition intensities may be performed when only nested case-control data are available for all or some of the transitions. The regression parameter estimates and the estimates of baseline transition intensities are combined to give estimates of the integrated transition intensities for specified covariate histories, and from these estimates covariate-dependent Markov transition probabilities are derived.

*Key words:* Aalen-Johansen estimator, absolute risk, competing risks, Markov chains, multi-state models, multivariate survival analysis, nested case-control studies, risk set sampling.

## 1 Introduction

In most clinical and epidemiological studies, the focus is on one endpoint, often death due to a certain disease. There is, however, an increasing interest in studying more than one endpoint using multi-state models; see, e.g., the review by Andersen and Keiding (2002). A common situation of this form is when two or more competing causes of death are considered. Another situation is when an individual may move from a healthy state to a diseased state and then to dead. Quite generally, an individual may move among a finite, usually small, number of states defined by specific conditions of health, including death. Usually each individual will have a set of covariates measured, and one may then want to estimate the probabilities of transitions between the states and how these transition probabilities depend on the covariates.

There are two steps in this problem. The first is to estimate how the covariates influence the transition intensities between the various states. This may be done by fitting a Cox regression model, or another proportional hazards model, for each of the possible transitions. The second step is to combine the estimated transition intensities for a given set of covariates into appropriate transition probabilities. When the transitions between the states can be modeled by a Markov chain, this can be done using the empirical transition matrix for time-continuous non-homogeneous Markov chains of Aalen and Johansen

(1978). For cohort data, this program for estimation of covariate-dependent Markov transition probabilities has been carried through by Andersen *et al.* (1991), see also Andersen *et al.* (1993, Section VII.2.3).

Estimation of the transition intensities using proportional hazards models is based on a partial likelihood which, at each observed transition between two states of health, compares the covariate values of the individual making the transition to the covariate values of all individuals who are at risk of making the transition at the time when the transition occurs. In large epidemiologic cohort studies, use of proportional hazards models requires the investigator to collect complete covariate information on all individuals in the cohort, even when only a small fraction of these actually experiences (some of) the transitions that are being studied. This may be very expensive, or even logistically impossible. Nested case-control sampling, where the relevant covariate information is collected for all individuals making a transition (“cases”), but only for a sample of those who are not (“controls”), then offers a useful alternative which may drastically reduce the resources that need to be allocated to a study.

Important progress has been achieved during the last decade both with respect to the theoretical foundation of the nested case-control design, and in the development of new methodology of practical importance. The key to this progress has been to use counting processes to model jointly the occurrence of events and the sampling of controls (Borgan *et al.* 1995). This counting process formulation not only allows for a theoretical study of the nested case-control design paralleling the one for the full cohort of Andersen and Gill (1982). It also solves the problem of how to estimate, from nested case-control data, the baseline hazard rate (Borgan and Langholz 1993) and the absolute risk of death from a disease in the presence of competing causes of death (Langholz and Borgan 1997). The purpose of the present paper is to extend the results of Langholz and Borgan (1997) to any time-continuous non-homogeneous Markov chain with a finite state space, thus providing a method for estimation of covariate-dependent Markov transition probabilities from nested case-control data.

To motivate the general situation, we in Section 2 consider in detail the competing risks model with deaths from a disease  $D$  and from other causes. We assume that only nested case-control data are available for deaths from disease  $D$ , while the cause specific hazard of deaths due to other causes does not depend on covariates. The results are illustrated by a study of absolute risk of lung cancer death among a cohort of uranium miners. The material in Subsections 2.1-2.4 closely follows Langholz and Borgan (1997), but in our presentation we use notation which is adapted to our more general needs. In Subsection 2.5 we show how the covariate-dependent transition probabilities for the competing risks model may be expressed as a matrix product. This provides a stepping stone to Section 3, where the situation with a general multi-state model is considered. In Subsections 3.1-3.3 we assume, for the ease of presentation, that only nested case-control data are available for all the transitions between the states, and that the covariates influence the intensities of all the possible transitions. The modifications needed when we have full cohort data for one or more of the transitions (or even all of them), or when some transitions do not depend on covariates, are discussed in Subsection 3.4. Some concluding comments are given in Section 4.

The estimators of the covariate-dependent Markov transition probabilities presented in Section 3 are of the Aalen-Johansen type, and their properties may be derived by com-

binning the arguments of Andersen *et al.* (1993, Section VII.2.3) on estimation of covariate-dependent transition probabilities from cohort data with those of Borgan *et al.* (1995) on regression parameter and baseline hazard estimation for nested case-control data. In order to keep the presentation as non-technical as possible, no proofs are given in the main body of the paper. An outline of the proofs for all the results given in the paper is provided in the Appendix.

## 2 Two competing risks

We first consider estimation of the absolute risk of death from a specific disease  $D$  taking into account the competing risk of death from other causes. This is a situation of considerable interest in its own right. Furthermore it makes it possible to introduce many of the ideas of Section 3 without worrying about the full generality of the situation considered there. In order to keep the presentation simple, we assume in this section that the cause specific hazard of deaths due to other causes does not depend on covariates. The necessary modifications when this is the case are deferred to Section 3. The notation of the present section is somewhat more involved than would have been needed if our only interest had been in the situation with two competing causes of death. The notation is chosen to make the relation to the general situation as transparent as possible.

### 2.1 The model

We model the situation with two competing causes of death (the disease  $D$  and other causes) by a process with one transient state 1, corresponding to “alive,” and two absorbing states 2 and 3 corresponding to “dead from disease  $D$ ” and “dead from other causes,” respectively, as illustrated in Figure 1.

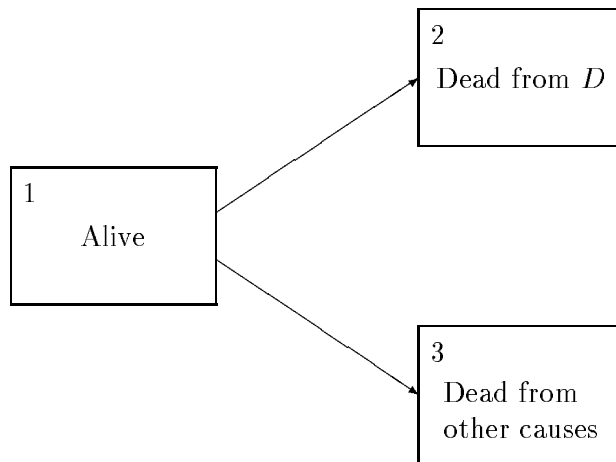


Figure 1: A model for two competing causes of death.

The cause specific hazard of death due to disease  $D$  (a  $1 \rightarrow 2$  transition) for an individual with vector of covariates  $\mathbf{z}(t)$  is denoted  $\alpha_{12}(t; \mathbf{z})$ . It describes the instantaneous risk dying from disease  $D$  for this individual. The time-variable  $t$  may be age, time since employment, or some other time-scale relevant to the problem at hand. The covariates may be time-fixed or time-dependent.

We assume that  $\alpha_{12}(t; \mathbf{z})$  is related to the vector of covariates  $\mathbf{z}(t)$  by the proportional hazards model

$$\alpha_{12}(t; \mathbf{z}) = \alpha_{12}(t) r_{12}(\boldsymbol{\beta}, \mathbf{z}(t)). \quad (1)$$

Here  $\alpha_{12}(t)$  is a baseline cause-specific hazard for deaths due to disease  $D$ ,  $r_{12}(\boldsymbol{\beta}, \mathbf{z}(t))$  is the relative risk for an individual with vector of covariates  $\mathbf{z}(t)$ , and  $\boldsymbol{\beta}$  is a vector of regression coefficients describing the effects of the covariates. We assume that the relative risk function is standardized such that  $r_{12}(\boldsymbol{\beta}, \mathbf{0}) = 1$ . For the special choice  $r_{12}(\boldsymbol{\beta}, \mathbf{z}(t)) = \exp\{\mathbf{z}(t)' \boldsymbol{\beta}\}$ , formula (1) gives the usual Cox regression model for deaths due to the disease  $D$ . Other possibilities include the linear relative risk function  $r_{12}(\boldsymbol{\beta}, \mathbf{z}(t)) = 1 + \boldsymbol{\beta}' \mathbf{z}(t)$  and the excess relative risk model considered in Subsection 2.4.

The cause specific hazard of deaths due to other causes (a  $1 \rightarrow 3$  transition) is, in this section, assumed not to depend on the covariates, and it is denoted  $\alpha_{13}(t)$ .

