Mathematical Statistics
Stockholm University

Proportional hazards model for matched failure time data

Johan Zetterqvist

Examensarbete 2013:1

Postal address:
Mathematical Statistics
Dept. of Mathematics
Stockholm University
SE-106 91 Stockholm
Sweden

Internet:
http://www.math.su.se/matstat

Mathematical Statistics
Stockholm University
Examensarbete **2013:1**,
http://www.math.su.se/matstat

Proportional hazards model for matched failure time data

Johan Zetterqvist*

February 2013

Abstract

The aim was to compare proportional hazards models for matched failure time data in terms of underlying assumptions and requirements for causal inference. We also wanted to compare the models in terms of power and performance under model misspecification. Four models were compared: the stable-Weibull frailty model, the stratified Cox model, the marginal Weibull model and the marginal Cox model. Strict assumptions behind the stable-Weibull model makes it less useful for unmatched data. The stratified Cox model is the most appropriate model when one wants to make causal inference from unmatched clustered data. Matched data from different frailty models were generated to compare the methods empirically. The power of the stable-Weibull model was significantly larger than the power of the stratified Cox. The performance of the stable-Weibull model was sensitive to misspecification of the baseline hazard. The marginal Weibull model and the marginal Cox model performed very similar. For small intra-cluster dependence, they performed similar to the stable-Weibull model in terms of power. When misspecified, the marginal Cox performed well for small to moderate dependence.

*Postal address: Mathematical Statistics, Stockholm University, SE-106 91, Sweden. E-mail:johan.zetterqvist@ki.se. Supervisor: Arvid Sjölander.

Preface

This report is a thesis for the degree of master worth 30 ECTS in mathematical statistics at the Department of Mathematics at Stockholm University.

I want to express my deepest gratitude to my supervisor Arvid Sjölander at the department of Medical Epidemiology and Biostatistics at Karolinska Institutet for formalizing this project and for supervising me. Thank you for all the hours of discussions we had, your valuable feedback, your pedagogical explanations, your support and your patience with me. I have learned a lot during this time.

I would also like to thank Henrik Larsson and Paul Lichtenstein for your interest and for your support.

Contents

1 Introduction

Mathematical statistics gives us models to measure statistical associations. In applied science, on the other hand, one is often primarily concerned with causal relations. The fact that there is an association between two factors does not, in itself, entail that there is a causal relation between the factors. In order to give a causal interpretation of a statistical association, one has to be sure that the association is not confounded by other factors. For example, suppose a statistical association between coffee consumption and lung cancer is found. Although it might not be wrong to say that high consumers of coffee run a higher risk of getting lung cancer, we shouldn't conclude that coffee or caffeine *causes* lung cancer. It is well established that smoking causes cancer. It is also known that high consumers of coffee generally smoke more than low consumers and non coffee drinkers. Therefore the association between coffee drinking and smoking may, to some extent, be caused by the higher prevalence of smoking among high consumers of coffee than among low consumers. If the variable smoking is measured, this measure may be used to estimate the effect of coffee drinking conditional on smoking status. This effect is then unconfounded by the variable smoking.

To increase efficiency, matching is often used in observational studies. For binary outcomes (event/no event) it's possible to match on the outcome. In 1:k matched case-control studies each case (an individual with an event) is matched to k unrelated controls (individuals without an event) with the same values of the confounding variables. This procedure aims to maximize the efficiency when adjusting for the confounding variables. In the above example, this corresponds to matching each lung cancer case to k controls without lung cancer such that the cases have a similar history of smoking as the controls.

In 1:k matched cohort studies the matching is instead made for a binary exposure. Each exposed individual is then matched to k unrelated controls with the same values of the confounding variables. In this way, the distribution of the confounding variables will be the same among the exposed as among the unexposed. Thus the association between the confounding variables and the exposure is eliminated. In terms of the example, this corresponds to matching the high consumers of coffee with low consumer of coffee with the same smoking habits. In this report we will only consider 1:k cohort studies.

