Advances in frailty models

Balan, T.A.

Chapter 4

Ascertainment correction in frailty models for recurrent events data

Abstract

In retrospective studies involving recurrent events, it is common to select individuals based on their event history up to the time of selection. In this case, the ascertained subjects might not be representative for the target population, and the analysis should take the selection mechanism into account. The purpose of this chapter is two-fold. First, to study what happens when the data analysis is not adjusted for the selection, and second, to propose a corrected analysis. Under the Andersen-Gill and shared frailty regression models, we show that the estimators of covariate effects, incidence and frailty variance can be biased if the ascertainment is ignored, and we show that with a simple adjustment of the likelihood, unbiased and consistent estimators are obtained. The proposed method is assessed by a simulation study and is illustrated on a data set comprising recurrent pneumothoraces.

4.1 Introduction

In the study of recurrent events it is of interest to model the rate at which the events occur in time, along with estimating the effects of different factors which may influence

This chapter has been published as: T.A. Balan, M.A. Jonker, P.C. Johannesma, H. Putter (2016). Ascertainment correction in frailty models for recurrent events data. *Statistics in Medicine* 35(23), 4183-4201.

this rate, such as treatments or individual covariates. Usually, it is assumed that the process which generates the recurrent events starts at a time 0; this can be, for example, the diagnosis of a certain disease, a medical intervention, or birth, and that it is ended by some form of right-censoring. Ultimately, the research aims to extrapolate the conclusions from a sample of individuals to a larger "population at risk". The sample selection process is critical for the validity and interpretation of the results.

In a prospective study a random sample is drawn from the target population at time 0, and then followed up for the occurrence of events, whereas in a retrospective study design, the sample is selected at a time later than 0, with the data up to the time of selection being collected on the ascertained individuals. Prospective studies are desirable although they may require a long time to be conducted. Their main advantage is that all aspects of the data collection are under the control of the researcher. Retrospective studies are usually observational in nature. While cheaper and shorter than prospective studies, they are associated with less control on the sample selection process. Ideally, the sampling mechanism should lead to a sample that can be viewed as a random representation of the full population of interest at time 0.

When the sampling happens at a time point after 0, the probability for a subject to be included in the study may depend on the subject's event history. For example, registries are often kept only for patients who experienced some recurrent events, not on the whole population that is at risk to experience these events. Such a sample can not be regarded as representative for the target population. The necessity to adjust the analysis to take the selection mechanism into account has been underlined in the context of recurrent events in Cook and Lawless (2007, ch. 7.3), although most approaches for this problem are ad-hoc in nature.

In the motivating example of this chapter, only subjects who experienced at least one occurrence of the event of interest between 1990 and 2014 were registered. Hence, subjects who only experience events before 1990 or after 2014 are not included in the study. As a consequence, the individuals who have a higher rate of recurrent events are over-represented in the sample. If not adjusted for, the ascertainment can bias the estimation of model parameters in the statistical analysis of such data. Selection bias is a known problem in epidemiology (Hernán, Hernández-Díaz, and Robins, 2004). Several paradoxical results in studies involving recurrent events might be explained by a closely related "index bias" (Dahabreh and Kent, 2011).

The effects of the selection scheme are more difficult to disentangle when random effects are used to model additional heterogeneity or correlation structures present in the data. The frailty model (Vaupel, Manton, and Stallard, 1979) is commonly used for recurrent events or clustered failures data (Hougaard, 2000; Cook and Lawless, 2007). When the selection of individuals depends on the previous history of events, it might also depends on the value of the random effect, further complicating the estimation of frailty models.

Most of the literature on event-based selection in this context has focused on models for the waiting times (gaps) between the events. For example, Scheike, Petersen, and

Martinussen (1999) use an adjusted frailty model to analyze time to pregnancy. In their case, the selection scheme reflects itself in truncation of the observed gap times. Their approach can be used when the selection of individuals is directly related to the length of the waiting times, rather than a specified calendar time interval. It is worth mentioning that a closely related problem is that of frailty models for clustered survival data in the presence of left truncation. Jensen et al. (2004) proposed a corrected likelihood for family data when observations are collected only on the failure has not occured before some time. Several papers followed up on this work (Van den Berg and Drepper, 2011; Eriksson, Martinussen, and Scheike, 2015). Sun and Li (2004) used a frailty model to analyze clustered survival data, where random effects are used to describe a familial structure. In their work, a family is ascertained when at least two members have experience the failure before a certain age. They also provide an "ascertainment-adjused" likelihood, with the focus lying on estimating parameters which describe the latent structure. The selection scheme in the present motivating example is similar. However, we focus on quantities which are of more interest in the recurrent event context, such as covariate effects or the intensity of the recurrent event processes.

We show that the selection based on event history may lead to a sample which is not representative for the initial population at risk, even in very simple ascertainment scenarios, and that this may lead to biased estimates when not properly accounted for in the analysis. The novelty of this chapter lies in the fact that we analyze the effects of ignoring the selection process in the context of recurrent events, along with comparing the adjusted and unadjusted estimators. In Section 4.2, we review the Andersen-Gill and the shared frailty models, we discuss the general idea of constructing a likelihood for models which take the selection mechanism into account, and we propose estimation procedures for both parametric and semiparametric models. The proposed methods are evaluated through a simulation study in Section 4.3, where we investigate properties of the estimators of the baseline intensities, regression coefficients and frailty variances under several scenarios. Finally, we illustrate the considerations of this chapter on a data set on recurrent pneumothoraces in Section 4.4, and we lay out our concluding remarks in Section 4.5.

4.2 Methods

This section is outlined as follows: in 4.2.1 we review the Andersen-Gill and the shared frailty models, and in 4.2.2 we adapt these models to take the selection mechanism into account. In 4.2.3 we discuss the estimation of the proposed adjusted models for parametric specification and we introduce a novel approach for their semiparametric estimation.

4.2.1 Statistical models

The canonical framework for recurrent events is that of counting processes, and particularly that of Poisson processes (Cook and Lawless, 2007, ch. 2). The history of events

of an individual i is "counted" by a stochastic process $N_i(t)$ for $t \geq 0$, with an intensity function $\lambda_i(t)$.

In the Andersen-Gill (AG) model, N_i is assumed to follow the specifications of a non-homogeneous Poisson process with intensity

$$\lambda_i(t;\, \beta, \phi) = \lambda_0(t; \phi) \exp\left(\beta' \mathbf{x}_i(t)\right) \tag{4.1}$$

where $\mathbf{x}_i$ is a vector of possibly time dependent covariates, β is a vector of regression coefficients and ϕ is a vector of parameters which characterize the "baseline" intensity λ_0. In the shared frailty model, N_i is assumed to follow the specifications of a non-homogeneous Poisson process conditional on the unobserved "frailty" $Z_i = z_i$:

$$\lambda_i(t|z_i; \beta, \phi) = z_i \lambda_0(t; \phi) \exp\left(\beta' \mathbf{x}_i(t)\right) \tag{4.2}$$

where it is assumed that $Z_i > 0$ and that the Z_i's are i.i.d. distributed with some density f_θ. In both cases we assume that the censoring is independent given the covariates, and for the shared frailty model we also assume that it is non-informative for the frailty.

Although there is a wide variety of distributions that can be used for Z_i, the most common are the gamma distribution and the log-normal distribution. In the rest of this chapter, we will consider f_θ as the gamma density with expectation 1 and variance θ,

$$f_\theta(z) = \frac{1/\theta^{1/\theta}}{\Gamma(1/\theta)} z^{1/\theta - 1} \exp(-z/\theta), \tag{4.3}$$

with $\theta > 0$ and for $z > 0$. This choice is particularly convenient because the marginal features can be obtained in closed form; see Nielsen et al. (1992) and Murphy (1995a). The AG model can be seen as a limiting case of the shared frailty model when $\theta \longrightarrow 0$; indeed, it can be seen that in this case all z_i's are equal to 1 and (4.2) simplifies to (4.1).

For a subject i, let n_i be the number of observed events. We denote the follow-up time as t_i, and the observed recurrent event times as t_{ij} with $j \in 1...n_i$. Traditionally, the observed data of a certain individual i is denoted as O_i and it represents the probabilistic event "n_i events at $t_{i1} < ... < t_{in_i}$ over the observation time $(0, t_i)$". In the absence of any event-dependent sampling, the construction of likelihoods based on counting processes is detailed in Kalbfleisch and Prentice (2002, ch. 6). In the AG model, from λ_i defined as in (4.1), $P(O_i)$ can be written as

$$P(O_i; \beta, \phi) = \prod_{j=1}^{n_i} \lambda_i(t_{ij}; \beta, \phi) \exp\{-\Lambda_i(t_i; \beta, \phi)\} \tag{4.4}$$

where $\Lambda_i(t_i; \beta, \phi)$ is the cumulative intensity, i.e. $\Lambda_i(t_i; \beta, \phi) = \int_0^{t_i} \lambda_i(s; \beta, \phi) ds$. This leads to the log-likelihood

$$\ell^O(\beta, \phi) = \sum_{i=1}^{n} \sum_{j=1}^{n_i} \{\log \lambda_0(t_{ij}; \phi) + \beta' \mathbf{x}_i(t_{ij})\} - \sum_{i=1}^{n} \int_0^{t_i} \exp(\beta' \mathbf{x}_i(s)) \lambda_0(s) ds. \tag{4.5}$$

In the shared frailty model, a similar expression as (4.4) is obtained conditional on the frailty $Z_i = z_i$, by using the conditional intensity (4.2). The unconditional marginal probability is obtained by taking the expectation over the random effects:

$$\begin{aligned} P(O_i; \beta, \phi, \theta) &= \mathrm{E}_{Z_i} P(O_i | Z_i; \beta, \phi) \\ &= \int_0^\infty \prod_{j=1}^{n_i} \lambda_i(t_{ij}|z_i; \beta, \phi) \exp\{-\Lambda_i(t_i|z_i; \beta, \phi)\} f_\theta(z_i) dz_i. \end{aligned} \tag{4.6}$$

If the frailty follows the gamma distribution with density (4.3), this leads to the log-likelihood

$$\begin{aligned} \ell^O(\beta, \phi, \theta) = &\sum_{i=1}^{n} \sum_{j=1}^{n_i} \left\{ \log \lambda_0(t_{ij}; \phi) + \beta' \mathbf{x}_i(t_{ij}) \right\} + \\ &+ \sum_{i=1}^{n} \left[-(1/\theta + N_{i.}) \log \left\{ 1/\theta + \int_0^{t_i} \exp(\beta' \mathbf{x}_i(s)) \lambda_0(s) ds \right\} + g_i(\theta) \right], \end{aligned} \tag{4.7}$$

where $g_i(\theta) = 1/\theta \log(1/\theta) + \log \Gamma(1/\theta + N_{i.}) - \log \Gamma(1/\theta)$ and $N_{i.}$ represents the total number of events observed for subject i; see Nielsen et al. (1992) for a rigorous and more detailed derivation of this expression.