If all the time-dependent covariates in (1) are external, we may consider the model we get by conditioning on a specified (i.e. deterministic) covariate history  $\mathbf{z}_0(t)$  (Kalbfleisch and Prentice 1980, Section 5.3). The transition intensities  $\alpha_{12}(t; \mathbf{z}_0)$  and  $\alpha_{13}(t)$  for this (conditional) model, define a Markov chain with state space  $\{1, 2, 3\}$ . For this Markov chain, we write  $P_{hj}(s, t; \mathbf{z}_0)$  for the probability that an individual who is in state  $h$  at time  $s$  will be in state  $j$  at a later time  $t$ ;  $h, j = 1, 2, 3$ . Then  $P_{12}(s, t; \mathbf{z}_0)$  is the *absolute risk* of death from disease  $D$  between time  $s$  and  $t$  for an individual with covariate history  $\mathbf{z}_0(t)$ , while  $P_{13}(s, t; \mathbf{z}_0)$  is the absolute risk of death from other causes. Note also that  $P_{11}(s, t; \mathbf{z}_0)$  is the probability of surviving at least to time  $t$  for an individual with covariate history  $\mathbf{z}_0(t)$  who is alive at time  $s$ . The transition probabilities may be expressed by the transition intensities as follows:

$$P_{11}(s, t; \mathbf{z}_0) = \exp \left\{ - \int_s^t [\alpha_{12}(u; \mathbf{z}_0) + \alpha_{13}(u)] du \right\}, \quad (2)$$

$$P_{12}(s, t; \mathbf{z}_0) = \int_s^t P_{11}(s, u; \mathbf{z}_0) \alpha_{12}(u; \mathbf{z}_0) du, \quad (3)$$

$$P_{13}(s, t; \mathbf{z}_0) = \int_s^t P_{11}(s, u; \mathbf{z}_0) \alpha_{13}(u) du. \quad (4)$$

## 2.2 Estimation of regression parameters and baseline hazards

We consider a cohort of  $n$  individuals. The individuals may be followed over different periods of time, so our observations of their survival times may be subject to left-truncation and/or right censoring. A crucial assumption, however, is that truncation and censoring is independent so that the entry and censoring times do not carry any information on the risk of death from the two causes; see Andersen *et al.* (1993, Sections III.2-3) for a general discussion. We assume that there are no tied deaths (a few ties may be broken at random). The times when deaths (of any cause) are observed are denoted  $t_1 < t_2 < \dots$ , and we let  $i_k$  be the index of the individual dying at  $t_k$ . Furthermore, we let  $d_{12k} = 1$  if a

death due to disease  $D$ , i.e., a  $1 \rightarrow 2$  transition, is observed at  $t_k$ ;  $d_{12k} = 0$  otherwise; and we define  $d_{13k}$  similarly for deaths from other causes. The number of individuals at risk (i.e. in state 1) at time  $t_k$  is assumed known, and is denoted  $n_1(t_k)$ .

In order to estimate the transition probabilities (2)-(4) for an individual with a specified covariate history  $\mathbf{z}_0(t)$ , we first need to estimate the regression parameter  $\beta$  and the baseline cause specific hazard  $\alpha_{12}(t)$  in (1), and the cause specific hazard  $\alpha_{13}(t)$  of death due to other causes.

Consider first deaths due to other causes. Here we use the Nelson-Aalen estimator to estimate the integrated cause specific hazard  $A_{13}(t) = \int_0^t \alpha_{13}(u)du$ . For our purpose it is convenient to write this as

$$\hat{A}_{13}(t) = \sum_{t_k \leq t} \hat{\alpha}_{13}(t_k) \quad (5)$$

where  $\hat{\alpha}_{13}(t_k) = d_{13k}/n_1(t_k)$ ; see, e.g., Borgan (1998b) for a review of the Nelson-Aalen estimator.

We then consider deaths due to the disease  $D$ . Here we do not have complete covariate information. Rather we assume that only nested case-control data are available (Thomas 1977). Thus, at each  $t_k$  with  $d_{12k} = 1$ , a random sample of controls of size  $m - 1$  is selected without replacement from those at risk, i.e., from the individuals who are in state 1 “just before” time  $t_k$ . We let  $\tilde{\mathcal{R}}_k$  denote the “sampled risk set” at  $t_k$  consisting of the case  $i_k$  together with its sampled controls. Covariate information is then collected for the individuals in the sampled risk sets.

Then the regression parameter  $\beta$  in (1) is estimated by  $\hat{\beta}$ , the value of  $\beta$  which maximizes the partial likelihood

$$\mathcal{L}_{12}(\beta) = \prod \left\{ \frac{r_{12}(\beta, z_{i_k}(t_k))}{\sum_{l \in \tilde{\mathcal{R}}_k} r_{12}(\beta, \mathbf{z}_l(t_k))} \right\} \quad (6)$$

(Oakes 1981). The product in (6) extends over all  $t_k$  with  $d_{12k} = 1$ , and  $\mathbf{z}_l(t_k)$  is the covariate vector for individual  $l$  at time  $t_k$ . The integrated baseline cause specific hazard  $A_{12}(t) = \int_0^t \alpha_{12}(u)du$  is estimated by the Breslow-type estimator

$$\hat{A}_{12}(t) = \sum_{t_k \leq t} \frac{d_{12k}}{\sum_{l \in \tilde{\mathcal{R}}_k} r_{12}(\hat{\beta}, \mathbf{z}_l(t_k)) w_{12l}(t_k)} \quad (7)$$

where the weights  $w_{12l}(t_k) = n_1(t_k)/m$  are inversely proportional to the fraction of those at risk who are selected to the sampled risk set  $\tilde{\mathcal{R}}_k$  (Borgan and Langholz 1993).

### 2.3 Estimation for a specified covariate history

For a specified covariate history  $\mathbf{z}_0(t)$ , the integrated cause specific hazard of deaths due to disease  $D$  is  $A_{12}(t; \mathbf{z}_0) = \int_0^t \alpha_{12}(u; \mathbf{z}_0)du$ . This may be estimated by

$$\hat{A}_{12}(t; \mathbf{z}_0) = \sum_{t_k \leq t} \hat{\alpha}_{12}(t_k; \mathbf{z}_0), \quad (8)$$

where

$$\hat{\alpha}_{12}(t_k; \mathbf{z}_0) = \frac{r_{12}(\hat{\beta}, \mathbf{z}_0(t_k))}{\sum_{l \in \tilde{\mathcal{R}}_k} r_{12}(\hat{\beta}, \mathbf{z}_l(t_k)) w_{12l}(t_k)} d_{12k} \quad (9)$$

is equal to  $r_{12}(\hat{\beta}, \mathbf{z}_0(t_k))$  times the increment at  $t_k$  of the Breslow-type estimator (7). The estimator (8) is approximately normally distributed with a variance which may be estimated as described in Subsection 3.3.

We may now use the estimators of the (integrated) cause specific hazards to obtain estimators for the covariate-dependent transition probabilities (2)-(4). The survival probability (2) is estimated by the Kaplan-Meier type estimator

$$\hat{P}_{11}(s, t; \mathbf{z}_0) = \prod_{s < t_k \leq t} \{1 - [\hat{\alpha}_{12}(t_k; \mathbf{z}_0) + \hat{\alpha}_{13}(t_k)]\}, \quad (10)$$

while the cause specific absolute risks of death (3) and (4) are estimated by

$$\hat{P}_{12}(s, t; \mathbf{z}_0) = \sum_{s < t_k \leq t} \hat{P}_{11}(s, t_{k-1}; \mathbf{z}_0) \hat{\alpha}_{12}(t_k; \mathbf{z}_0), \quad (11)$$

$$\hat{P}_{13}(s, t; \mathbf{z}_0) = \sum_{s < t_k \leq t} \hat{P}_{11}(s, t_{k-1}; \mathbf{z}_0) \hat{\alpha}_{13}(t_k). \quad (12)$$

The estimator for the survival probability makes sense provided that all the factors are positive in the product on the right hand side of (10). This is the case when  $\hat{\alpha}_{12}(t_k; \mathbf{z}_0) + \hat{\alpha}_{13}(t_k) < 1$  for all  $t_k$  between  $s$  and  $t$ . Due to the weights  $w_{12l}(t_k) = n_1(t_k)/m$  in the denominator of (9), this will almost always be the case, unless one tries to estimate the survival probability for (typically large) values of  $t$  where there are few individuals at risk.

Note that the estimators (10)-(12) are of the same form as the transition probabilities (2)-(4), but with the exponential function in (2) replaced by the product in (10), and the integrals in (3) and (4) replaced by the sums in (11) and (12). The estimators (10)-(12) are special cases of the general estimator derived in Section 3. There we will discuss how to estimate the variances of the covariate-dependent transition probabilities; cf. the formulas in the last paragraph of Subsection 3.3.