In contrast to data matched by design, data may be "naturally" matched,

as in sibling studies. Siblings share a lot of potential confounders, e.g. childhood environment and DNA. For instance, monozygotic twins share all of their DNA, while other full siblings share (on average) 50 % of their DNA. By choosing sibling pairs which are discordant on the exposure of interest (that is one exposed and one unexposed), we get 1:1 matched cohort data. In this way, the data is matched on many potential confounders, some of which may be hard or impossible to measure.

Matched data contain clusters of individuals sharing the same values of the matching variables. Because of this, there will be a correlation between members of the same cluster. Statistical analysis of matched designs have to account for this correlation in data. Otherwise estimates may be biased or the standard errors might be wrong.

One way to deal with this correlation is to measure the association between the exposure and outcome conditional on cluster membership. The association is assumed to be of the same magnitude on the chosen scale across all clusters (e.g. constant conditional odds ratio). Such an association is often called a "cluster-specific" association or a "within" association. Sjölander et al. (2012) have shown that, for 1:1 matched cohorts with binary exposure and binary outcomes, if the set of matching variables contain all confounders, the (logistic) within association can be interpreted as the causal effect in a sub-population defined by the distinct levels of the matching variables. In general, these methods assume independence within clusters. This is generally true for data matched by design. For "naturally" matched data, like twin or sibling data, this is not necessarily true. For instance, if an event in a twin has an influence on the probablity of an event in his/her co-twin, their outcomes are dependent conditional on the cluster.

One may also focus on the marginal (over clusters) association between exposure and outcome, treating the dependence within clusters a nuisance. Such models are often described as "population averaged" models. For 1:k matched cohorts, the association between the matching variables and exposure is blocked. Therefore, assuming that the set of matching variables contains all confounders, the marginal association can be interpreted as a causal effect in a sub-population where the matching variables are distributed as among the exposed subjects (Sjölander and Greenland, 2013). The standard errors will have to be adjusted for the correlation within data due to clustering. This is typically done by a sandwich formula or bootstrapping.

In general, the conditional association and the marginal association do not coincide, even when both the conditional and the marginal model are

correctly specified for the problem. This is often referred to as the "non-collapsability" of effect measures (Greenland et al., 1999). One exception is the common linear regression model.

For survival outcomes it is common to model the association between the exposure and outcome as a proportional hazards model, in which exposure and other covariates are assumed to act multiplicatively on the hazard rate and with the associations quantified by logarithms of hazard ratios. Throughout, we let β denote the log hazard ratio of the exposed to the unexposed, conditional on the cluster. We let β^* denote the marginal (over clusters) log hazard ratio of the exposed to the unexposed.

Several modifications of Cox's original proportional hazards model (Cox, 1972) have been proposed to account for clustered data. Some of these models estimate β and some estimate β^*. Both of these types of models may be parametric or semi-parametric.

In the stratified Cox regression model, introduced by Holt and Prentice (1974), the effect of belonging to a certain cluster (the cluster effect) is modeled with a separate baseline hazard for each cluster which is cancelled out using a separate conditional likelihood for each cluster.

A common alternative is the family of frailty models, where the clustering is accounted for by a random effect (Clayton, 1978; Oakes, 1982, 1986; Hougaard, 1986). This random effect is modeled parametrically and is mostly assumed to act multiplicatively on a baseline hazard. Clusters with more "frail" individuals will have a higher hazard, modeled as a higher value of the random effect. For this reason, the random effect is commonly denoted "frailty". Frailty models may be fully parametric (modelling the baseline hazard) or semi-parametric (leaving the baseline hazard unspecified). The most common versions of the frailty models will target β like the stratified Cox model. In addition, the frailty models also give us a measure of the effect of the cluster and to test if there really is a cluster effect.