4.2.2 Ascertainment adjustment

Ascertainment schemes The specification of A_i depends on the design of the study. We introduce three examples to provide the intuition behind this concept.

1. (Left truncation) At the time of the selection, data for subject i is available only if no event occurrences were observed until the age t_{Ri}, In this case, A_i is the probabilistic event "no events occurred between $t_{Li} = 0$ and t_{Ri}", and $P(A_i) = P(N_i(t_{Ri}) = 0)$.

2. In the case of recurrent events, registry data is available on subjects who experienced at least one occurrence in the last k years before the sampling time. Denote the age of individual i at selection as t_i. In this case, A_i is the probabilistic event "at least one event occurred in (t_{Li}, t_{Ri})" where $t_{Li} = t_i - k$ and $t_{Ri} = t_i$, and $P(A_i) = P(N_i(t_i) - N_i(t_i - k) > 0)$.

3. A population is at risk for recurrent events, although only a fraction actually experience an occurrence during follow-up. After the first event, the subjects enter a database where all subsequent recurrences are collected. If data is collected retrospectively from this database, at a time point where subject i's age is t_i, then A_i is the probabilistic event "at least one event in (t_{Li}, t_{Ri})" with $t_{Li} = 0$ and $t_{Ri} = t_i$, and $P(A_i) = P(N_i(t_i) > 0)$.

As in the motivating example, it is more often the case that, in studies involving recurrent events, subjects must experience at least one event during a certain time period, which we refer to as "one-in" ascertainment. This is illustrated by scenarios 2 and 3 above. We define the ascertainment interval (t_{Li}, t_{Ri}) with $0 \leq t_{Li} < t_{Ri} \leq t_i$. Left truncation is a particular selection scenario that, for clarity purposes, we will not develop further, and focus instead on the "one-in" ascertainment.

The adjusted likelihood For now, denote the parameters of the chosen model (AG or shared frailty) as η. We define the event A_i as the ascertainment event, the sampling of subject i from the "population at-risk". The case of interest is when A_i depends on the event history of subject i, and implicitly on the intensity of the counting process N_i. In this case, A_i is a more general event than O_i, since an individual needs to be ascertained in order for O_i to be observed, therefore $O_i \subset A_i$. This implies that $A_i \cap O_i = O_i$ and the likelihood contribution of subject i is given by

$$P(O_i|A_i;\eta) = \frac{P(O_i \cap A_i;\eta)}{P(A_i;\eta)} = \frac{P(O_i;\eta)}{P(A_i;\eta)}. \tag{4.8}$$

Heuristically, the meaning of (4.8) is that each contribution is weighted so that subjects with a low chance of being ascertained (small $P(A_i)$) receive more weight, as they are representative for a part of the population of interest which is under-represented in the ascertained sample.

We define the (ascertainment) adjusted likelihood as the product over the individual contributions (4.8). The adjusted log-likelihood for n individuals is given by

$$\ell(\eta) = \sum_{i=1}^{n} \log P(O_i;\eta) - \sum_{i=1}^{n} \log P(A_i;\eta). \tag{4.9}$$

We will refer to $\ell^O = \sum_{i=1}^n \log P(O_i;\eta)$ as the unadjusted log-likelihood. For the AG and shared frailty mode, this is given by (4.5) and (4.7). We denote the remaining part, $\ell^A = \sum_{i=1}^n \log P(A_i;\eta)$, as the ascertainment adjustment, so that $\ell(\eta) = \ell^O(\eta) - \ell^A(\eta)$.

One-in ascertainment: Andersen-Gill We define

$$\Lambda_{Ai}(\beta,\phi) = \int_{t_{Li}}^{t_{Ri}} \lambda_i(s;\beta,\phi)ds \tag{4.10}$$

with λ_i as specified in (4.1). The probability of no events in (t_{Li}, t_{Ri}) is $P(A_i;\beta,\phi) = \exp(-\Lambda_{Ai}(\beta,\phi))$. Therefore, when subjects are ascertained only when they experience at least one event in (t_{Li}, t_{Ri}),

$$P(A_i;\ \beta,\phi) = 1 - \exp\left(-\Lambda_{Ai}(\beta,\phi)\right).$$

This yields the adjusted log-likelihood

$$\ell(\beta,\phi) = \ell^O(\beta,\phi) - \sum_{i=1}^{n} \log\{1 - \exp\left(-\Lambda_{Ai}(\beta,\phi)\right)\} \tag{4.11}$$

with ℓ^O as defined in (4.5). As the window of observation, or more precisely Λ_{Ai} becomes smaller, the ascertainment correction becomes larger in magnitude. The score functions provide insight into the effect of the ascertainment adjustment.

Denote $S_i^{(1)}(s,t) = \int_s^t \mathbf{x}_i(u)\exp(\beta'\mathbf{x}_i(u))\lambda_0(u)du$. The derivatives of (4.11) with respect to the components of β are

$$U_\beta(\beta,\phi) = \sum_{i=1}^{n}\sum_{j=1}^{n_i}\mathbf{x}_i(t_{ij}) - \sum_i S_i^{(1)}(0,t_i) - \sum_{i=1}^{n}\frac{\exp\left(-\Lambda_{Ai}(\beta,\phi)\right)}{1-\exp\left(-\Lambda_{Ai}(\beta,\phi)\right)}S_i(t_{Li},t_{Ri})$$

where $S_i(s,t) = \int_s^t \mathbf{x}_i(u)\exp(\beta'\mathbf{x}_i(u))\lambda_0(u)du$. The last term in this expression arises as from the ascertainment adjustment, and omitting it would lead to a biased estimate of β. In principle, similar considerations apply also for ϕ, the parameters which describe the baseline intensity. For example, if we consider the Breslow estimator for λ_0, i.e. ϕ is a vector of elements $\phi_k = \lambda_0(s_k)$ where s_k is a time point at which an event is observed, the score vector for ϕ is composed of elements

$$U_{\phi_k}(\beta,\phi) = \frac{N_k}{\phi_k} - \sum_{i=1}^{n} Y_i(s_k)\exp(\beta'\mathbf{x}_i(s_k)) - \\ - \sum_{i=1}^{n} Y_i^A(s_k)\exp(\beta'\mathbf{x}_i(s_k))\frac{\exp\left(-\Lambda_{Ai}(\beta,\phi)\right)}{1-\exp\left(-\Lambda_{Ai}(\beta,\phi)\right)} \tag{4.12}$$

where $Y_i(t)$ is an indicator function which is 1 as long as subject i is at risk at t and 0 otherwise, and $Y_i^A(t)$ is an indicator function which is 1 as long as $t \in (t_{Li}, t_{Ri})$. The second sum term in (4.12) appears due to the ascertainment correction and omitting it would lead to a biased estimate of ϕ.

One-in ascertainment: shared frailty Here we use the same definition for Λ_{Ai} as in (4.10) which can be interpreted as the integrated intensity over the ascertainment interval with the frailty fixed to 1. Conditional on the frailty, the probability of no events in (t_{Li}, t_{Ri}) is $\exp\left(-z_i\Lambda_{Ai}(\beta,\phi)\right)$. The unconditional probability is obtained by integrating over the random effect,

$$E_{Z_i}\left\{\exp\left(-Z_i\Lambda_{Ai}(\beta,\phi)\right)\right\} = \int_0^\infty \exp\left(-z_i\Lambda_{Ai}(\beta,\phi)\right) f_\theta(z_i)dz_i \\ = \frac{1/\theta^{1/\theta}}{\left(1/\theta+\Lambda_{Ai}(\beta,\phi)\right)^{1/\theta}}.$$

Therefore, when subjects are ascertained only when they experience at least one event in (t_{Li}, t_{Ri}),

$$P(A_i;\beta,\phi,\theta) = 1 - \frac{1/\theta^{1/\theta}}{\left(1/\theta+\Lambda_{Ai}(\beta,\phi)\right)^{1/\theta}}$$

yielding the adjusted log-likelihood

$$\ell(\beta, \phi, \theta) = \ell^O(\beta, \phi, \theta) - \sum_{i=1}^{n} \log \left\{ 1 - \frac{1/\theta^{1/\theta}}{(1/\theta + \Lambda_{Ai}(\beta, \phi))^{1/\theta}} \right\}, \tag{4.13}$$

with ℓ^O as defined in (4.7).

Similarly to the Andersen-Gill case, the ascertainment adjustment gives rise to an extra term involving Λ_{Ai} in the score functions for β, and this can be seen to be true also in the score function for θ. The extent of the bias which appears if the ascertainment adjustment is ignored depends in this case also on θ, in addition to ϕ and β. If, again, we consider the Breslow estimator with $\phi_k = \lambda_0(s_k)$ for s_k event time points, we obtain

$$U_{\phi_k}(\beta, \phi, \theta) = \frac{N_k}{\phi_k} - \sum_{i=1}^{n} Y_i(s_k) \exp\left(\beta' \mathbf{x}_i(s_k)\right) h_1(\beta, \phi, \theta) - \\ - \sum_{i=1}^{n} Y_i^A(s_k) \exp\left(\beta' \mathbf{x}_i(s_k) h_2(\beta, \phi, \theta)\right) \tag{4.14}$$

with

$$h_1(\beta, \phi, \theta) = \frac{1/\theta + N_{i.}}{1/\theta + \Lambda_i(t_i; \beta, \phi)}$$

and

$$h_2(\beta, \phi, \theta) = \frac{1/\theta^{1/\theta+1}}{(1/\theta + \Lambda_{Ai}) \left\{ (1/\theta + \Lambda_{Ai})^{1/\theta} - 1/\theta^{1/\theta} \right\}}.$$

4.2.3 Estimation of λ_0

The "baseline" intensity λ_0 is seen as parametrized by a vector of parameters ϕ. Our intention is to cover two cases: fully parametric models (where ϕ is low-dimensional) and semiparametric models (where ϕ is infinite-dimensional).