## 2.4 Lung cancer deaths for uranium miners

We illustrate the methods of Subsections 2.1-2.3 using a nested case-control sample from a cohort of uranium miners from the Colorado Plateau. This cohort was assembled to study the effects of radon exposure and smoking on mortality and has been described in detail in earlier publications; see e.g. Hornung and Meinhardt (1987). We will focus on lung cancer mortality ( $D$ ) taking into account death from other causes.

The cohort consists of 3,347 Caucasian male miners recruited between 1950 and 1960. The cohort was traced for mortality outcomes through December 31, 1982, by which time 258 lung cancer deaths and 2087 deaths from other cause were observed. Although covariate information is available on all cohort subjects, in order to illustrate the methods, each lung cancer case was matched to a single control sampled randomly from subjects who were on study at the case's age of death (a few tied failure times were randomly broken). 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. In our analyses, we summarize these exposures for the cases and their controls into cumulative radon and cumulative packs of cigarettes lagged by two years, which we denote by  $\mathbf{z}(t) = (R(t), S(t))'$ .

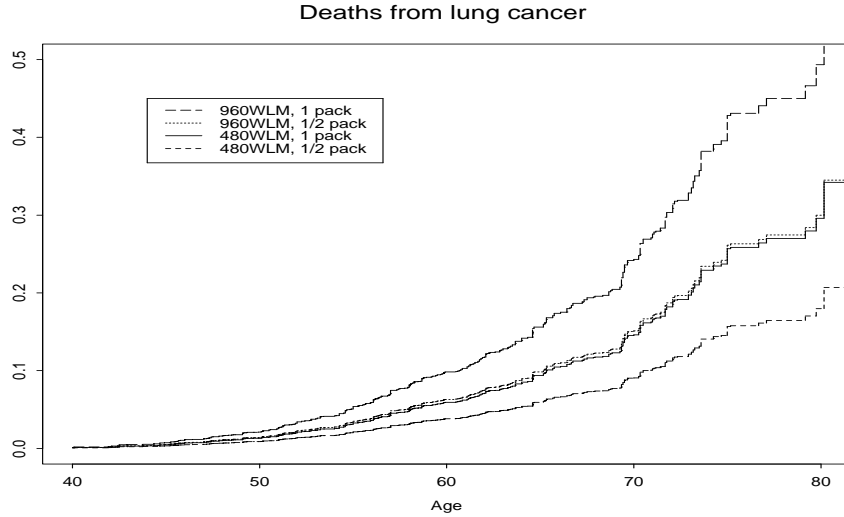


Figure 2: Estimated integrated hazard for lung cancer death for four exposure histories.

As has been the case in previous analyses of these data, we assume that the lung cancer mortality at age  $t$  may be described by the relative excess risk model

$$\alpha_{12}(t; \mathbf{z}) = \alpha_{12}(t)(1 + \beta_R R(t))(1 + \beta_S S(t)) \quad (13)$$

(Langholz and Borgan 1997), while the mortality  $\alpha_{13}(t)$  from other causes, as an approximation, is assumed not to depend on the radon and smoking exposures. The parameter estimates from the nested case-control data were  $\hat{\beta}_R = 0.42$  per 100 WLM (s.e. = 0.20), and  $\hat{\beta}_S = 0.23$  per 1000 packs of cigarettes (s.e. = 0.10), with a correlation between the estimates of 0.15.

We will assume that the cumulative radon and smoking exposures (lagged by two years) are not influenced by the lung cancer deaths, i.e. that they are external time-dependent covariates. It then makes sense consider the absolute risk of lung cancer death conditional on specific (i.e. deterministic) radon and smoking histories; cf. Subsection 2.1. We will consider four such covariate histories. For all the four situations considered, we assume a constant exposure intensity for radon between ages 20 to 50. Thus, the two year lagged cumulative radon exposure  $R(t)$  is zero for  $t < 22$ , then increases linearly up to the total dose of radon at  $t = 52$ , and is constant at the total dose thereafter. Smoking is described by the number of packs per day, and we assume that smoking begin at age 20 and continues throughout life at the same level. The following four covariate histories are considered:

- Total radon dose 960 WLM, smoking 1 pack per day.
- Total radon dose 960 WLM, smoking 1/2 pack per day.
- Total radon dose 480 WLM, smoking 1 pack per day.
- Total radon dose 480 WLM, smoking 1/2 pack per day.

Figure 2 shows the estimated integrated hazards (8) for the four exposure histories. According to these estimates, the effect on the lung cancer death rates of a doubling of the

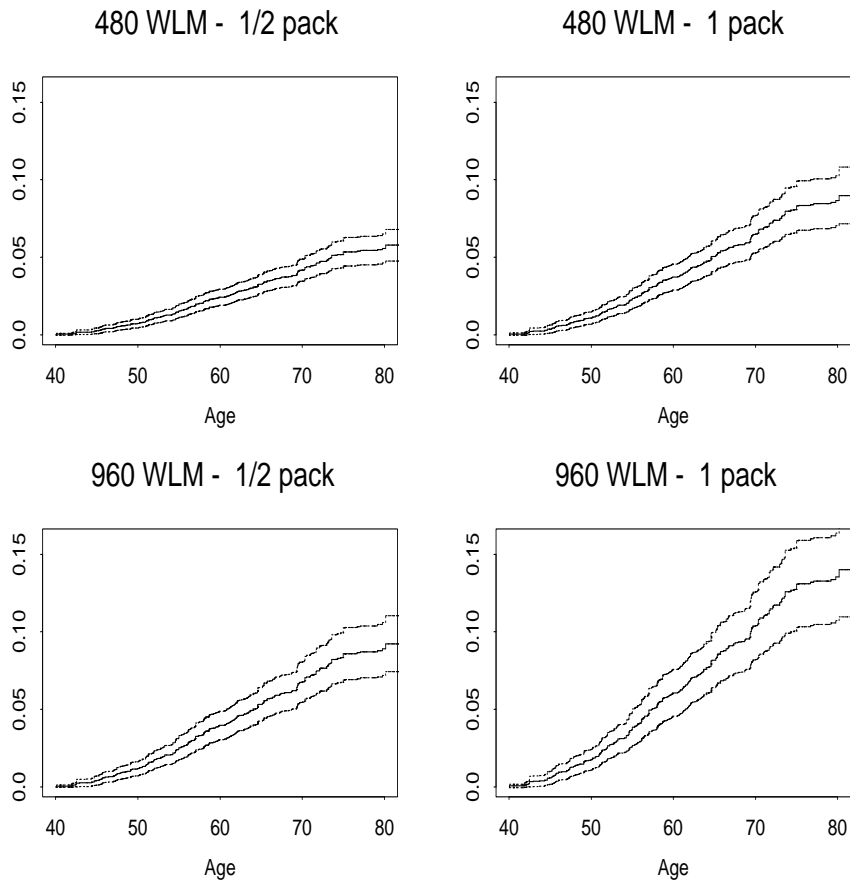


Figure 3: Estimated absolute risk of lung cancer death for four exposure histories.

radon exposure is about the same as the effect of a doubling of the amount of smoking. The estimates (11) of the absolute risk of lung cancer death are shown in Figure 3. In the figure we also give 95% point-wise confidence intervals using the asymptotic normality of (11) and the variance estimate given in Subsection 3.3. The estimated absolute risk of lung cancer death before age 80 years, is about 5% for a miner who smokes 1/2 a pack a day and is exposed to a total dose of radon of 480 WLM between age 20 years and 50 years. If both the smoking and radon exposure is doubled, the estimated absolute risk is about 13%.

## 2.5 The transition probability matrix

It is possible to rewrite the estimators (10)-(12) as a matrix product. This is useful, since the matrix formulation may readily be extended to the general situation considered in Section 3. To see how this can be done, we first introduce the covariate-dependent transition probability matrix  $\mathbf{P}(s, t; \mathbf{z}_0)$  as the  $3 \times 3$  matrix having the transition probability  $P_{hj}(s, t; \mathbf{z}_0)$  as its  $(h, j)$ -th element. More specifically, since the states 2 and 3 are



absorbing:

$$\mathbf{P}(s, t; \mathbf{z}_0) = \begin{pmatrix} P_{11}(s, t; \mathbf{z}_0) & P_{12}(s, t; \mathbf{z}_0) & P_{13}(s, t; \mathbf{z}_0) \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

This transition probability matrix is estimated by the matrix  $\widehat{\mathbf{P}}(s, t; \mathbf{z}_0)$  obtained by replacing the transition probabilities  $P_{1j}(s, t; \mathbf{z}_0)$  in the first row by their estimators (10)-(12).