Another approach is to target β^*, assuming proportional marginal hazards, and adjust the standard errors to account for the correlations induced by the clustering by a sandwich formula. β^* may be estimated semi-parametrically as in Cox original model (Lee et al., 1992) or parametrically (Huster et al., 1989).

The advantages and disadvantages of each of these two approaches have been discussed in many articles. However, they are not comparable except in special cases, since the assumption of proportionality of marginal hazards and the assumption of proportionality of conditional hazards are not equivalent.

Therefore, models estimating β may be misspecified for data where models estimating β^* are correctly specified, and vice versa. Even when both types of models are correctly specified, β and β^* are generally different, due to the non-collapsability of the hazard ratio (Greenland, 1996).

Several studies have compared different survival models for clustered data. Lorino et al. (2004) compared marginal models with semi-parametric frailty models in terms of power and robustness for clustered data. The frailty distribution was modeled both as gamma and log-normal. Manatunga and Oakes (1999) compared the stratified Cox model, the parametric frailty model and the marginal Cox model in terms of efficiency. By simulating from the stable-Weibull frailty model proposed by Hougaard (1986), all models were correctly specified, so comparisons between marginal and conditional effect estimates were fair.

In this report we will provide a detailed comparison between the stratified Cox model, the stable-Weibull frailty model, the marginal Cox model and a marginal Weibull model in terms of underlying assumption and their implications for inference. Using simulations, we will also compare the models in terms of power when both models are correctly specified. Further, we will compare the models in terms of performance under model misspecification.

2 Survival models for clustered data

Suppose we observe a cohort of individuals in N clusters, each of size n. Let T_{ij} and X_{ij} denote the survival time and exposure respectively for the jth individual in cluster i. The survival function for $T_{ij}|X_{ij}$ is defined as

$$S(t_{ij}|X_{ij}) = P(T_{ij} > t_{ij}|X_{ij})$$

and the hazard function for $T_{ij}|X_{ij}$ is defined as

$$h(t_{ij}|X_{ij}) = \lim_{d\to 0} \frac{P(t_{ij} \leq T_{ij} < t_{ij} + d|X_{ij}, T_{ij} \geq t_{ij})}{d}$$

Suppose that we have right censoring. We model the censoring with a variable C_{ij}, denoting the time at which individual j in cluster i is censored. Define the failure indicator as $\Delta_{ij} = I(T_{ij} \leq C_{ij})$ and let $V_{ij} = min(T_{ij}, C_{ij})$. What we actually observe are the realisations $(v_{ij}, \delta_{ij}, x_{ij})$ of the random vectors $(V_{ij}, \Delta_{ij}, X_{ij})$. Note that this model is counterfactual, since a failure (or death) and censoring cannot occur for the same individual. Thus

the random vectors (T_{ij}, C_{ij}) cannot be observed. The lifetimes T_{ij} can be considered to be hypothetical quantities that would all be realized if no censoring had occured. Under certain conditions the parameters indexing the distributions for the lifetimes T_{ij} are still identifiable from the observerved vectors $(v_{ij}, \delta_{ij}, x_{ij})$. We will discuss this further in sections 2.1.1 and 2.2.1 below.

It will be convenient to use boldface symbols to denote random vectors. We will use $\boldsymbol{T}_i$, $\boldsymbol{C}_i$, and $\boldsymbol{X}_i$ to denote the vectors $(T_{i1}, ..., T_{in})$, $(C_{i1}, ..., C_{in})$ and $(X_{i1}, ..., X_{in})$ respectively.