Parametric models Parametric specifications of λ_0, such as exponential or Weibull, which lead to closed forms of the log-likelihood can be estimated with general purpose optimization software such as the function `optim` in `R`. Such software also provides an estimate of the Hessian matrix at the maximum likelihood estimate from which standard errors can be obtained in a straight-forward way. In this chapter we choose a flexible piecewise constant specification for λ_0, where we consider the baseline intensity to be constant on a small number of intervals which partition the follow-up time. The limits of these intervals are determined so that they contain a roughly equal number of events.

Semiparametric models A semiparametric estimator for λ_0 is, for example, the Breslow estimator, which is obtained by solving the score equations corresponding to the score functions (4.12) for AG and (4.14) for the shared frailty model. Let $\lambda_0(t) = \phi_t$ for

t a known event time and 0 otherwise. The baseline intensity is then parametrized by $\phi = (\phi_1, ..., \phi_N)$ where N is the number of distinct event time points in the data. The difficulties induced by this specification are that the dimension of ϕ may be large, and it becomes larger as there are more unique event time points in the data set. Therefore, direct maximization of the log-likelihood is not usually feasible. We propose a new two-step iterative algorithm to obtain maximum likelihood estimates for semiparametric models.

Without any ascertainment adjustment, the high-dimensional ϕ parameter can be profiled out directly in the AG model (Johansen, 1983), and indirectly, within an Expectation-Maximization algorithm, in the shared frailty model (Nielsen et al., 1992). With ascertainment adjustment, these methods are not available. We propose to alternate between maximizing the log-likelihood with respect to the low-dimensional parameters β and θ for fixed ϕ and updating the high-dimensional parameter by solving a set of "pseudo score equations", which we derive from the score functions (4.12) and (4.14). If we denote the parameters of the model as η, then the score function for ϕ_k takes the form $U_{\phi_k}(\eta) = \frac{N_k}{\phi_k} - h(\eta)$ where h depends on whether the AG or the shared frailty model is used, and whether the likelihood is adjusted for ascertainment. For the adjusted models this can be seen in (4.12) and (4.14). We define the pseudo-score function as

$$\widetilde{U}_{\phi_k}(\phi_k|\tilde{\eta}) = \frac{N_k}{\phi_k} - h(\tilde{\eta}) \tag{4.15}$$

where η is seen as fixed to $\tilde{\eta}$. Solving the equation $\widetilde{U}_{\phi_k} = 0$ with respect to $\phi_k$leads to

$$\hat{\phi}_k = \frac{N_k}{h(\tilde{\eta})},$$

increases the log-likelihood if η is regarded as fixed to $\tilde{\eta}$. Finally, the algorithm follows the following steps. First, choose initial values β^0, θ^0 and ϕ^0, and fix a small $\varepsilon > 0$ as the desired precision.

1. At the ith iteration, maximize $\ell(\beta, \theta|\phi^{(i-1)})$, with ϕ fixed to $\phi^{(i-1)}$, with a general optimization software, e.g. `optim` in R. Obtain the updated $\beta^{(i)}$ and $\theta^{(i)}$.
2. Denote $\tilde{\eta} = \left(\beta^{(i)}, \theta^{(i)}, \phi^{(i-1)}\right)$ and solve the pseudo-score equations (4.15). Obtain the updated $\phi^{(i)}$.
3. Repeat steps 1 and 2 until $\ell(\beta^{(i)}, \theta^{(i)}, \phi^{(i)}) - \ell(\beta^{(i-1)}, \theta^{(i-1)}, \phi^{(i-1)}) < \varepsilon$

The advantage of this procedure is that it can estimate any model with a semiparametric baseline intensity for which an explicit expression of the pseudo-score (4.15) exists. The initial values β^0, θ^0 and ϕ^0 can be obtained from a Cox model ignoring any possible dependence between observations or ascertainment correction. A similar algorithm was proposed for frailty models without ascertainment correction (Gorfine, Zucker, and Hsu, 2006). Simulations which we do not show here indicate that the log-likelihood increases

with each iteration. Furthermore, for the semiparametric AG and shared frailty models without ascertainment adjustment, the estimates of the proposed algorithm coincide with the estimates provided by the available standard software. In the remainder of this chapter we fix the convergence criterion to correspond to $\varepsilon = 10^{-6}$.

Standard error estimates can be obtained from the "non-parametric information matrix", obtained by taking the second derivatives of the log-likelihood $\ell(\eta)$ with respect to all the parameters, including ϕ. This has been shown to lead to valid standard error estimates for the shared frailty model without ascertainment correction (Andersen, Klein, et al., 1997). As long as the estimates in the ascertainment-adjusted model enjoy similar asymptotic properties as the ones in the shared frailty model, a similar reasoning may be applied for this case. Since the semiparametric model can be seen as a limiting case of a parametric model with a piecewise constant baseline, with the piecewise intervals becoming smaller, it is to be expected that inverting the non-parametric information matrix will lead to correct estimates of the standard errors.

4.3 Simulation study

4.3.1 Toy example

We first consider a basic example to illustrate the bias which arises by not taking the ascertainment into account. For this we consider a "full" data set and a "truncated" version of the same data set where the subjects who have not experienced any event are removed. This reflects a simple situation where the selection of subjects is based on a registry where only individuals with at least one occurrence are present, as described by case 3 in Section 4.2.2.

First, we simulate subjects under a scenario where 300 individuals have the same risk, with $\lambda_i(t) = \lambda_i = 1$, without covariates or frailty. The cumulative intensity is then $\Lambda_i(t) = t$ for all subjects. In Figure 41 (left) the estimates $\hat{\Lambda}_i$ based on 20 full and truncated data sets are shown, and with the black line the true value is plotted. The cluster around this line are estimates based on the AG model on the full data set and the other lines are the estimates from the truncated data sets. It can be seen that if the subjects which do not experience any events during follow-up are not part of the data set, the uncorrected estimates are biased upwards.

Next, we simulate data sets and truncate them as before, this time with a binary covariate from a Bernoulli(1/2) distribution, as in (4.1). We repeat this procedure 30 times for a grid of values of β, we estimate $\widehat{\beta}$, and we collect the bias $\widehat{\beta} - \beta$. For every value of β a boxplot of the bias is shown in Figure 41 (center). It can be seen that, for $\beta < 0$, the estimate has a positive bias and for $\beta > 0$ the bias is negative. The absolute value of the bias is larger as β is further away from 0, and for negative values this phenomenon is more severe.

Finally, we simulate a large data set of 3000 patients from the frailty model (4.2), with a gamma distributed random effect (4.3) with $\theta = 1$. We truncate this data set in

the same way as described before. The individual frailty value is unknown, and it can be inferred from the conditional distribution of Z_i given the data. A particularity of the gamma shared frailty model is that this "posterior" distribution is also gamma, however with mean $1/\hat{\theta} + N_{i.}$ and variance $1/\hat{\theta} + \hat{\Lambda}_i(t_i)$, see (Nielsen et al., 1992). From the full data set, we compute the "posterior" frailty estimates, equal to the expectation of this conditional distribution:

$$\widehat{z_i} = \frac{1/\hat{\theta} + N_{i.}}{1/\hat{\theta} + \hat{\Lambda}_i(t_i)}.$$

The logarithm of these estimates are shown in a histogram in Figure 41 (right, above). We show the estimated frailties of the subjects who are part of the truncated data set in Figure 41 (right, below), clearly indicating that the one-in ascertainment favors the selection of individuals with a high frailty value. This is because a high frailty value is associated with a higher rate of recurrent events, leading to an ascertained sample which is less heterogeneous and not representative of the population at risk.

4.3.2 Set up

The idea of the simulation study is to first simulate a random sample of M subjects, from which the estimates of baseline intensity, regression coefficients and eventually frailty variance are obtained. These are regarded as the "correct" estimates. Next, from this data set we obtain 3 different "ascertained" data sets, by selecting only a subset of the M subjects. On these "ascertained" data sets, we perform two analyses: one ignoring the ascertainment correction, in order to assess the extent of the bias induced by event-dependent selection, and one in which the correct ascertainment correction is used, to evaluate how the estimates obtained from the adjusted likelihood compare to the "correct" ones.

By S0 we will refer to the full-data scenario, comprising the M simulated individuals. The three "ascertained" data sets are obtained from the following scenarios:

- By S1 we refer to the situation where only subjects who experience at least one event during follow-up are selected in the sample.
- In S2 only the subjects who experience at least one event during an observation window at the end of the follow-up are kept in the sample.
- In S3 only subjects who experience at least one event in an observation window in the middle of the follow-up are kept; this pertains to a selection scheme similar to S2, where the ascertained subjects are also followed until the end of their follow-up.

In all three cases we assume that the whole follow-up, including the events outside the ascertainment window, is known for the subjects in the sample.