To see how  $\widehat{\mathbf{P}}(s, t; \mathbf{z}_0)$  can be expressed as a matrix product, we define  $\widehat{\alpha}_{11}(t_k; \mathbf{z}_0) = -[\widehat{\alpha}_{12}(t_k; \mathbf{z}_0) + \widehat{\alpha}_{13}(t_k; \mathbf{z}_0)]$ ; cf. (5) and (9); and introduce the  $3 \times 3$  matrix

$$\widehat{\boldsymbol{\alpha}}(t_k; \mathbf{z}_0) = \begin{pmatrix} \widehat{\alpha}_{11}(t_k; \mathbf{z}_0) & \widehat{\alpha}_{12}(t_k; \mathbf{z}_0) & \widehat{\alpha}_{13}(t_k; \mathbf{z}_0) \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}. \quad (14)$$

It is then a simple exercise in matrix calculus to see that we may write

$$\widehat{\mathbf{P}}(s, t; \mathbf{z}_0) = \prod_{s < t_k \leq t} \{\mathbf{I} + \widehat{\boldsymbol{\alpha}}(t_k; \mathbf{z}_0)\}, \quad (15)$$

where  $\mathbf{I}$  is the  $3 \times 3$  identity matrix, and the matrix product is taken in the order of increasing  $t_k$ s.

### 3 Multi-state models

In this section we extend the results for two competing risks to the situation where an individual may move among a finite number of states defined by specific conditions of health, including death. One such example is when an individual may move from a healthy state to a diseased state and then to death. This situation may be modeled by an illness-death model with the states 1 = “healthy,” 2 = “diseased,” and 3 = “dead,” as illustrated in Figure 4. More involved multi-state models are discussed, e.g., by Andersen *et al.* (1993, Sections I.3.3 and I.3.4) and Andersen and Keiding (2002).

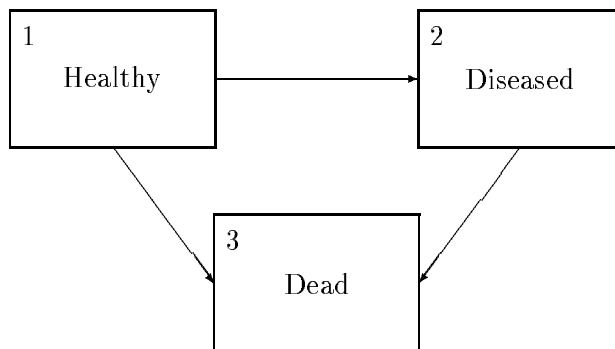


Figure 4: An illness-death model.

### 3.1 The model

We have a cohort with  $n$  individuals, and we want to study how the individuals move between  $K$  states of health. To this end we use a multi-state model with state space  $\{1, 2, \dots, K\}$ , and denote by  $\alpha_{hj}(t; \mathbf{z})$  the transition intensity from state  $h$  to state  $j$ ;  $j \neq h$ ; for an individual with vector of covariates  $\mathbf{z}(t)$ . The transition intensities describe the instantaneous risks of transitions between the states, i.e.,  $\alpha_{hj}(t; \mathbf{z})dt$  is the probability that an individual with vector of covariates  $\mathbf{z}(t)$  who is in state  $h$  just before time  $t$  will make a transition to state  $j$  in the small time interval  $[t, t + dt)$ .

We assume that the transition intensities are related to the vector of covariates  $\mathbf{z}(t)$  by proportional hazards models

$$\alpha_{hj}(t; \mathbf{z}) = \alpha_{hj}(t) r_{hj}(\boldsymbol{\beta}, \mathbf{z}(t)); \quad (16)$$

cf. (1). In connection with (16), it is important to note that separate relative risk functions are used for the different transitions, but that the vector of regression coefficients is the same for all of them. This way of formulating the model covers a number of situations.

Assume, as an illustration, that we for the competing risks model of Figure 1 have two time-fixed covariates,  $\mathbf{z} = (z_1, z_2)'$ , and that these influence the cause specific hazards of deaths from the disease  $D$  as well as from other causes. If we assume Cox regression models with different effect of the covariates for the two causes, this is obtained by specifying the relative risk functions as  $r_{12}(\boldsymbol{\beta}, \mathbf{z}) = \exp(\beta_1 z_1 + \beta_2 z_2)$  and  $r_{13}(\boldsymbol{\beta}, \mathbf{z}) = \exp(\beta_3 z_1 + \beta_4 z_2)$  and the vector of regression coefficients as  $\boldsymbol{\beta} = (\beta_1, \beta_2, \beta_3, \beta_4)'$ . If, e.g., the effect of the covariate  $z_1$  is assumed to be the same for the two causes,  $r_{12}(\boldsymbol{\beta}, \mathbf{z})$  remains the same, while  $r_{13}(\boldsymbol{\beta}, \mathbf{z}) = \exp(\beta_1 z_1 + \beta_3 z_2)$  and  $\boldsymbol{\beta} = (\beta_1, \beta_2, \beta_3)'$ . Note also that the model formulation (16) covers the situation where different relative risk functions are used for the two causes, e.g., we may use an excess relative risk model (cf. Section 2.4) for deaths from the disease  $D$  and a Cox model for deaths from other causes.

If all the time-dependent covariates in (16) are external, we may, as in Subsection 2.1, consider the model we get by conditioning on a specified (i.e. deterministic) covariate history  $\mathbf{z}_0(t)$ . The transition intensities  $\alpha_{hj}(t; \mathbf{z}_0)$  for this (conditional) model define a Markov chain with state space  $\{1, 2, \dots, K\}$ . The transition probabilities  $P_{hj}(s, t; \mathbf{z}_0)$  of this Markov chain give the probabilities that an individual with covariate history  $\mathbf{z}_0(t)$  who is in a state  $h$  at time  $s$  will be in a state  $j$  at a later time  $t$ . The  $K \times K$  matrix with entry  $(h, j)$  equal to  $P_{hj}(s, t; \mathbf{z}_0)$  is denoted  $\mathbf{P}(s, t; \mathbf{z}_0)$ . This is the transition probability matrix of the Markov chain.

Only for simple situations, like the competing risks model of Section 2, is it possible to give explicit expressions for the  $P_{hj}(s, t; \mathbf{z}_0)$  in terms of the transition intensities, cf. (2)-(4). However, for all Markov chains, the transition probability matrix may be expressed as a “continuous matrix product,” or more formally as a *product-integral*, of the matrix of transition intensities; see, e.g., Andersen *et al.* (1993, Sections II.6 and IV.4) for a formal discussion or Borgan (1998a) for a less technical review. This product-integral representation is the key to the study of the estimator  $\hat{\mathbf{P}}(s, t; \mathbf{z}_0)$  of the transition probability matrix described in the next subsection; cf. the Appendix.

### 3.2 Estimation of transition intensities and probabilities

The  $n$  individuals in the cohort may be followed over different periods of time, so our observations of their life-histories may be subject to left-truncation and/or right censoring. It is crucial that truncation and censoring is independent; cf. Subsection 2.2. We assume that exact times for transitions between the states are recorded, and denote by  $t_1 < t_2 < \dots$  the times when transitions between the states are observed. Furthermore, assuming no tied transition times, we let  $i_k$  be the index of the individual who makes a transition at  $t_k$ , and we let  $d_{hjk} = 1$  if a  $h \rightarrow j$  transition is observed at  $t_k$ ;  $d_{hjk} = 0$  otherwise. The number of individuals in state  $h$  at time  $t_k$  is assumed known, and is denoted  $n_h(t_k)$ ;  $h = 1, 2, \dots, K$ .

For all transitions, we assume that only nested case-control data are available (see, however, Subsection 3.4). Thus at each  $t_k$  with  $d_{hjk} = 1$ , a random sample of controls of size  $m - 1$  is selected without replacement from the  $n_h(t_k) - 1$  individuals (the case excluded) who are in state  $h$  “just before” time  $t_k$ . The “sampled risk set”  $\tilde{\mathcal{R}}_k$  at  $t_k$  consists of the individual  $i_k$  making the transition (the case) together with its sampled controls. Covariate information is collected for the cases and their controls, but is not needed for the other individuals in the cohort.

Then the regression parameter  $\beta$  in (16) is estimated by maximizing the partial likelihood

$$\mathcal{L}(\beta) = \prod_{h \neq j} \mathcal{L}_{hj}(\beta), \quad (17)$$

where  $\mathcal{L}_{hj}(\beta)$  is defined as in (6), but with the states  $h, j$  replacing the states 1, 2. A formal derivation of (17) is given in the Appendix. There we also show that  $\hat{\beta}$  is asymptotically multivariate normally distributed around the true value of  $\beta$  with a covariance matrix that may be estimated by  $\mathcal{I}(\hat{\beta})^{-1}$ , where  $\mathcal{I}(\beta) = -(\partial^2/\partial\beta^2) \log \mathcal{L}(\beta)$  is the observed information matrix.