2.1 Conditional models

2.1.1 Assumptions

It will be convenient to introduce a scalar random variable Z_i, which is supposed to summarize all observed and unobserved factors (i.e. the set of matching variables) that are shared among the members of cluster i. Conditioning on cluster i is then equivalent with conditioning on Z_i. We will refer to Z_i as the "frailty" of cluster i. Since individuals in the same cluster have the same frailty, their observed outcomes will be dependent (marginally over the frailty). However, we need to assume that there are no dependencies between clusters, that is

$$(\boldsymbol{T}_1, \boldsymbol{C}_1, \boldsymbol{X}_1, Z_1), ..., (\boldsymbol{T}_N, \boldsymbol{C}_N, \boldsymbol{X}_N, Z_N) \text{ are independent} \tag{1}$$

This assumption is generally true for data both matched by design (like 1:k matched cohort designs) and for naturally matched data (like twin data), since any two clusters are typically unrelated.

We will also assume independence within clusters, conditional on the frailty, that is

$$(T_{i1}, C_{i1}, X_{i1})|Z_i, ..., (T_{in}, C_{in}, X_{in})|Z_i \text{ are independent for i=1,...,N} \tag{2}$$

which implies that

$$p(\boldsymbol{T}_i, \boldsymbol{C}_i|\boldsymbol{X}_i, Z_i) = \prod_{j=1}^{n} p(T_{ij}, C_{ij}|X_{ij}, Z_i)$$

for each cluster i. Assumption (2) is generally true for data matched by design, since the frailty (i.e. the set of matching variables) is all that "ties

together" individuals from the same cluster. However, it is not necessarily true for naturally matched data. For instance, if the event in one member of a twin pair affects the event of the other twin, then the survival times are not independent. To be able to identify parameters indexing the distribution for $T_{ij}|X_{ij}, Z_i$, we also need to assume random censoring within clusters

$$T_{ij} \perp C_{ij} | X_{ij}, Z_i \tag{3}$$

This assumption does not necessarily hold. Further, it is not possible to test this assumption from any observed data (Tsiatis, 1975).

Finally, we assume proportionality between the conditional (on Z_i) hazards

$$\frac{h(t_{ij}|X_{ij}, Z_i)}{h(t_{ij'}|X_{ij'}, Z_i)} = e^{\beta(X_{ij} - X_{ij'})} \tag{4}$$

2.1.2 The Stratified Cox model

In the stratified Cox model for clustered data, the hazards in each cluster are modeled separately. The hazards are assumed to have the form

$$h(t_{ij}|X_{ij}, Z_i) = h_{0i}(t_{ij}) e^{\beta X_{ij}}$$

where Z_i is absorbed into the unspecified function $h_{0i}(t)$. This model targets the conditional log hazard ratio β, assuming (4). The assumptions (1)-(3) allow us to estimate the parameter β by maximizing the partial likelihood

$$\begin{aligned} PL(\beta) &= \prod_{i=1}^{N} \prod_{j=1}^{n} \left\{ \frac{h(v_{ij}|X_{ij}, Z_i; \beta)}{\sum_{(i,j') \in r_{ij}} h(v_{ij'}|X_{ij'}, Z_i; \beta)} \right\}^{\delta_{ij}} \\ &= \prod_{i=1}^{N} \prod_{j=1}^{n} \left\{ \frac{h_{0i}(v_{ij}) e^{\beta X_{ij}}}{\sum_{(i,j') \in r_{ij}} h_{0i}(v_{ij}) e^{\beta X_{ij'}}} \right\}^{\delta_{ij}} \\ &= \prod_{i=1}^{N} \prod_{j=1}^{n} \left\{ \frac{e^{\beta X_{ij}}}{\sum_{(i,j') \in r_{ij}} e^{\beta X_{ij'}}} \right\}^{\delta_{ij}} \end{aligned} \tag{5}$$

with respect to β. Here $r_{ij} = \{(i, j') | v_{ij'} \geq v_{ij}\}$ denotes the risk set for t_{ij} in cluster i. Note that (5) does not contain the cluster specific hazards

h_{0i}. Therefore, the estimate $\widehat{\beta}$ does not depend on the Z_i's. For uncensored matched pair data with no ties, Holt and Prentice (1974) derived simple expressions for the estimate and its asymptotic variance. Let T_{i0} and T_{i1} denote lifetimes of unexposed and exposed individuals respectively. Then the partial likelihood estimate can be written