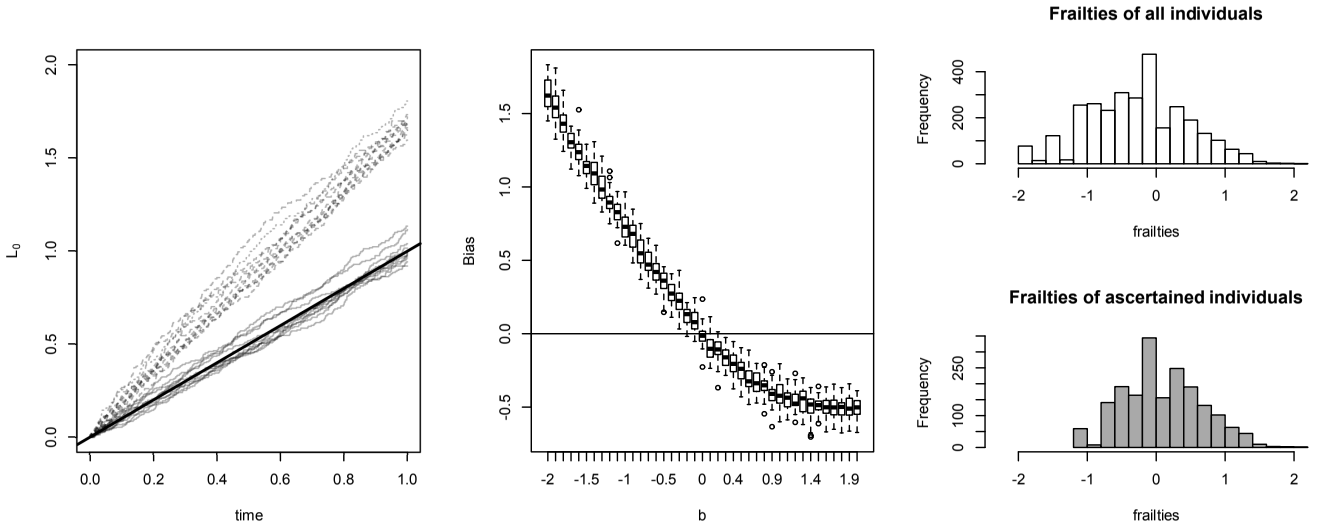

Figure 41: Left: estimates of Λ_0 based on the full data set clustered around the true value (black line) and from the truncated data set (the higher sloped lines); center: bias in log-hazard ratio for two groups when the ascertainment is ignored in the truncated data sets, from the model $\lambda_i(t) = e^{\beta' x_i}\lambda_0(t)$ with x_i an indicator variable for whether the subject belongs to the high-risk group, for values of β between -2 and 2; right: histograms of the posterior log-frailties estimates from the full data set analysis (above) and the frailties of the subjects which experience one event or more during follow-up (below).

The general set-up of the simulations is as follows: we take the baseline intensity $\lambda_0(t) \equiv \lambda_0 = 2$ for each individual. Two covariates are generated independently from a Bernoulli distribution with $P(X_{iq} = 1) = P(X_{iq} = 0) = 0.5$ for $i = 1...M$ and $q = 1, 2$. The regression coefficients used for the simulation are $\beta_1 = 1$ and $\beta_2 = -0.5$. The subjects are censored at time $t_C = 1$ or by a "drop-out process" determined by an exponential distribution with mean $2\exp(x_{i1})$, whichever comes first. For the frailty model (4.2), we simulate M gamma-distributed random variables according to (4.3) with $\theta = 0.5$, and subsequently with $\theta = 1$, and assume that the dropout does not depend on the frailty.

The simulations consist of 1000 replicated data sets, each with $M = 500$ counting processes (individuals) that are simulated from a Poisson process, with the intensities (4.1) for the no-frailty case and (4.2) for the frailty cases, i.e.

$$\lambda_i(t) = 2\exp(\beta_1 x_{i1} + \beta_2 x_{i2})$$

for the $\theta = 0$ (AG) case and

$$\lambda_i(t|z_i) = 2z_i\exp(\beta_1 x_{i1} + \beta_2 x_{i2})$$

for the $\theta = 0.5$ and $\theta = 1$ cases.

For scenario S2 the observation window is chosen as $(0.7, 1.0]$ and for scenario S3 as $(0.3, 0.5)$. Because the data sets used in S1, S2 and S3 are subsets of the full data set S0, they contain fewer individuals. The average sizes of the truncated data sets is shown in Table 41.

Table 41: Data set sizes for S0, S1, S2 and S3 in terms of average number of individuals (M) and average number of events per individual ($N./M$)

		S0	S1	S2	S3
$\theta = 0$					
	M	500	384.15	181.64	173.72
	$N./M$	2.34	3.05	4.08	4.00
$\theta = 0.5$					
	M	500	342.59	161.28	157.93
	$N./M$	2.35	3.43	4.87	4.87
$\theta = 1$					
	M	500	308.37	146.07	144.06
	$N./M$	2.34	3.80	5.56	5.64

For the regression parameters and the frailty variance, the estimates are evaluated according to systematic bias, root mean-squared error and . The systematic bias is defined as

$$\frac{1}{1000}\sum_{j=1}^{1000}(\widehat{\beta}_{qj} - \beta_q)$$

for $q \in 1, 2$, where $\widehat{\beta}_{qj}$ is the estimate of β_q from the j-th simulation and β_q is the true value of the parameter. The root mean-squared error (RMSE) is defined as

$$\frac{1}{1000}\sqrt{\sum_{j=1}^{1000}(\widehat{\beta}_{qj} - \beta_q)^2}$$

and the coverage is the percentage of times that the 95% confidence interval of the estimates contains the true value of the parameter. For the regression coefficients, the estimated standard error obtained from the maximization software is used to construct symmetric confidence intervals. A special case is represented by θ, which is restricted to positive values. Instead, an unrestricted estimate $\log\theta$ is obtained from the maximization of the likelihood together with a standard error $\text{se}(\widehat{\log\theta})$. A symmetric 95% confidence interval for $\log\theta$ can then be constructed on the log-scale as

$$\left[\widehat{\log\theta} - 1.96 * \text{se}(\widehat{\log\theta}),\ \widehat{\log\theta} + 1.96 * \text{se}(\widehat{\log\theta})\right].$$

After exponentiating these bounds, a 95% asymmetric confidence interval is obtained for $\widehat{\theta}$.

Both ascertainment unadjusted and adjusted estimates are obtained from a self-written algorithm which maximizes the likelihoods (4.11) and (4.13), using a piecewise constant parametric form for λ_0. Throughout the simulations, the time axis is split into 8 intervals, which are chosen so that each interval includes roughly the same number of events. This implies that the intervals themselves may vary from simulation to simulation.

4.3.3 Simulation results

Andersen-Gill The estimates of the baseline intensity λ_0 from the unadjusted and adjusted AG model are shown in Figure 42, the estimate from each simulation being represented by a piece-wise constant function. The 8 intervals are of roughly similar length across the simulations due to the Poisson specification in Section 4.3.2. It can be seen that scenario S1 induces an upward bias that seems to be reasonably constant throughout time. S2 and S3 seem to bias the analysis at all time points, to a similar extent as S1, however with peaks during the observation window. The adjusted estimates are unbiased; nevertheless, for the heavier ascertainment scenarios S2 and S3 the estimates seem to exhibit a noticeably larger variance.

Boxplots of the unadjusted and adjusted estimates of β_1 and β_2 are shown in Figure 43. It can be seen that the adjusted estimates are unbiased, although they exhibit a larger variance. Indeed, the ascertainment-adjusted estimates also have higher estimated standard errors (not shown here).

The unadjusted estimators are also associated with an underestimated standard error (not shown here). This reflects itself in the poor coverage properties of the estimators. The simulation results of the $\theta = 0$ (AG) case are summarized in Table 2. It can be

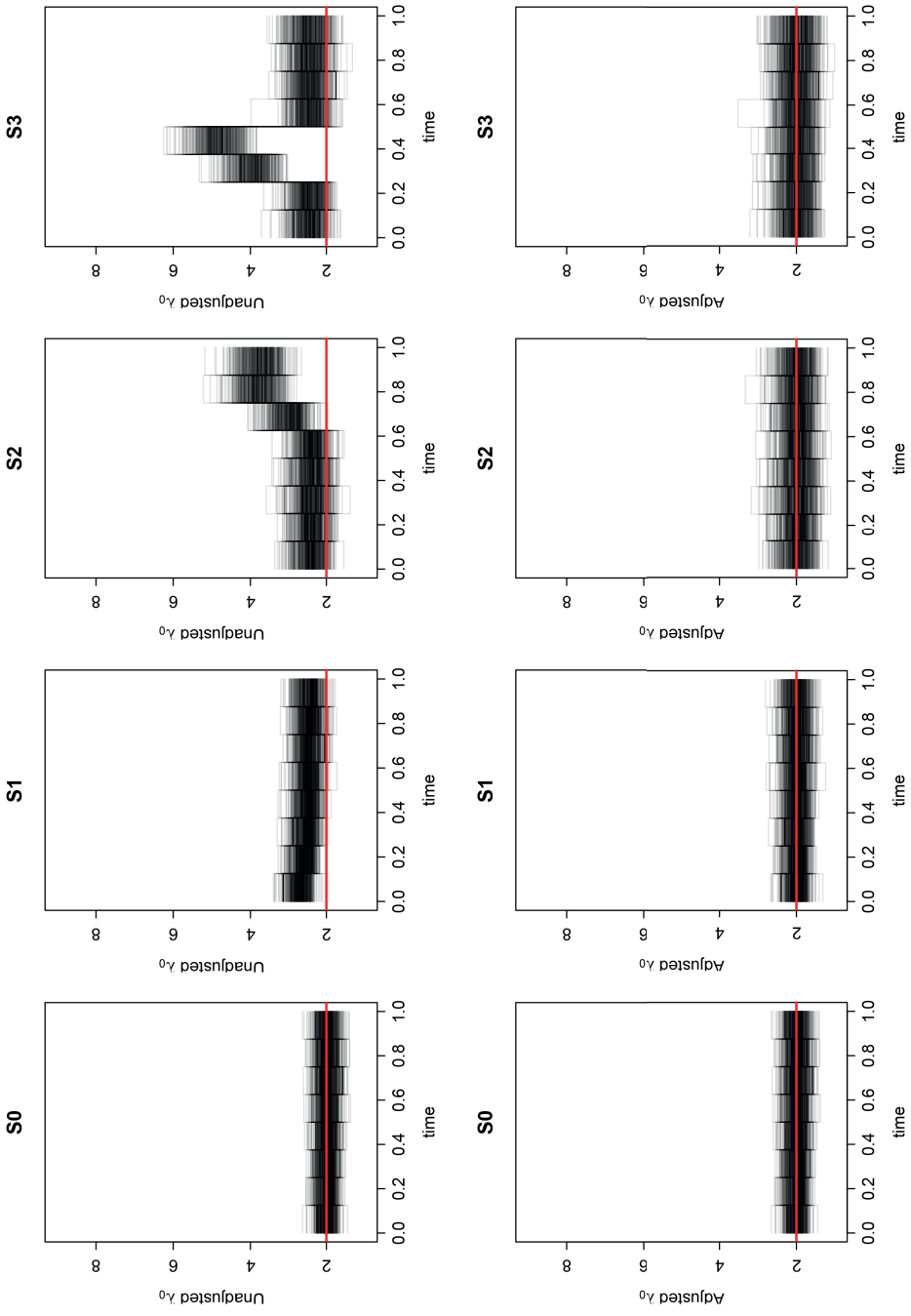

Figure 42: Baseline intensities, AG model; the horizontal line at $y = 2$ corresponds to the true $\lambda_0 = 2$.