A Breslow type estimator for the integrated baseline intensity for the  $h \rightarrow j$  transition is obtained by replacing the states 1, 2 in (7) by the states  $h, j$ . For a specified covariate history  $\mathbf{z}_0(t)$ , we may therefore estimate the integrated transition intensity  $A_{hj}(t; \mathbf{z}_0) = \int_0^t \alpha_{hj}(u; \mathbf{z}_0) du$  by

$$\hat{A}_{hj}(t; \mathbf{z}_0) = \sum_{t_k \leq t} \hat{\alpha}_{hj}(t_k; \mathbf{z}_0), \quad (18)$$

where

$$\hat{\alpha}_{hj}(t_k; \mathbf{z}_0) = \frac{r_{hj}(\hat{\beta}, \mathbf{z}_0(t_k))}{\sum_{l \in \tilde{\mathcal{R}}_k} r_{hj}(\hat{\beta}, \mathbf{z}_l(t_k)) w_{hjl}(t_k)} d_{hjk} \quad (19)$$

with weights  $w_{hjl}(t_k) = n_h(t_k)/m$ ; cf. (8) and (9). A formal derivation of (18) is given in the Appendix.

An estimator of the transition probability matrix is then  $\hat{\mathbf{P}}(s, t; \mathbf{z}_0)$  given by (15). But in the general situation considered here,  $\mathbf{I}$  on the right hand side of (15) is the  $K \times K$  identity matrix, while  $\hat{\alpha}(t_k; \mathbf{z}_0)$  is the  $K \times K$  matrix with entry  $(h, j)$  equal to  $\hat{\alpha}_{hj}(t_k; \mathbf{z}_0)$  for  $h \neq j$  and entry  $(h, h)$  equal to  $\hat{\alpha}_{hh}(t_k; \mathbf{z}_0) = -\sum_{j \neq h} \hat{\alpha}_{hj}(t_k; \mathbf{z}_0)$ . The estimator makes sense provided that  $\hat{\alpha}_{hh}(t_k; \mathbf{z}_0) > -1$  for all  $t_k$  between  $s$  and  $t$ ; cf. the discussion in Subsection 2.3.

### 3.3 Asymptotic distribution and estimation of variances

In the Appendix we show that  $\widehat{A}_{hj}(t; \mathbf{z}_0)$  given by (18) is asymptotically normally distributed around  $A_{hj}(t; \mathbf{z}_0)$ , with a variance that may be estimated by

$$\widehat{\sigma}_{hj}^2(t; \mathbf{z}_0) = \widehat{\tau}_{hj}^2(t; \mathbf{z}_0) + \widehat{\mathbf{C}}_{hj}(t; \mathbf{z}_0)' \mathcal{I}(\widehat{\boldsymbol{\beta}})^{-1} \widehat{\mathbf{C}}_{hj}(t; \mathbf{z}_0). \quad (20)$$

Here

$$\widehat{\tau}_{hj}^2(t; \mathbf{z}_0) = \sum_{t_k \leq t} \widehat{\alpha}_{hj}(t_k; \mathbf{z}_0)^2, \quad (21)$$

where  $\widehat{\alpha}_{hj}(t_k; \mathbf{z}_0)$  is given by (19), and

$$\widehat{\mathbf{C}}_{hj}(t; \mathbf{z}_0) = \sum_{t_k \leq t} \widehat{\mathbf{c}}_{hj}(t_k; \mathbf{z}_0), \quad (22)$$

where

$$\widehat{\mathbf{c}}_{hj}(t_k; \mathbf{z}_0) = \left\{ \frac{\dot{\mathbf{r}}_{hj}(\widehat{\boldsymbol{\beta}}; \mathbf{z}_0(t_k))}{\sum_{l \in \widetilde{\mathcal{R}}_k} r_{hj}(\widehat{\boldsymbol{\beta}}; \mathbf{z}_l(t_k)) w_{hjl}(t_k)} - \frac{r_{hj}(\widehat{\boldsymbol{\beta}}; \mathbf{z}_0(t_k)) \sum_{l \in \widetilde{\mathcal{R}}_k} \dot{\mathbf{r}}_{hj}(\widehat{\boldsymbol{\beta}}; \mathbf{z}_l(t_k)) w_{hjl}(t_k)}{\left\{ \sum_{l \in \widetilde{\mathcal{R}}_k} r_{hj}(\widehat{\boldsymbol{\beta}}; \mathbf{z}_l(t_k)) w_{hjl}(t_k) \right\}^2} \right\} d_{hjk} \quad (23)$$

with  $\dot{\mathbf{r}}_{hj}(\boldsymbol{\beta}; \mathbf{z}(t)) = \frac{\partial}{\partial \boldsymbol{\beta}} r_{hj}(\boldsymbol{\beta}; \mathbf{z}(t))$ .

In the Appendix we also show that the estimated transition probability matrix  $\widehat{\mathbf{P}}(s, t; \mathbf{z}_0)$  is asymptotically multivariate normally distributed. In particular we have, for given  $h, j = 1, 2, \dots, K$ , that the estimated covariate-dependent transition probability  $\widehat{P}_{hj}(s, t; \mathbf{z}_0)$  is asymptotically normally distributed around  $P_{hj}(s, t; \mathbf{z}_0)$  with a variance that may be estimated by

$$\widehat{\text{Var}} \left( \widehat{P}_{hj}(s, t; \mathbf{z}_0) \right) = \widehat{\text{Var}}_1 \left( \widehat{P}_{hj}(s, t; \mathbf{z}_0) \right) + \widehat{\text{Var}}_2 \left( \widehat{P}_{hj}(s, t; \mathbf{z}_0) \right),$$

where

$$\widehat{\text{Var}}_2 \left( \widehat{P}_{hj}(s, t; \mathbf{z}_0) \right) = \sum_{g=1}^K \sum_{l \neq g} \sum_{s < t_k \leq t} \left[ \widehat{P}_{hg}(s, t_k; \mathbf{z}_0) \{ \widehat{P}_{lj}(t_k, t; \mathbf{z}_0) - \widehat{P}_{gj}(t_k, t; \mathbf{z}_0) \} \widehat{\alpha}_{gl}(t_k; \mathbf{z}_0) \right]^2, \quad (24)$$

and

$$\widehat{\text{Var}}_1 \left( \widehat{P}_{hj}(s, t; \mathbf{z}_0) \right) = \widehat{\mathbf{G}}_{hj}(s, t; \mathbf{z}_0)' \mathcal{I}(\widehat{\boldsymbol{\beta}})^{-1} \widehat{\mathbf{G}}_{hj}(s, t; \mathbf{z}_0). \quad (25)$$

Here

$$\widehat{\mathbf{G}}_{hj}(s, t; \mathbf{z}_0) = \sum_{g=1}^K \sum_{l=1}^K \sum_{s < t_k \leq t} \widehat{P}_{hg}(s, t_k; \mathbf{z}_0) \widehat{P}_{lj}(t_k, t; \mathbf{z}_0) \widehat{\mathbf{c}}_{gl}(t_k; \mathbf{z}_0) \quad (26)$$

with  $\widehat{\mathbf{c}}_{gl}(t_k; \mathbf{z}_0)$  given by (23) for  $g \neq l$ , and  $\widehat{\mathbf{c}}_{gg}(t_k; \mathbf{z}_0) = -\sum_{l \neq g} \widehat{\mathbf{c}}_{gl}(t_k; \mathbf{z}_0)$ .

For a specific model, the formulas (24) and (26) may often be rewritten in a less compact form. In particular, for the estimated probability of a  $1 \rightarrow 2$  transition in the model of Figure 1 with two competing causes of death, they become

$$\widehat{\text{Var}}_2\left(\widehat{P}_{12}(s, t; \mathbf{z}_0)\right) = \sum_{s < t_k \leq t} \left[ \widehat{P}_{11}(s, t_k; \mathbf{z}_0) \{1 - \widehat{P}_{12}(t_k, t; \mathbf{z}_0)\} \widehat{\alpha}_{12}(t_k; \mathbf{z}_0) \right]^2 + \sum_{s < t_k \leq t} \left[ \widehat{P}_{11}(s, t_k; \mathbf{z}_0) \widehat{P}_{12}(t_k, t; \mathbf{z}_0) \widehat{\alpha}_{13}(t_k; \mathbf{z}_0) \right]^2,$$

and

$$\widehat{\mathbf{G}}_{12}(s, t; \mathbf{z}_0) = \sum_{s < t_k \leq t} \widehat{P}_{11}(s, t_k; \mathbf{z}_0) \left[ 1 - \widehat{P}_{12}(t_k, t; \mathbf{z}_0) \right] \widehat{\mathbf{c}}_{12}(t_k; \mathbf{z}_0).$$

### 3.4 Modifications

In Subsections 3.1-3.3, we have assumed that only nested case-control data are available for all the transitions between the states, and that the covariates influence the risk of all the possible transitions.