$$\hat{\beta} = log\left(\frac{\hat{p}}{1-\hat{p}}\right)$$

where $\hat{p} = \frac{\sum_{i=1}^{N} I(t_{i1}<t_{i0})}{N}$ is the non-parametric maximum likelihood estimate of the probability $p = P(T_{i1} < T_{i0})$. They also showed that in the stratified Cox model, β is indeed equal to $log\left(\frac{p}{1-p}\right)$. This means that we only need the rank statistics $I(T_{i1} < T_{i0})$ for $i = 1, ..., N$ to estimate β. The asymptotic variance of this estimate is

$$Var(\hat{\beta}) = \frac{(1+e^{\beta})^2}{Ne^{\beta}}$$

Note that this variance only depends on the real value of β and the number of clusters N and not on the nuisance functions h_{0i}. In general, we cannot use these formulas in the presence of censoring, since in this case the distribution of the observed ranks $I(v_{i1} < v_{i0})$ does not coincide with the ranks $I(t_{i1} < t_{i0})$ of the survival times. However, as Holt and Prentice (1974) showed, simple formulas are still possible in the case where the censoring time is the same for both pair members (e.g. Type I right censoring). When both pair members are censored, no contribution is made to the partial likelihood. All other cases will still contribute information about the statistic $I(T_{i1} < T_{i0})$. Let $U = \{k|max(\delta_{k1}, \delta_{k0}) = 1\}$. Then the estimate $\widehat{\beta}$ is the same as in the uncensored case, but with $\hat{p} = \frac{1}{|U|}\sum_{k\in U} I(v_{k1} < v_{k0})$. The asymptotic variance can be written

$$Var(\hat{\beta}) = \frac{(1+e^{\beta})^2}{|U|e^{\beta}} \tag{6}$$

2.1.3 Parametric frailty models

The most common version of the frailty model is the shared frailty extension of the proportional hazards model, where the cluster-specific baseline hazards have the form

$$h(t_{ij}|X_{ij}, Z_i) = Z_i h_0(t_{ij}) e^{\beta X_{ij}}$$

where the Z_i's are assumed to be independent samples of some parametric distribution, most commonly gamma, lognormal or positive stable. For identification purposes, it makes sense to restrict the paramater space of the frailty distribution to one dimension, such that the frailty distribution only depends on one parameter. In frailty models, the conditional hazards are proportional, in line with assumption (4). The conditional survival functions have the form

$$S(t_{ij}|X_{ij}, Z_i) = exp\{-Z_i H_0(t_{ij}) e^{\beta X_{ij}}\} \tag{7}$$

where $H_0(t) = \int_0^t h_0(s)ds$. Note that the frailty model can be seen a special case of the stratified Cox model with $h_{i0}(t) = Z_i h_0(t)$.

The baseline hazard $h_0(t)$ may be left unspecified as in Cox models, giving us a semi-parametric model. The conditional effect β can then be estimated using the EM algorithm (Klein, 1992). If the baseline hazard is modeled as the hazard for some parametric distribution, then we get a fully parametric model. It is then possible to derive the densities marginalized over the frailties. β can then be estimated by maximizing the likelihood based on these densities. We will henceforth focus on fully parametric frailty models.

There are two assumptions that are required by frailty models, but not required by stratified Cox models. These two assumptions are often unstated in the literature. For instance, these assumptions are not mentioned in the standard textbooks by Hougaard (2000) and Duchateau and Janssen (2008). The first assumption is that the exposure is independent of the frailty:

$$\boldsymbol{X}_i \perp Z_i \quad \text{, for } i = 1, ..., N \tag{8}$$

If this assumption is violated, the effect estimate $\widehat{\beta}$ is no longer consistent for β. The assumption (8) is not generally true. For instance, (8) is violated if Z contains confounders. However, in matched cohort studies, (8) holds automatically, since the vector $\boldsymbol{X}_i$ is constant over i. For instance, in 1:1 matched studies, $(X_{i1}, X_{i2}) = (1, 0)$, since the ordering of X_{i1} and X_{i2} is unimportant. In section 3 we will discuss the implications of violations of (8).