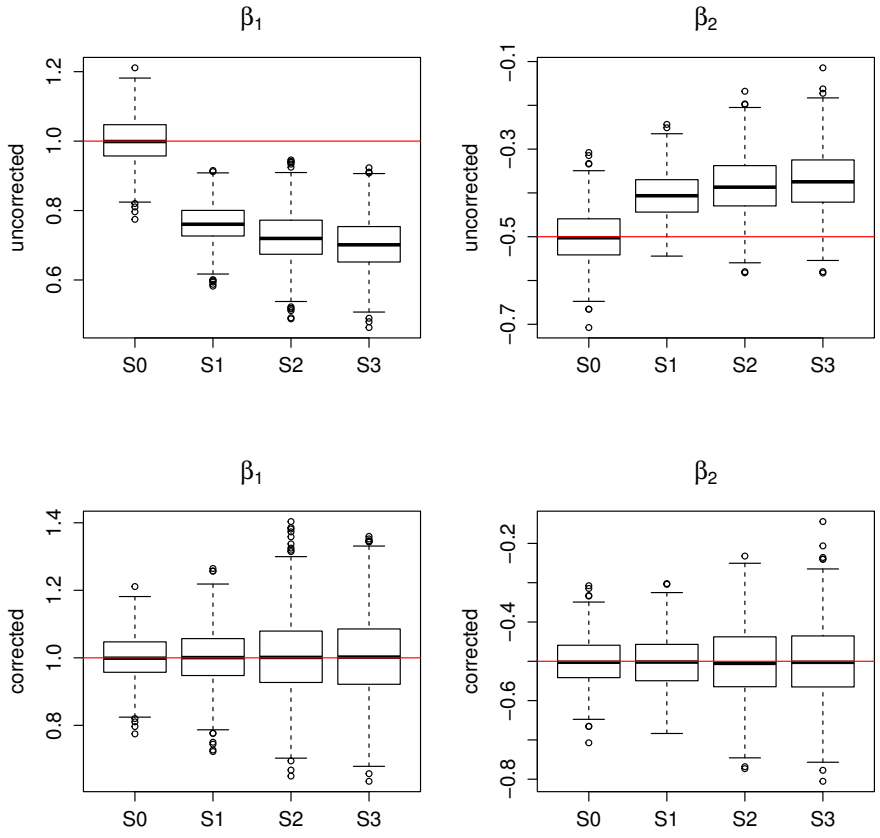

Figure 43: Point estimates for regression coefficients, AG model. The horizontal lines correspond to the true value of the parameters, $\beta_1 = 1.0$ and $\beta_2 = -0.5$.

observed that, if not corrected for ascertainment, the estimates of β_1 are affected more than those for β_2. This is due to an imbalance which is caused in the data set: because x_1 also influences the risk of censoring as specified in Section 4.3.2, it is less likely that the subjects experience events during the ascertainment window, simply because they have less time at risk. The correct model, which adjusts for the ascertainment, provides unbiased estimates and show a drastic reduction of the RMSE from their unadjusted counterparts.

Table 42: Simulation results, AG model

		Uncorrected			Corrected		
		S0	S1	S2	S3	S1	S2	S3
β_1								
	Bias	0.003	-0.232	-0.273	-0.293	0.010	0.012	0.011
	RMSE	0.062	0.238	0.283	0.303	0.085	0.117	0.127
	Coverage	0.966	0.036	0.124	0.106	0.950	0.956	0.952
β_2								
	Bias	0.000	0.096	0.119	0.127	0.000	0.002	0.001
	RMSE	0.057	0.109	0.136	0.146	0.067	0.089	0.091
	Coverage	0.968	0.678	0.690	0.690	0.952	0.956	0.962

Shared frailty We employ two scenarios for the frailty models, one of high heterogeneity ($\theta = 1$) and one of medium heterogeneity ($\theta = 0.5$). For the $\theta = 1$ scenario, the estimates of the baseline intensities are shown in Figure 44. By contrast with Figure 42, the 8 intervals are more different across simulations. This is due to the fact that the frailty-induced heterogeneity induces a more uneven spread of the events in time. Therefore, when determining the piecewise constant intervals as described in Section 4.3.2, the outcome can vary more than in the AG case. The larger heterogeneity, as represented by $\theta = 1$, also leads to more bias when the ascertainment is not adjusted for. However, the adjusted estimates are still unbiased and exhibit a larger variance, similarly with those shown in Figure 42. The baseline intensity estimates with $\theta = 0.5$ (not shown here) show a similar behavior.

The results are summarized in Tables 43 and 44, as well as in Figures 45 and 46. In terms of the regression coefficients, the bias is more severe in the higher heterogeneity scenario, even if only slightly so. As in the case of the AG model, the corrected estimates are unbiased. The major difference lies in terms of the estimate of θ. The unadjusted estimates show a large bias towards 0, notably more acute when $\theta = 1$. The large bias seems to lead to a very large variance of the adjusted estimators, as can be seen in Figure 45. Nevertheless, in the corrected estimators consistently provide major improvements in terms of RMSE and coverage over their unadjusted counterparts.

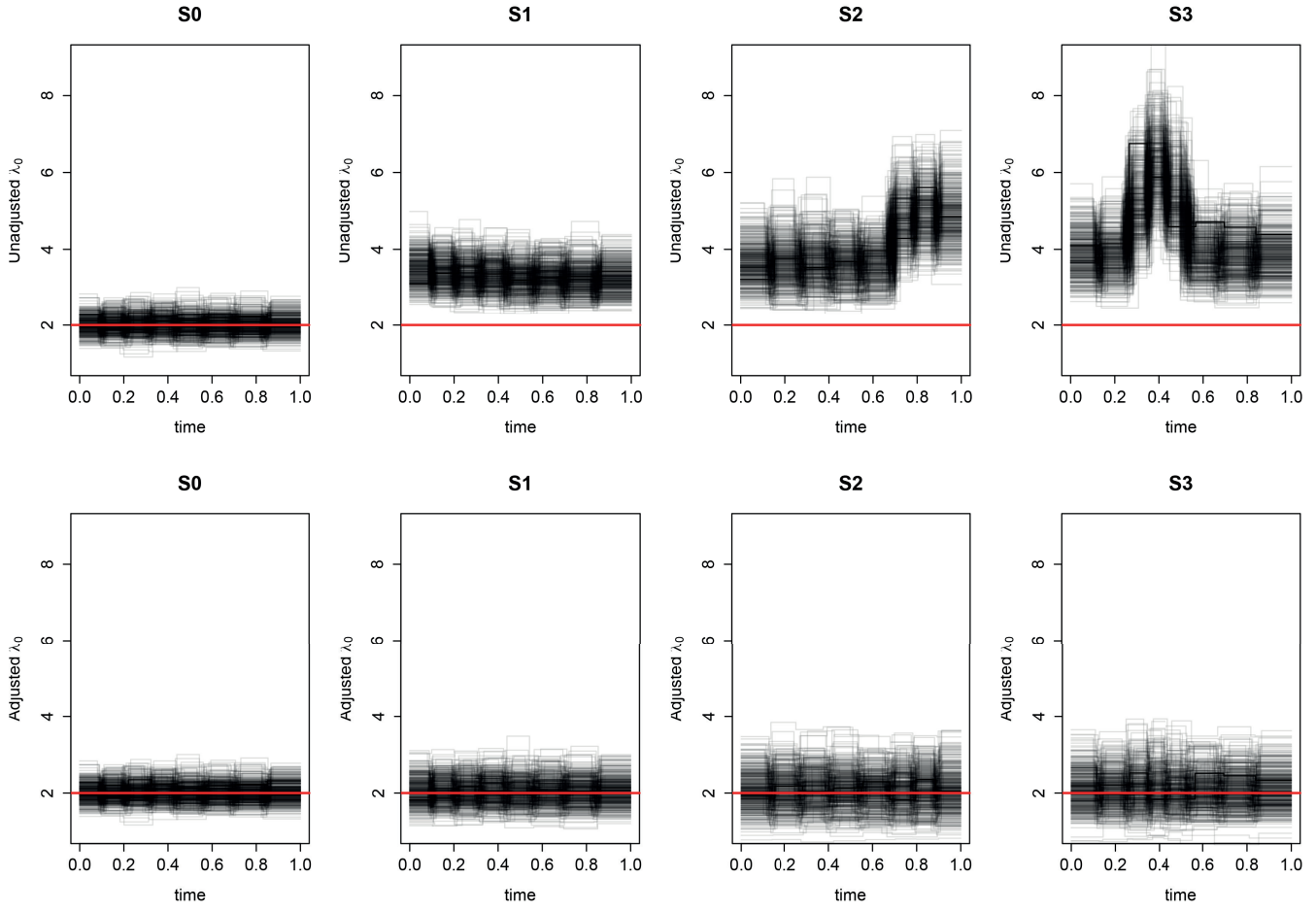

Figure 44: Baseline intensities, shared frailty model with $\theta = 1$; the horizontal line at $y = 2$ corresponds to the true $\lambda_0 = 2$.