If we have full cohort information for one or more of the transitions (or even all of them), the results of Subsections 3.2 and 3.3 continue to hold with only minor modifications. The only formulas which need to be modified, are (19) and (23). Assume, e.g., that we have full cohort information for the  $m \rightarrow r$  transition. We then, for this transition, replace the sampled risk set  $\widetilde{\mathcal{R}}_k$  by the set of everyone at risk in state  $m$  “just before” time  $t_k$ , and use the weights  $w_{mrl}(t_k) = 1$  in (19) and (23).

If one or more of the transitions do not depend on covariates (as is the case for death from other causes in Sections 2), this is covered by our results by letting the relative risk function in (16) be identically equal to one for such transitions. As a consequence, the quantities in (23), and therefore also in (22), (25), and (26), vanish for such transitions.

## 4 Discussion

In Section 3 we have, for the ease of presentation, assumed that we select the same number of controls for all transitions. If the number of controls for a  $h \rightarrow j$  transition is  $m_{hj} - 1$ , all the results of Section 3 continue to hold if we use the weights  $w_{hjl}(t_k) = n_h(t_k)/m_{hj}$  in (19) and (23). In the Appendix we show that the results also remain valid if the controls are selected using the counter-matched design of Langholz and Borgan (1995), or one of the other sampling designs described by Borgan *et al.* (1995, Section 5), provided that the appropriate weights for the designs are used in (19) and (23), and also the partial likelihood [cf. (6) and (17)] is modified by the introduction of these weights in the numerator and the denominator.

In the formulation of the proportional hazards models (16), we have assumed separate baseline intensities for all the possible transitions between the states. In some situations it may be useful to assume that the intensities for two or more transitions are proportional. This is not immediately covered by the results we present in Section 3. It is possible to modify the partial likelihood and the estimators of the covariate-dependent transition intensities and probabilities in order to accommodate this situation as well, but it is outside the scope of this paper to go into details about how this can be done.

We have assumed that the effect of covariates can be described by the proportional hazards models (16). Another option is to use Aalen’s additive model. Aalen *et al.* (2001) show

how estimates of integrated transition intensities from Aalen's model, possibly based on nested case-control data, may be used to obtain estimates of covariate-dependent Markov transition probabilities. For Aalen's model one estimate (integrated) regression *functions* for each of the covariates, rather than regression parameters as in the proportional hazards models (16). An advantage of this is that the effects of the covariates are allowed to change over time. The cost is increased variability in the estimates. In particular for nested case-control data, a higher number of controls per case is needed for Aalen's model than when proportional hazards models are used (Zhang and Borgan 1999).

An alternative to the nested case-control design, is to use the case-cohort design (Prentice 1986). For this design one, at the start of the study, selects at random a subcohort from the full cohort. The individuals in the subcohort who are in a given state  $h$  at time  $t$ , are then used as controls if a transition out of state  $h$  is observed at time  $t$ . When integrated transition intensities are estimated from case-cohort data, one may still use a matrix of the form (15) to estimate covariate-dependent Markov-transition probabilities. However, as estimation is no longer based on a partial likelihood, the large sample properties of these estimators will become more involved. The individuals in the subcohort may be used as controls for all transitions for which they are at risk. Therefore it is often claimed that the case-cohort design is to be preferred to the nested case-control design for situations with multiple endpoints as considered in the present paper. However, the reuse of the control individuals in the subcohort will create a correlation in the estimates of the transition intensities, and this makes it hard to decide without further studies which of the two designs have the highest statistical efficiency for estimating covariate-dependent Markov-transition probabilities.

## References

- Aalen, O. O., Borgan, Ø. and Fekjær, H. (2001). Covariate adjustment of event histories estimated from Markov chains: the additive approach. *Biometrics* **57**, 993–1001.
- Aalen, O. O. and Johansen, S. (1978). An empirical transition matrix for nonhomogeneous Markov chains based on censored observations. *Scandinavian Journal of Statistics* **5**, 141–150.
- Andersen, P. K., Borgan, Ø., Gill, R. D. and Keiding, N. (1993). *Statistical Models Based on Counting Processes*. Springer-Verlag, New York.
- Andersen, P. K. and Gill, R. D. (1982). Cox's regression model for counting processes: A large sample study. *Annals of Statistics* **10**, 1100–1120.
- Andersen, P. K., Hansen, L. S. and Keiding, N. (1991). Non- and semi-parametric estimation of transition probabilities from censored observations of a non-homogeneous Markov process. *Scandinavian Journal of Statistics* **18**, 153–167.
- Andersen, P. K. and Keiding, N. (2002). Multi-state models for event history analysis. *Statistical Methods in Medical Research*, to appear.
- Borgan Ø. (1998a). Aalen-Johansen estimator. In: P. Armitage and T. Colton (eds.) *Encyclopedia of Biostatistics* **1**, 5–10. Wiley, Chichester.
- Borgan Ø. (1998b). Nelson-Aalen estimator In: P. Armitage and T. Colton (eds.) *Encyclopedia of Biostatistics* **4**, 2967–2972. Wiley, Chichester.
- Borgan, Ø., Goldstein L., and Langholz (1995). Methods for the analysis of sampled cohort data in the Cox proportional hazards model. *Annals of Statistics* **23**, 1749–1778.

- Borgan, Ø. and Langholz, B. (1993). Non-parametric estimation of relative mortality from nested case-control studies. *Biometrics* **49**, 593-602.
- Hornung, R. and Meinhardt, T. (1987). Quantitative risk assessment of lung cancer in U. S. uranium miners. *Health Physics*, **52**, 417-30.
- Kalbfleisch, J. D. and Prentice, R. L. (1980). *The Statistical Analysis of Failure Time Data*. Wiley, New York.
- Langholz, B. and Borgan, Ø. (1995). Counter-matching: A stratified nested case-control sampling method. *Biometrika* **82**, 69-79.
- Langholz, B. and Borgan, Ø. (1997). Estimation of absolute risk from nested case-control data, *Biometrics* **53**, 768-775.
- Oakes, D. (1981). Survival times: Aspects of partial likelihood (with discussion). *International Statistical Review* **49**, 235-264.
- Prentice, R. L. (1986). A case-cohort design for epidemiologic cohort studies and disease prevention trials. *Biometrika* **73**, 1-11.
- 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. *Journal of the Royal Statistical Society. Ser. A* **140**, 469-491.
- Zhang, J. and Borgan, Ø. (1999). Aalen's linear model for sampled risk set data: a large sample study. *Lifetime Data Analysis* **5**, 351-369.

## Appendix

The results summarized in the main body of the paper may be derived by combining the arguments of Borgan *et al.* (1995) for regression parameter and baseline hazard estimation from nested case-control data with those of Andersen *et al.* (1993, Sections IV.4 and VII.2.3) for estimation of Markov transition probabilities from cohort data. In this appendix we outline the main steps in these derivations, tacitly assuming that the necessary regularity conditions hold.

To begin with we assume that only nested case-control data are available for all the possible transitions of the multi-state model. At the end we comment on which modifications are needed when other sampling plans are used for selecting the controls, or when we have complete data for some of the transitions.

The model and the data are described in Subsections 3.1 and 3.2. Furthermore, we let  $Y_{hi}(t) = 1$  if individual  $i$  is in state  $h$  "just before" time  $t$ , and  $Y_{hi}(t) = 0$  otherwise;  $i = 1, 2, \dots, n$ . Then  $n_h(t) = \sum_{i=1}^n Y_{hi}(t)$  is the number of individuals in state  $h$  at  $t-$ . If individual  $i$  makes a  $h \rightarrow j$  transition at time  $t$ , we select by simple random sampling  $m - 1$  controls among the  $n_h(t) - 1$  individuals in state  $h$  (individual  $i$  excluded). The probability that the sampled risk set becomes the set  $\mathbf{r}$  is then

$$\pi_t(\mathbf{r}|h, j, i) = \binom{n_h(t) - 1}{m - 1}^{-1} = \binom{n_h(t)}{m}^{-1} w_{hji}(t), \quad (\text{A.1})$$

with  $w_{hji}(t) = n_h(t)/m$ , provided that  $\mathbf{r}$  is a set of size  $m$  which contains  $i$  and is a subset of the set of individuals in  $h$  at  $t-$ .