The second assumption is that censoring is independent of the frailty conditional on the exposure, that is

$$C_{ij} \perp Z_i | X_{ij} \tag{9}$$

The assumption (9) does not necessarily hold for real data. In twin studies for instance, Z_i represents the shared DNA (amongst other factors). It is not

unreasonable to assume that the shared DNA in a cluster is associated with elevated levels of competing risks represented by $C_{i1}, ..., C_{1n}$.

Together with the assumptions (1)-(3), the assumptions (8) and (9) are required in order to compute the full likelihood based on the observed data. To clarify the roles of the assumptions (8) and (9), we will now derive the likelihood for a fully parametric frailty model. We emphasize that the detailed derivation (starting from observed data) is typically not provided in standard textbooks on frailty models, e.g. Hougaard (2000) and Duchateau and Janssen (2008). Suppose that we have observed the vector $(v_{ij}, \delta_{ij}, x_{ij})$ for $i = 1, ..., n$ and $j = 1, ..., n$. Now

$$V_{ij} = v_{ij}, \Delta_{ij} = \delta_{ij}, X_{ij} = x_{ij}, Z_i = z_i$$

$$\Longleftrightarrow$$

$$\begin{cases} T_{ij} = v_{ij}, C_{ij} > v_{ij}, X_{ij} = x_{ij}, Z_i = z_i & \text{when } \delta_{ij} = 1 \\ T_{ij} > v_{ij}, C_{ij} = v_{ij}, X_{ij} = x_{ij}, Z_i = z_i & \text{when } \delta_{ij} = 0 \end{cases}$$

so the likelihood can be expressed in terms of T_{ij}, C_{ij}, X_{ij}, and Z_i. By assumption (1) the likelihood factorizes to clusterwise contributions. To make notation more reader friendly, we can therefore drop the cluster specific subscripts i. Let k denote the number of observations in a cluster that are not censored. Since the ordering of the observations is arbitrary, we can assume that $\delta_1 = ... = \delta_k = 1$ and $\delta_{k+1} = ... = \delta_n = 0$. Then, conditional on $\boldsymbol{X}$, the clusters contribution to the likelihood is

$$\begin{aligned} &p(T_1 = v_1, ..., T_k = v_k, T_{k+1} > v_{k+1}, ..., T_n > v_n, \\ &\qquad C_1 > v_1, ..., C_k > v_k, C_{k+1} = v_{k+1}, ..., C_n = v_n | \boldsymbol{X}) \\ &= E_{Z|\boldsymbol{X}}\{p(T_1 = v_1, ..., T_k = v_k, T_{k+1} > v_{k+1}, ..., T_n > v_n, \\ &\qquad C_1 > v_1, ..., C_k > v_k, C_{k+1} = v_{k+1}, ..., C_n = v_n | \boldsymbol{X}, Z)\} \end{aligned}$$

By assumption (2) of conditional independence, this equals

$$E_{Z|\boldsymbol{X}} \left\{ \prod_{j=1}^{k} p(T_j = v_1, C_j > v_j | X_j, Z) \prod_{j=k+1}^{n} p(T_j > v_1, C_j = v_j | X_j, Z) \right\} \quad (10)$$

By assumption (3) the lifetimes and censoring times are independent conditional on Z and X. Therefore

$$p(T_j = v_j, C_j > v_j | X_j, Z) = p(T_j = v_j | X_j, Z) p(C_j > v_j | X_j, Z)$$

and

$$p(T_j > v_1, C_j = v_j | X_j, Z) = p(T_j > v_j | X_j, Z) p(C_j = v_j | X_j, Z)$$