Table 43: Simulation results, shared frailty model, $\theta = 1$

		Uncorrected				Corrected	
	S0	S1	S2	S3	S1	S2	S3
β_1							
Bias	0.004	-0.296	-0.302	-0.305	0.001	0.009	0.005
RMSE	0.116	0.311	0.327	0.332	0.148	0.190	0.196
Coverage	0.946	0.143	0.343	0.352	0.946	0.945	0.934
β_2							
Bias	0.002	0.140	0.147	0.145	0.001	0.001	-0.002
RMSE	0.115	0.172	0.193	0.193	0.139	0.175	0.180
Coverage	0.947	0.660	0.758	0.765	0.948	0.956	0.937
θ							
Bias	-0.007	-0.657	-0.695	-0.716	0.001	0.004	0.011
RMSE	0.108	0.659	0.697	0.718	0.237	0.344	0.404
Coverage	0.963	0.000	0.000	0.000	0.963	0.963	0.946

Table 44: Simulation results, shared frailty model, $\theta = 0.5$

		Uncorrected				Corrected	
	S0	S1	S2	S3	S1	S2	S3
β_1							
Bias	0.001	-0.286	-0.297	-0.301	0.000	0.003	0.003
RMSE	0.096	0.298	0.316	0.320	0.127	0.163	0.168
Coverage	0.937	0.069	0.240	0.241	0.935	0.941	0.934
β_2							
Bias	0.000	0.127	0.136	0.137	-0.003	-0.002	-0.005
RMSE	0.092	0.151	0.171	0.173	0.113	0.145	0.148
Coverage	0.941	0.630	0.717	0.707	0.954	0.954	0.946
θ							
Bias	-0.005	-0.304	-0.327	-0.34	-0.004	-0.007	-0.007
RMSE	0.066	0.306	0.330	0.343	0.109	0.154	0.169
Coverage	0.958	0.000	0.000	0.000	0.958	0.957	0.941

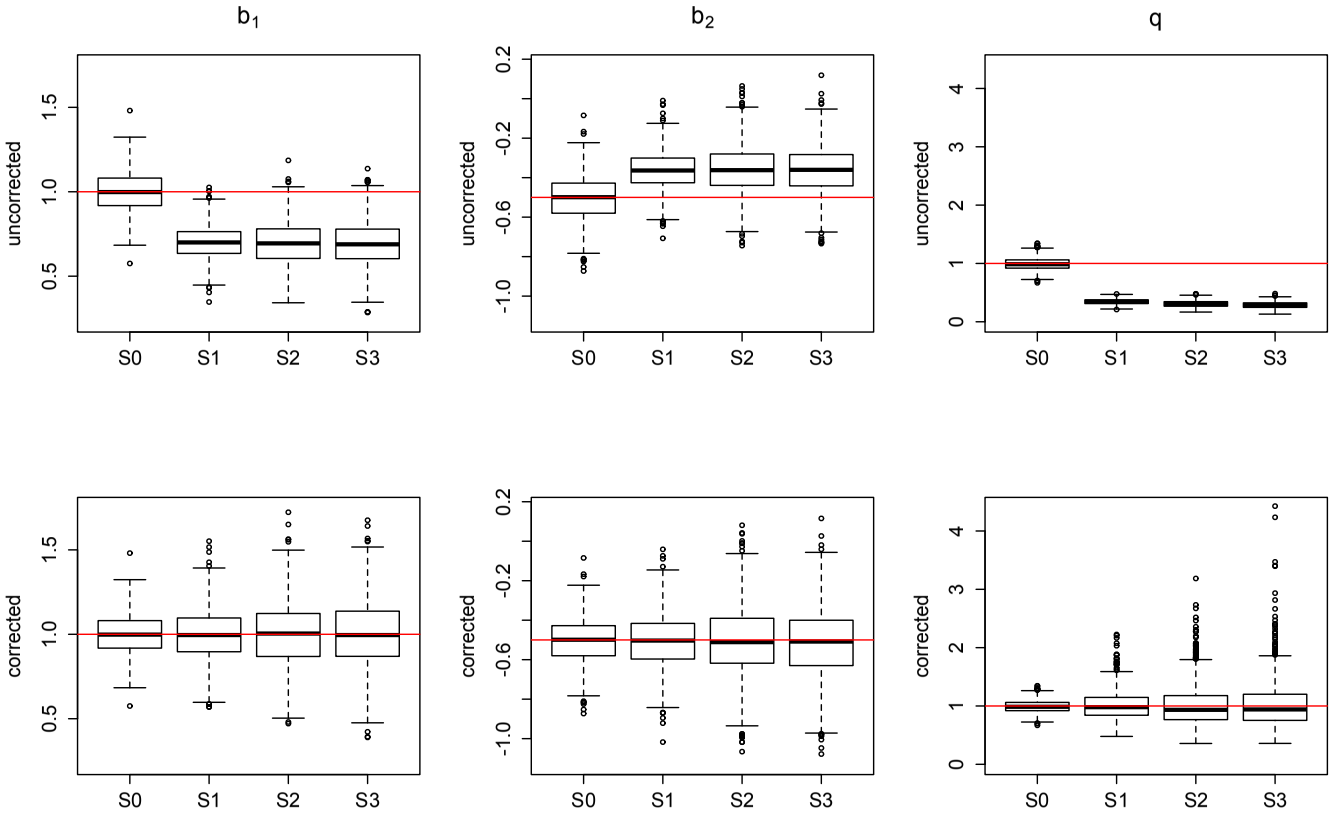

Figure 45: Point estimates, $\theta = 1$. The horizontal line corresponds to the true value of the parameters. Top: without ascertainment correction, bottom: with ascertainment correction.

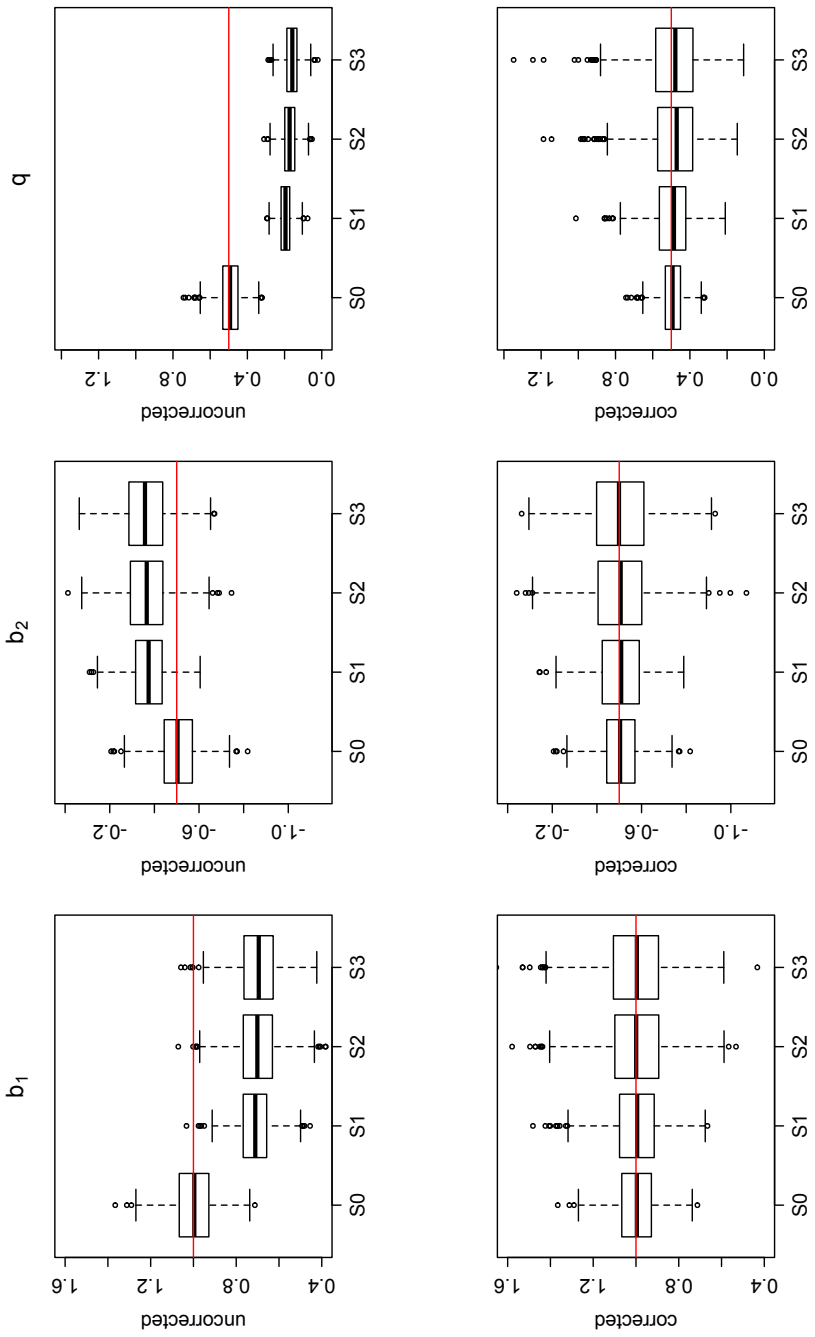

Figure 46: Point estimates, $\theta = 0.5$. The horizontal line corresponds to the true value of the parameters. Top: without ascertainment correction, bottom: with ascertainment correction.

4.3.4 Incomplete history

In the selection schemes described in Section 4.2, it is essential that, if an individual is selected for the study, the whole history of events outside the observation window is collected. As will be seen in the motivating example in Section 4.4, this might not always be the case. To assess the performance of the indicated adjusted models, we consider the scenario when the data on the history previous to the beginning of the observation period is (partly) missing. Using the same data sets that were simulated before obtained under selection schemes 2 and 3, we induce this incomplete character of the data in the following way. In the "mild incompleteness" scenario, 10% of the individuals are randomly selected from the ascertained data sets. For them, a "recollection time" is generated from a uniform distribution between 0 and the left time point of the observation window (0.7 in S2 and 0.3 in S3). The events before this time point are subsequently removed from the data set and the adjusted and unadjusted analyses are performed. In the "heavy incompleteness" scenario, the same is repeated with 50% of the individuals in the data sets. The results for $\theta = 1$ are summarized in Table 45. For $\theta = 0.5$, similar results were observed and are not shown here.

The corrected estimates of the frailty variance θ seem to be the most severely affected, particulary in scenario S2. The large bias (0.114 and 0.668), coupled with very wide confidence intervals (with coverages of 0.975 and 0.998) indicate that the standard errors are overestimated. In Figure 47 we plot the ascertainment-adjusted estimated baseline hazards for this case. By comparison with 44, it can be seen that the ascertainment adjustment does not work as well. The missing event history before time 0.7 leads to the underestimation of the intensity of the recurrent events process, mostly visible in the 50% missing case.

The general conclusion is that, when unadjusted for ascertainment, the incompleteness seems to slightly aggravate the problems illustrated in Tables 43 and 44. Nevertheless, the bias, RMSE and coverage remain comparable. The ascertainment adjustment seems to work well also with the incomplete data sets in terms of regression coefficients, at the price of a small increase in bias in and a slight increase in RMSE. In terms of the estimation of θ, we remark that the adjustment induces a positive bias to the estimates. In addition, the overly optimistic results of the coverage, in conjunction with the increase in RMSE and the bias results, reveals an over estimation of the standard errors. We conclude that, when the events outside the observation window are not completely collected, the ascertainment correction is robust in regards to the regression coefficients, however the frailty variance parameter should be interpreted with caution.