We introduce the processes  $N_{(h,j,i,\mathbf{r})}(t)$  counting the number of times individual  $i$  makes a  $h \rightarrow j$  transition *and* the sampled risk set equals  $\mathbf{r}$  in  $[0, t]$ . Since we assume no tied transitions, the collection of all these counting processes is a multivariate counting process. The intensity process of  $N_{(h,j,i,\mathbf{r})}(t)$  takes the form

$$\lambda_{(h,j,i,\mathbf{r})}(t) = Y_{hi}(t)\alpha_{hj}(t)r_{hj}(\boldsymbol{\beta}_0, \mathbf{z}_i(t))\pi_t(\mathbf{r}|h, j, i),$$

where we use  $\beta_0$  to denote the true value of  $\beta$ .

Arguing as in Borgan *et al.* (1995, pp. 1753-54), and using (A.1), it is seen that the partial likelihood for  $\beta$  takes the form

$$\mathcal{L}(\beta) = \prod_{h \neq j} \prod_t \prod_{\mathbf{r} \in \mathcal{P}} \prod_{i \in \mathbf{r}} \left\{ \frac{Y_{hi}(t) r_{hj}(\beta, \mathbf{z}_i(t)) w_{hji}(t)}{S_{h_j \mathbf{r}}^{(0)}(\beta, t)} \right\}^{\Delta N_{(h,j,i,\mathbf{r})}(t)}, \quad (\text{A.2})$$

where  $\mathcal{P}$  is the set of all possible sampled risk sets, and

$$S_{h_j \mathbf{r}}^{(0)}(\beta, t) = \sum_{l \in \mathbf{r}} Y_{hl}(t) r_{hj}(\beta, \mathbf{z}_l(t)) w_{hjl}(t). \quad (\text{A.3})$$

For simple random sampling of the controls, the weights  $w_{hjl}(t) = n_h(t)/m$  do not depend on the individual  $l$ , and therefore cancel in the numerator and denominator of (A.2). The result is the partial likelihood (17). By an argument along the lines of Borgan *et al.* (1995, pp. 1754-56 and Theorems 1 and 2) one may show that  $\sqrt{n}(\hat{\beta} - \beta_0)$  converges weakly, as  $n \rightarrow \infty$ , to a mean zero multivariate normal distribution. The covariance matrix of this distribution may be estimated consistently by  $n\mathcal{I}(\hat{\beta})^{-1}$ , where  $\mathcal{I}(\beta) = -(\partial^2/\partial\beta^2) \log \mathcal{L}(\beta)$  is the observed information matrix.

Let  $N_{h_j \mathbf{r}}(t) = \sum_{i \in \mathbf{r}} N_{(h,j,i,\mathbf{r})}(t)$  count the total number of  $h \rightarrow j$  transitions having sampled risk set equal to  $\mathbf{r}$  in  $[0, t]$ . A similar argument as the one in Borgan *et al.* (1995, pp. 1756-57) shows that the integrated baseline transition intensities  $A_{hj}(t)$ ;  $h \neq j$ ; may be estimated by  $\hat{A}_{hj}(t) = \hat{A}_{hj}(t, \hat{\beta})$ , where

$$\hat{A}_{hj}(t, \beta) = \int_0^t \sum_{\mathbf{r} \in \mathcal{P}} \frac{dN_{h_j \mathbf{r}}(u)}{S_{h_j \mathbf{r}}^{(0)}(\beta, u)}. \quad (\text{A.4})$$

For a specified covariate history  $\mathbf{z}_0(t)$ , we may then estimate the integrated transition intensities  $A_{hj}(t; \mathbf{z}_0)$  by

$$\hat{A}_{hj}(t; \mathbf{z}_0) = \int_0^t r_{hj}(\hat{\beta}, \mathbf{z}_0(u)) d\hat{A}_{hj}(u) = \int_0^t \sum_{\mathbf{r} \in \mathcal{P}} \frac{r_{hj}(\hat{\beta}, \mathbf{z}_0(u))}{S_{h_j \mathbf{r}}^{(0)}(\hat{\beta}, u)} dN_{h_j \mathbf{r}}(u). \quad (\text{A.5})$$

Note that the estimator (18) is of this form.

To derive the asymptotic distribution of the estimators (A.5), we introduce, for  $h \neq j$ , the vectors

$$\hat{\mathbf{B}}_{hj}(t, \beta) = -\frac{\partial}{\partial \beta} \hat{A}_{hj}(t, \beta) = \int_0^t \sum_{\mathbf{r} \in \mathcal{P}} \frac{\mathbf{S}_{h_j \mathbf{r}}^{(1)}(\beta, u)}{S_{h_j \mathbf{r}}^{(0)}(\beta, u)^2} dN_{h_j \mathbf{r}}(u), \quad (\text{A.6})$$

where

$$\mathbf{S}_{h_j \mathbf{r}}^{(1)}(\beta, t) = \frac{\partial}{\partial \beta} S_{h_j \mathbf{r}}^{(0)}(\beta, t) = \sum_{l \in \mathbf{r}} Y_{hl}(t) \dot{\mathbf{r}}_{hj}(\beta, \mathbf{z}_l(t)) w_{hjl}(t) \quad (\text{A.7})$$

with  $\dot{\mathbf{r}}_{hj}(\beta, \mathbf{z}_l(t)) = \frac{\partial}{\partial \beta} r_{hj}(\beta, \mathbf{z}_l(t))$ . Let  $\mathbf{B}_{hj}(t, \beta)$  be the uniform limit in probability, as  $n \rightarrow \infty$ , of (A.6). Arguing as in Andersen *et al.* (1993, Theorem VII.2.3) and Borgan *et al.* (1995, Theorem 3), we may then show that  $\sqrt{n}(\hat{\beta} - \beta_0)$  and the processes

$$U_{hj}(\cdot) = \sqrt{n} \left( \hat{A}_{hj}(\cdot) - A_{hj}(\cdot) \right) + \sqrt{n}(\hat{\beta} - \beta_0)' \mathbf{B}_{hj}(\cdot, \beta_0) \quad h \neq j$$

are asymptotically independent, and that the limiting distribution of  $U_{hj}$  is that of a zero mean Gaussian martingale with a variance function which may be estimated uniformly consistently, as  $n \rightarrow \infty$ , by  $n$  times

$$\hat{\omega}_{hj}^2(t) = \int_0^t \sum_{\mathbf{r} \in \mathcal{P}} \frac{dN_{h_j \mathbf{r}}(u)}{S_{h_j \mathbf{r}}^{(0)}(\hat{\beta}, u)^2}. \quad (\text{A.8})$$



From this it follows by a Taylor expansion as in Andersen *et al.* (1993, pp. 505-506 and 513-514) that, for  $h \neq j$ , the processes

$$\sqrt{n} \left( \widehat{A}_{hj}(\cdot; \mathbf{z}_0) - A_{hj}(\cdot; \mathbf{z}_0) \right)$$

are asymptotically equivalent to

$$X_{hj}^{(1)}(\cdot) + X_{hj}^{(2)}(\cdot).$$

Here

$$X_{hj}^{(1)}(t) = \sqrt{n}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)' \mathbf{C}_{hj}(t; \mathbf{z}_0)$$

with

$$\mathbf{C}_{hj}(t; \mathbf{z}_0) = \int_0^t \dot{\mathbf{r}}_{hj}(\boldsymbol{\beta}_0; \mathbf{z}_0(u)) \alpha_{hj}(u) du - \int_0^t r_{hj}(\boldsymbol{\beta}_0; \mathbf{z}_0(u)) d\mathbf{B}_{hj}(u, \boldsymbol{\beta}_0),$$

while

$$X_{hj}^{(2)}(t) = \int_0^t r_{hj}(\boldsymbol{\beta}_0; \mathbf{z}_0(u)) dU_{hj}(u).$$

The processes  $X_{hj}^{(1)}(\cdot)$  and  $X_{hj}^{(2)}(\cdot)$  are asymptotically independent.