By assumption (9), censoring is independent of the frailty, so

$$p(C_j > v_j | X_j, Z) = p(C_j > v_j | X_j)$$

and

$$p(C_j = v_j | X_j, Z) = p(C_j = v_j | X_j)$$

Therefore, (10) equals

$$E_{Z|\boldsymbol{X}} \left\{ \prod_{j=1}^{k} p(T_j = v_j | X_j, Z) \prod_{j=k+1}^{n} p(T_j > v_j | X_j, Z) \right\}$$
$$\prod_{j=1}^{k} p(C_j > v_j | X_j) \prod_{j=k+1}^{n} p(C_j = v_j | X_j) \tag{11}$$

Thus, the censoring times only contributes to the likelihood with a factor that does not dependend on the parameters of the distributions for the lifetimes T_j and the frailty Z. Without assumptions (3) and (9), we would not be able to factorize out the "censoring" part from the expectation in (10).

Assumption (8) of independence between $\boldsymbol{X}$ and Z now implies that we can use the unconditional distribution of Z when integrating out Z, that is

$$E_{Z|\boldsymbol{X}} \left\{ \prod_{j=1}^{k} p(T_j = v_j | X_j, Z) \prod_{j=k+1}^{n} p(T_j > v_j | X_j, Z) \right\}$$
$$= E_Z \left\{ \prod_{j=1}^{k} p(T_j = v_j | X_j, Z) \prod_{j=k+1}^{n} p(T_j > v_j | X_j, Z) \right\} \tag{12}$$

If the distribution of $Z|\boldsymbol{X}$ is unknown, we will not be able to compute the likelihood marginalized over $Z|\boldsymbol{X}$. The assumption (8) means that $Z|\boldsymbol{X}$ have the same distribution as Z.

Since

$$p(T_j > v_j|X_j, Z) = S(t_j|X_j, Z)$$

and

$$p(T_j = v_j|X_j, Z) = -\frac{\partial}{\partial t_j} S(t_j|X_j, Z)$$

we can rewrite (12) in terms of the conditional survival functions:

$$E_Z \left\{ (-1)^k \frac{\partial^k}{\partial v_1, ..., v_k} \prod_{j=1}^{n} S(v_j|X_j, Z) \right\}$$

By switching integration and derivation we then get

$$E_Z \left\{ (-1)^k \frac{\partial^k}{\partial v_1, ..., v_k} \prod_{j=1}^{n} S(v_j|X_j, Z) \right\} = (-1)^k \frac{\partial^k}{\partial v_1, ..., v_k} E_Z \left\{ \prod_{j=1}^{n} S(v_j|X_j, Z) \right\}$$

Now,

$$\begin{aligned}
&E_Z \left[\prod_{j=1}^{n} S(v_j|X_j, Z) \right] \\
&= E_Z \left[\prod_{j=1}^{n} exp\{-Z H_0(v_j) e^{\beta X_j}\} \right] \\
&= E_Z \left[exp\{-Z \sum_{j=1}^{n} H_0(v_j) e^{\beta X_j}\} \right] \\
&= L_Z \left\{ \sum_{j=1}^{n} H_0(v_j) e^{\beta X_j} \right\}
\end{aligned}$$

where L_Z is the the Laplace transform of the density of Z. Thus the likelihood

contribution is

$$(-1)^k \frac{\partial^k}{\partial v_1, ..., v_k} L_Z \left\{ \sum_{j=1}^{n} H_0(v_j) e^{\beta X_j} \right\}$$
$$= (-1)^k \prod_{j=1}^{k} \left\{ h_0(v_j) e^{\beta X_j} \right\} L_Z^{(k)} \left\{ \sum_{j=1}^{n} H_0(v_j) e^{\beta X_j} \right\}$$