4.4 Data analysis

Data description The motivating data set comprises observations on primary spontaneous pneumothoraces (PSP). Risk factors for developing a PSP are male gender, smoking, and age, with a peak at 25-35 years of age; see Baumann and Noppen (2004) for an overview on the recurrent characteristics of PSPs.

Table 45: Simulation results, $\theta = 1$, with 10% and 50% missing data outside the observation window

		Uncorrected				Corrected			
		S2 10%	S2 50%	S3 10%	S3 50%	S2 10%	S2 50%	S3 10%	S3 50%
β_1									
	Bias	-0.310	-0.333	-0.308	-0.319	0.011	0.037	0.006	0.009
	RMSE	0.335	0.360	0.335	0.345	0.192	0.221	0.198	0.203
	Coverage	0.334	0.316	0.350	0.318	0.942	0.948	0.935	0.941
β_2									
	Bias	0.157	0.167	0.154	0.158	0.009	-0.003	0.007	0.003
	RMSE	0.205	0.215	0.200	0.204	0.188	0.206	0.183	0.189
	Coverage	0.719	0.715	0.744	0.738	0.944	0.959	0.929	0.929
θ									
	Bias	-0.689	-0.675	-0.716	-0.722	0.114	0.668	0.037	0.101
	RMSE	0.691	0.677	0.718	0.724	0.432	1.136	0.416	0.498
	Coverage	0.000	0.000	0.000	0.000	0.975	0.998	0.959	0.968

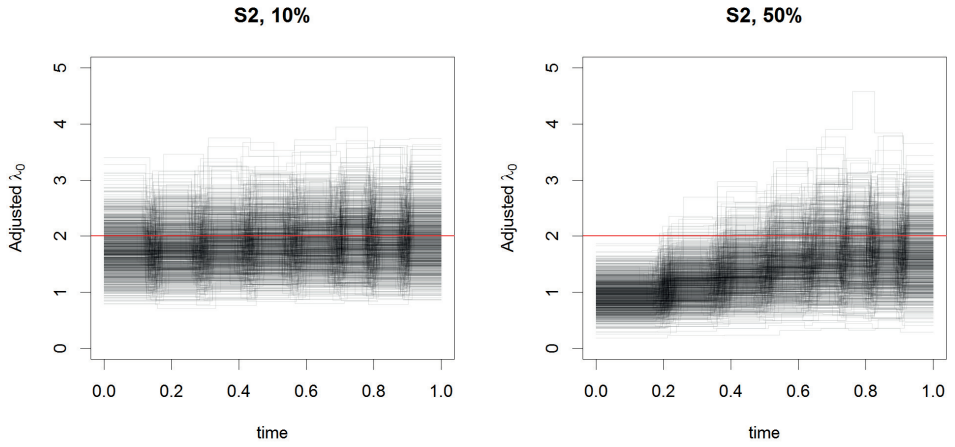

Figure 47: Ascertainment-adjusted estimates of baseline intensities, shared frailty model with $\theta = 1$, scenario S2, with 10% incompleteness (left) and 50% incompleteness (right) before the ascertainment window. the horizontal line at $y = 2$ corresponds to the true $\lambda_0 = 2$.

More recently, several genetic syndromes have been associated with an increased risk for (recurrent) episodes of spontaneous pneumothorax, like the Birt-Hogg-Dubé (BHD) syndrome, see Menko et al. (2009) and Johannesma et al. (2015). BHD is vastly under-diagnosed and usually patients with PSP do not receive a genetic test for this event, although (recurrent) PSP in BHD patients are caused by multiple cysts in the lower parts in the lung. By contrast, the non genetic PSP does not show these cysts at all on a thoracic CT-scan.

A variety of treatments are available for PSP, which we can divide into 3 categories: conservative (waiting, drainage, manual aspiration), sticking (pleurectomy, (chemical) pleurodesis) and cutting (lobectomy, bullectomy). Usually, the patients first receive a non-invasive (conservative) treatment, followed by a more invasive treatment for the next recurrent episodes.

The selection of the patients occurred as follows; between 2010 and 2014 a questionnaire was sent to the patients treated after 1990 for (recurrent) PSP. A number of respondents returned to the hospital and received a folliculin (*FLCN*) test to confirm BHD and then their PSP and treatment history was recorded. Information on the location of the PSP was not available, except for which lung the event took place in. The age of the patients at each event was recorded, approximated to the year. In total, the data set comprises 95 patients out of which 65 had PSP episodes only on one of the lungs, with a total of 220 episodes of PSP.

We define a tie as PSPs occurring in the same year in the same lung. This is observed in 26 of the 125 lungs with events in the data set. The presence of ties poses a difficulty

in analyzing the gap times, since 0-length gaps are not meaningful. Even if artificially spread out over a small interval of time, this might lead to a wrong impression on the length of the true gaps. This poses less of a problem if the intensity of the occurrences is treated as a non-homogeneous Poisson process, with age as time scale. This is the course that we follow in this analysis.

The data are shown in a Lexis diagram (Plummer and Carstensen, 2011) in Figure 48. The ascertainment process can be clearly seen. The general lack of events prior to 1990 casts some doubt on the completeness of the retrospective data collection, however this aspect is not further considered here. The effects of incomplete collection of the event history prior to the observation window were analyzed by simulation in Section 4.3.4.

Model construction The next step is the construction of a model. Each individual is represented by two counting processes corresponding to the two lungs, λ_i^L and λ_i^R. The intensities of these two processes can be influenced either by subject-level factors or by lung-specific history. A priori, there is no reason why one lung should be at a higher risk than the other. The general idea is that the subject-specific factors affect both lungs equally, while the difference in treatments between the lungs account for the observed differences between λ_i^L and λ_i^R.

We choose to model the events on the age time scale for which we take a baseline intensity common to all lungs from all subjects. The non-parametric estimate of the baseline intensity will have jumps only at event times, with the first event occurs at age 16. However, for the piecewise constant baseline a start of the recurrent events process must be explicitly defined. We choose this as the age of 15, since the risk of PSP before puberty is practically 0. We include BHD carrier indicator as a time-constant covariate in the model. At the lung level, we assume that the lungs may be in 4 states: "not under treatment", if there was no previous event, or under one of the 3 treatments: *conservative*, *sticking* or *cutting*. To account for differences between individuals, we include a gamma frailty which is shared for both λ_i^L and λ_i^R. These may be due to unmeasured covariates, such as shared environmental or behavioral variables.

In terms of treatments, *sticking* and *cutting* should be compared to *conservative*. To accommodate this, we assume that the intensity gets multiplied by $\exp(\beta_{\text{cons}})$, $\exp(\beta_{\text{cons}} + \beta_{\text{stick}})$ or $\exp(\beta_{\text{cons}} + \beta_{\text{cut}})$ according to which treatment the lung is under. In this case, $\exp(\beta_{\text{cons}})$ is the intensity ratio between one lung which had at least one event and is under conservative treatment and one lung which had no event and is not under treatment. This effect can also be interpreted as the intensity ratio between a lung which experienced PSPs and one which has not. On the other hand, $\exp(\beta_{\text{stick}})$ and $\exp(\beta_{\text{cut}})$ are intensity ratios between a lung which is under *sticking* or *cutting* and a lung under *conservative* treatment. For example, for individual i without BHD and with frailty z_i, which at time t has the left lung under *cutting* treatment and the right lung on the

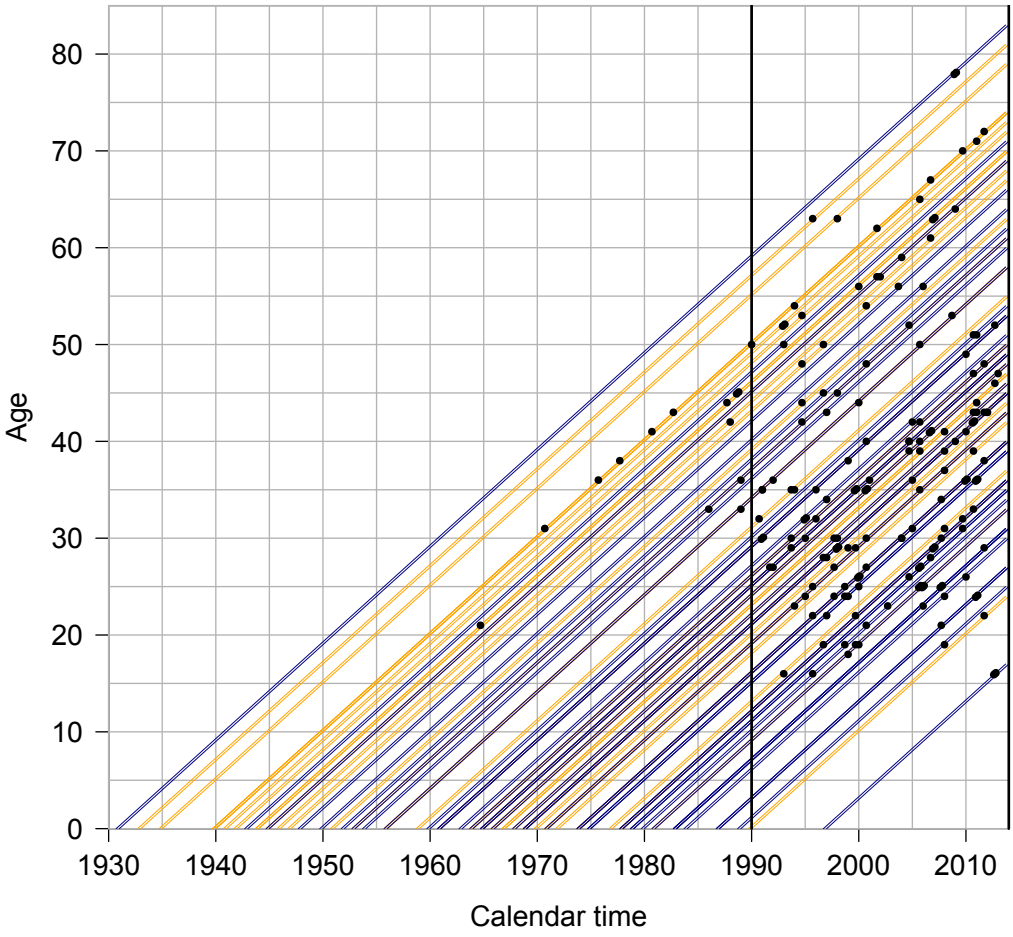

Figure 48: Lexis diagram of the PSP data. In blue the non-BHD patients. Dots represent observed events. The ascertainment window is marked between vertical lines.

conservative treatment, the intensities are

$$\lambda_i^L(t|z_i;\beta,\phi) = z_i \exp(\beta_{\text{cons}} + \beta_{\text{cut}})\lambda_0(t;\phi)$$
$$\lambda_i^R(t|z_i;\beta,\phi) = z_i \exp(\beta_{\text{cons}})\lambda_0(t;\phi)$$

A question of interest is whether the treatments perform equally for BHD and non-BHD patients. Finally, we also include interactions between these terms and BHD status.