In particular, for a fixed value of  $t$ , we have that  $\sqrt{n}(\widehat{A}_{hj}(t; \mathbf{z}_0) - A_{hj}(t; \mathbf{z}_0))$  is asymptotically normally distributed with mean zero and with a variance that may be estimated consistently by  $n$  times

$$\widehat{\sigma}_{hj}^2(t; \mathbf{z}_0) = \widehat{\tau}_{hj}^2(t; \mathbf{z}_0) + \widehat{\mathbf{C}}_{hj}(t; \mathbf{z}_0)' \mathcal{I}(\widehat{\boldsymbol{\beta}})^{-1} \widehat{\mathbf{C}}_{hj}(t; \mathbf{z}_0). \quad (\text{A.9})$$

Here [cf. (A.8)]

$$\widehat{\tau}_{hj}^2(t; \mathbf{z}_0) = \int_0^t r_{hj}(\widehat{\boldsymbol{\beta}}; \mathbf{z}_0(u))^2 d\widehat{\omega}_{hj}^2(u) = \int_0^t \sum_{\mathbf{r} \in \mathcal{P}} \left( \frac{r_{hj}(\widehat{\boldsymbol{\beta}}; \mathbf{z}_0(u))}{S_{hj\mathbf{r}}^{(0)}(\widehat{\boldsymbol{\beta}}, u)} \right)^2 dN_{hj\mathbf{r}}(u), \quad (\text{A.10})$$

while [cf. (A.4) and (A.6)]

$$\begin{aligned} \widehat{\mathbf{C}}_{hj}(t; \mathbf{z}_0) &= \int_0^t \dot{\mathbf{r}}_{hj}(\widehat{\boldsymbol{\beta}}; \mathbf{z}_0(u)) d\widehat{A}_{hj}(u) - \int_0^t r_{hj}(\widehat{\boldsymbol{\beta}}; \mathbf{z}_0(u)) d\widehat{\mathbf{B}}_{hj}(u, \widehat{\boldsymbol{\beta}}) \\ &= \int_0^t \sum_{\mathbf{r} \in \mathcal{P}} \left\{ \frac{\dot{\mathbf{r}}_{hj}(\widehat{\boldsymbol{\beta}}; \mathbf{z}_0(u))}{S_{hj\mathbf{r}}^{(0)}(\widehat{\boldsymbol{\beta}}, u)} - \frac{r_{hj}(\widehat{\boldsymbol{\beta}}; \mathbf{z}_0(u)) \mathbf{S}_{hj\mathbf{r}}^{(1)}(\widehat{\boldsymbol{\beta}}, u)}{S_{hj\mathbf{r}}^{(0)}(\widehat{\boldsymbol{\beta}}, u)^2} \right\} dN_{hj\mathbf{r}}(u). \end{aligned} \quad (\text{A.11})$$

The leading term on the right hand side of (A.9) is the contribution to the variance from  $X_{hj}^{(2)}(t)$ , while the last term is the contribution from  $X_{hj}^{(1)}(t)$ . Note that (A.10) and (A.11) equal, respectively, (21) and (22) in Subsection 3.3.

Let  $\mathbf{A}(t; \mathbf{z}_0)$  be the  $K \times K$  matrix with entry  $(h, j)$  equal to  $A_{hj}(t; \mathbf{z}_0)$  for  $h \neq j$  and entry  $(h, h)$  equal to  $A_{hh}(t; \mathbf{z}_0) = -\sum_{j \neq h} A_{hj}(t; \mathbf{z}_0)$ . Then the covariate-dependent transition probability matrix may be given as the product-integral  $\mathbf{P}(s, t; \mathbf{z}_0) = \mathcal{P}_{(s,t]} \{ \mathbf{I} + d\mathbf{A}(u, \mathbf{z}_0) \}$  (Andersen *et al.* 1993, Sections II.6 and IV.4). It is estimated by

$$\widehat{\mathbf{P}}(s, t; \mathbf{z}_0) = \mathcal{P}_{(s,t]} \left\{ \mathbf{I} + d\widehat{\mathbf{A}}(u, \mathbf{z}_0) \right\}, \quad (\text{A.12})$$

where  $\widehat{\mathbf{A}}(t; \mathbf{z}_0)$  is the  $K \times K$  matrix of estimated integrated transition intensities defined in the same way as  $\mathbf{A}(t; \mathbf{z}_0)$ . Note that (A.12) is the estimator given in the last paragraph of Subsection 3.2.

By the functional delta method, it follows as in Andersen *et al.* (1993, pp. 513-14) that the matrix-valued process

$$\sqrt{n} \left( \widehat{\mathbf{P}}(s, \cdot; \mathbf{z}_0) - \mathbf{P}(s, \cdot; \mathbf{z}_0) \right)$$

is asymptotically equivalent to the process equal in the point  $t$  to

$$\int_s^t \mathbf{P}(s, u; \mathbf{z}_0) d\mathbf{X}^{(1)}(u) \mathbf{P}(u, t; \mathbf{z}_0) + \int_s^t \mathbf{P}(s, u; \mathbf{z}_0) d\mathbf{X}^{(2)}(u) \mathbf{P}(u, t; \mathbf{z}_0).$$

Here  $\mathbf{X}^{(1)}(\cdot)$  and  $\mathbf{X}^{(2)}(\cdot)$  are asymptotically independent  $K \times K$  matrix-valued processes defined from the processes  $X_{hj}^{(1)}(\cdot)$  and  $X_{hj}^{(2)}(\cdot)$ , respectively, in the same way as described for  $\mathbf{A}(t; \mathbf{z}_0)$  above. Therefore, for fixed values of  $s$  and  $t$  with  $s < t$ , we have that  $\sqrt{n}(\widehat{\mathbf{P}}(s, t; \mathbf{z}_0) - \mathbf{P}(s, t; \mathbf{z}_0))$  is multivariate normally distributed with mean zero. As in Andersen *et al.* (1993, pp. 291-294 and 514-15) it follows that the elements of the  $K^2 \times K^2$  covariance matrix of this multivariate normal distribution may be estimated by  $n$  times

$$\widehat{\text{Cov}} \left( \widehat{P}_{hj}(s, t; \mathbf{z}_0), \widehat{P}_{mr}(s, t; \mathbf{z}_0) \right) = \widehat{\text{Cov}}_1 \left( \widehat{P}_{hj}(s, t; \mathbf{z}_0), \widehat{P}_{mr}(s, t; \mathbf{z}_0) \right) + \widehat{\text{Cov}}_2 \left( \widehat{P}_{hj}(s, t; \mathbf{z}_0), \widehat{P}_{mr}(s, t; \mathbf{z}_0) \right),$$

where

$$\widehat{\text{Cov}}_2 \left( \widehat{P}_{hj}(s, t; \mathbf{z}_0), \widehat{P}_{mr}(s, t; \mathbf{z}_0) \right) = \tag{A.13}$$

$$\sum_{g \neq l} \int_s^t \widehat{P}_{hg}(s, u; \mathbf{z}_0) \widehat{P}_{mg}(s, u; \mathbf{z}_0) \{ \widehat{P}_{lj}(u, t; \mathbf{z}_0) - \widehat{P}_{gj}(u, t; \mathbf{z}_0) \} \{ \widehat{P}_{lr}(u, t; \mathbf{z}_0) - \widehat{P}_{gr}(u, t; \mathbf{z}_0) \} d\widehat{\tau}_{gl}^2(u; \mathbf{z}_0)$$

and

$$\widehat{\text{Cov}}_1 \left( \widehat{P}_{hj}(s, t; \mathbf{z}_0), \widehat{P}_{mr}(s, t; \mathbf{z}_0) \right) = \widehat{\mathbf{G}}_{hj}(s, t; \mathbf{z}_0)' \mathcal{I}(\widehat{\boldsymbol{\beta}})^{-1} \widehat{\mathbf{G}}_{mr}(s, t; \mathbf{z}_0), \tag{A.14}$$

where

$$\widehat{\mathbf{G}}_{hj}(s, t; \mathbf{z}_0) = \sum_{g,l} \int_s^t \widehat{P}_{hg}(s, u; \mathbf{z}_0) \widehat{P}_{lj}(u, t; \mathbf{z}_0) d\widehat{\mathbf{C}}_{gl}(u; \mathbf{z}_0) \tag{A.15}$$

with the vector  $\widehat{\mathbf{C}}_{gl}(u; \mathbf{z}_0)$  given by (A.11) for  $g \neq l$ , and  $\widehat{\mathbf{C}}_{gg}(u; \mathbf{z}_0) = -\sum_{l \neq g} \widehat{\mathbf{C}}_{gl}(u; \mathbf{z}_0)$ . When  $(h, j) = (m, r)$ , (A.13)-(A.15) equal, respectively, (24)-(26) in Subsection 3.3.

Above we have assumed that only nested case-control data are available for all the possible transitions between the states. This assumption is taken care of by the specification of the sampling distribution  $\pi_t(\mathbf{r}|h, j, i)$ , which gives the probability of selecting  $\mathbf{r}$  as the sampled risk set if individual  $i$  makes a  $h \rightarrow j$  transition at time  $t$ . The specific form (A.1) of the sampling distribution is used to get the weights  $w_{hjl}(t) = n_h(t)/m$  in (A.2), (A.3), and (A.7), but does not otherwise influence the presentation in this appendix. Therefore all the results of this appendix remain valid for the general sampling schemes of Borgan *et al.* (1995) provided that the weights  $w_{hjl}(t)$  in (A.2), (A.3), and (A.7), are redefined as in formula (4.2) of that paper. The situation where we have full cohort information for one or more transitions, is also covered by the general set-up. For this corresponds to selecting the entire risk set with probability one, giving the weights  $w_{hjl}(t) = 1$ ; cf. Borgan *et al.* (1995, pp. 1757-58).