Below, we show the results from a parametric baseline with 7 piecewise constant intervals and a semiparametric model.

Results The results are described in Table 46. The p-values for the regression coefficients are obtained from a Wald test statistic. It can be seen that adjusting for the ascertainment does not drastically influence the point estimates of the regression coefficients, which change at most by one standard error. Also, the estimated standard errors are larger in the adjusted model, in both the piecewise constant and in the semiparametric models. The noticeable effect of this is that the main effect of the conservative treatment loses siginificance, with the p-value increasing form 0.02 to about 0.3.

Other than that, we remark that statistical significance at the $\alpha = 0.05$ level was not observed for any of the variables in the model. Nevertheless, the adjusted model does give slightly different results. It can be seen that BHD patients are at a higher risk for PSPs as compared to non-BHD patients. For the non-BHD group, it can be seen that all treatments elevate the intensity of the event process. Among the 3 treatments, the sticking seems to perform better. For the BHD group, the cutting treatment seems to perform better, relative to conservative or sticking. If one of the lungs does not have any events, the intensity of the treated lung relative to the untreated one is modified multiplicatively by a factor of $\exp(0.419 + 0.075) = 1.63$. Conversely, for sticking this is $\exp(0.419 - 0.187 + 0.075 - 0.166) = 1.15$ and for cutting $\exp(0.419 + 0.595 + 0.075 - 0.934) = 1.16$.

In Figure 49, the unadjusted and adjusted baseline intensity estimates are shown, for both the semiparametric and the piecewise constant models. Similarly with the results of the simulation study in Section 4.3, the unadjusted baseline is overall larger than the adjusted estimate.

The p-values are missing in the case of the frailty variance θ, because the null hypothesis $H_0 : \theta = 0$ is at the border of the parameter space and a Wald test would not be valid in this case. A Likelihood Ratio Test for $H_0 : \theta = 0$ based on a χ^2 distribution with 1 degree of freedom can be constructed by contrasting the estimated frailty model versus the AG model, which is seen as the limiting case when $\theta \longrightarrow 0$, see Therneau and Grambsch (2000) and Nielsen et al. (1992). The LRT statistics for this hypothesis in the unadjusted / adjusted models are < 0.01 / 10.64 (piecewise constant) and < 0.01 / 8.57 (semiparametric model), corresponding to p-values of 0.98 / < 0.01 (piecewise constant) and 0.99 / < 0.01 (semiparametric model). The large differences in significance can be explained by noting that, without correcting for ascertainment, the effect of the frailty is not captured at all, as was seen in the simulation study in Section 4.3. It can however

Table 46: Data analysis results

		Unadjusted			Adjusted		
		Coef.	SE	p	Coef.	SE	p
Piecewise constant							
	BHD	0.073	0.19	0.71	0.212	0.35	0.55
	Cons	0.866	0.36	0.02	0.500	0.44	0.26
	Stick	-0.202	0.42	0.69	-0.220	0.51	0.67
	Cut	0.525	0.48	0.28	0.567	0.60	0.35
	BHD:Cons	0.126	0.44	0.78	-0.001	0.52	1.00
	BHD:Stick	-0.167	0.52	0.75	-0.087	0.62	0.89
	BHD:Cut	-0.707	0.67	0.29	-0.958	0.85	0.26
	θ	< 0.001	*NA*	-	1.302	4.18	-
Semiparametric							
	BHD	0.069	0.19	0.72	0.191	0.22	0.39
	Cons	0.871	0.38	0.02	0.419	0.42	0.33
	Stick	-0.221	0.42	0.61	-0.187	0.46	0.69
	Cut	0.491	0.49	0.32	0.595	0.55	0.28
	BHD:Cons	0.173	0.46	0.71	0.075	0.49	0.88
	BHD:Stick	-0.198	0.53	0.71	-0.166	0.58	0.77
	BHD:Cut	-0.646	0.67	0.34	-0.934	0.77	0.23
	θ	< 0.001	0.07	-	1.73	3.51	-

be seen that the standard errors corresponding to the estimator of θ are very large in the adjusted models, which was also the case in the simulation study in Section 4.3.4. This is in line with the suspicion that the history before 1990 was incompletely collected. The same phenomenon of large estimates and very large standard errors was observed in the same context in Section 4.3.4.

4.5 Discussion

We have shown in this chapter that event-based ascertainment may lead to biased results when unaccounted for. This bias can be severe and it may lead to very weak coverage properties, and the true effect of various factors might not be captured at all. We can correct for this bias with the methods proposed in Section 4.2. The merit of the approach used in this chapter is that unbiased results can be obtained if the event-dependent selection conditions are correctly accounted for in the estimation method. Furthermore, it was seen in Section 4.3 that the adjusted estimators only exhibit a small increase in the root mean-squared error as compared to the full-data scenario, as seen in Table 2, despite a smaller sample size. This suggests that the same results can be obtained from

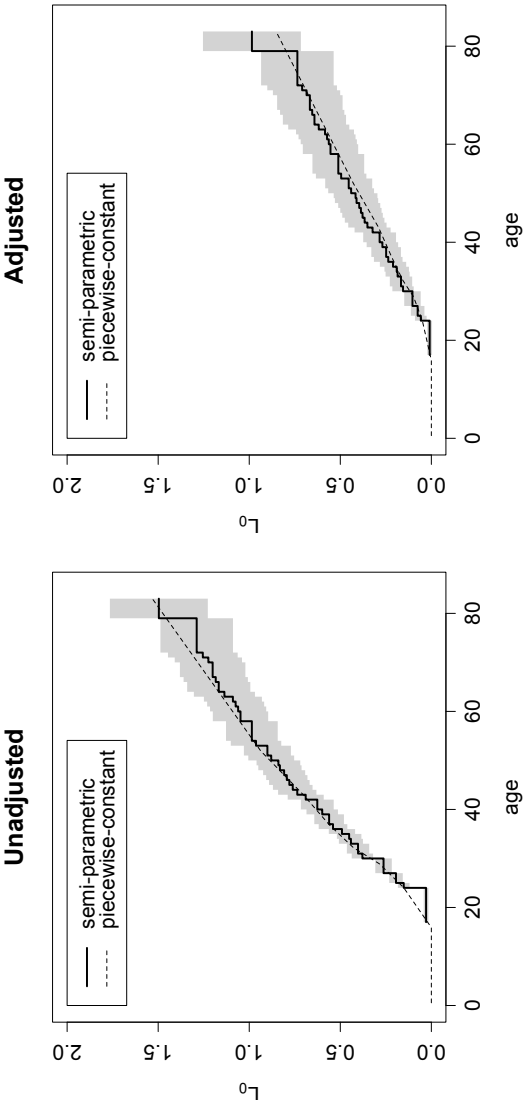

Figure 49: Adjusted and unadjusted estimates of the cumulative baseline intensity Λ_0. The gray band delimits the 95% confidence interval for the semiparametric estimate of Λ_0. With dotted lines, the parametric estimate with 7 piecewise constant intervals.

the two study designs, prospective study and retrospective study with event-dependent selection, as long as the ascertainment is correctly modeled. Hence, retrospective studies on recurrent events might prove to be a viable alternative to the prospective studies.

There are several limitations to the approach used throughout this chapter. First, we assumed that the whole event history of an individual can be collected at the time of the selection. This might not be true, especially when the event history has to be "remembered" by the patients. It can be seen in Figure 48 that very few events seem to happen before the start of the study (1990). It is possible that the subjects did not recall earlier events, or that registry data was not available for all the patients. However, the proposed methods showed promising results, as long as the complete history is collected for most individuals. The effects of incomplete collection of data outside the observation window are analyzed by simulation in Section 4.3.4.

Second, it is also common in the study of recurrent events that the subjects have to be alive at the time of the selection. If the rate of the recurrent events is associated with the terminal event, then joint models for recurrent and terminal events should be adopted; see, for example, Liu, Wolfe, and Huang (2004). In the data used in Section 4.4, it is reasonable to assume that death is not an event of interest, since the recurrent events in this case are not life-threatening. Nevertheless, Cook and Lawless (2007, ch. 7.3) provide some indication of possible strategies for this type of selection.

Third, as seen in Section 4.3, the adjusted estimators show a larger variance than their unadjusted counterparts, which is especially visible in the estimates of the frailty variance. This indicates that the frailty distribution itself is harder to identify than covariate effects in ascertainment-adjusted models.

Among the advantages of the approach presented in this chapter, is that several extensions can be obtained to accommodate more complicated models. First, in the framework outlined in Section 4.2, the distributional assumption for the frailty (gamma distribution) can be relaxed. Likelihoods can be constructed from (4.6) for a larger family of distributions, however these do not lead to closed form expressions; see Hougaard (2000).

Second, other similar models which lead to similar likelihood expressions as (4.4) or (4.6) could be accommodated with this approach. An example is a two-state Markov model for duration of recurrent episodes, where only subjects who have a first recurrence in an observation window are ascertained; see Cook and Lawless (2007, ch. 6.5).

Finally, we note that the framework introduced in Section 4.2 can be itself extended. As long as A_i is an event which is more general than O_i, in the sense of equation (4.8), a similar argumentation can be employed. This can be achieved by extending the definition of what amounts to the event history. Several examples can be found in Cook and Lawless (2007, ch. 7.3).

The promising results shown by the semiparametric estimation method proposed in Section 4.2.3 suggest that the properties of the algorithm should be further investigated, and completed by a proof of convergence. The `R` code used for the simulations in Section 4.3 is available upon request from the corresponding author. Future work, espe-

cially in terms of software development, would likely prove to be useful for clinicians. A focus of future research will be to provide an extension of R's `survival` package for ascertained and truncated data.