Reintroducing the cluster subscripts i and letting $D_i = \sum_{j=1}^{n} \delta_{ij}$, we can now write the full likelihood as

$$L(\beta) = \prod_{i=1}^{N} \left[(-1)^{D_i} \left\{ \prod_{j=1}^{n} h(v_{ij}|X_{ij})^{\delta_{ij}} \right\} L_Z^{(D_i)} \left\{ \sum_{j=1}^{n} H(v_{ij}|X_{ij}) \right\} \right]$$

2.1.4 The stable-Weibull model

In this report we will use a parametric frailty model with Weibull baseline hazards and positive stable frailties (Hougaard, 1986). The reason is that in this model, both the conditional hazards and the marginal hazards are proportional. In the stable-Weibull model, the baseline hazard is modeled parametrically as a Weibull hazard:

$$h_0(t) = \lambda_0 c t^{c-1}$$

This gives us the cumulative hazards

$$H(t|X) = \lambda_0 e^{\beta X} t^c$$

The frailty is assumed to belong to the family of positive stable distributions and has density

$$f_Z(z) = -\frac{1}{\pi z} \sum_{k=1}^{\infty} \frac{\Gamma(k\alpha + 1)}{k!} (-z^{-\alpha})^k sin(k\alpha\pi)$$

where $0 < \alpha < 1$. Fortunately, the Laplace transform has a simpler expression: $L(s) = e^{-s^\alpha}$. Therefore, the likelihood from each cluster with k uncensored observations is

$$(-1)^k \frac{\partial^k}{\partial v_1, ..., v_k} L_Z \left\{ \sum_{j=1}^{n} H_0(v_j) e^{\beta X_j} \right\}$$
$$= (-1)^k \frac{\partial^k}{\partial v_1, ..., v_k} exp \left\{ - \left(\sum_{j=1}^{n} \lambda_0 e^{\beta X_j} v_j^c \right)^\alpha \right\}$$

2.2 Marginal models

In marginal models for clustered data, the estimation of β^* is based on the marginal (over Z) density of the lifetimes. Standard errors are then corrected for the correlations due to clustering in data. In the parametric case, this means that the likelihood is based on the marginal distributions. In the semi-parametric case this means that we use the partial likelihood from the ordinary Cox regression model. It can be shown that this procedure gives consistent estimates of β^* both in the parametric case (Huster et al., 1989) and in the semi-parametric case (Lee et al., 1992). The standard errors are then corrected for the correlations within the clusters, using a sandwich formula or a grouped jackknife procedure.

2.2.1 Assumptions

Since marginal models does not include a random component Z different assumptions are needed for marginal models. We will still need to assume independence between clusters, that is

$$(\boldsymbol{T}_1, \boldsymbol{C}_1, \boldsymbol{X}_1), ..., (\boldsymbol{T}_N, \boldsymbol{C}_N, \boldsymbol{X}_N) \text{ are independent} \tag{13}$$

This assumption actually follows from (1).

We also need to assume that independent censoring for each individual, that is

$$T_{ij} \perp C_{ij} | X_{ij} \tag{14}$$

We will also need to assume proportional marginal hazards, that is

$$\frac{h(t|X)}{h(t|X')} = e^{\beta(X-X')} \tag{15}$$

Note that the assumption (15) is neither implied by nor implies the assumption (4) of conditional proportionality.

We don't need to assume that $\boldsymbol{X}$ is independent of cluster membership for valid inference. However, violations of this assumption will have consequences for the interpretation of the estimate. This will be further discussed in section 3.

2.2.2 The marginal Cox model

The marginal Cox model specifies that

$$h(t_{ij}|X_{ij}) = h_0(t_{ij})e^{\beta^* X_{ij}}$$

where h_0 is left unspecified. β^* is then estimated by maximizing the partial likelihood

$$\begin{aligned} PL(\beta^*) &= \prod_{i=1}^{N}\prod_{j=1}^{n}\left(\frac{h(v_{ij}|X_{ij})}{\sum_{v_{i',j'}\geq v_{ij}} h(v_{i',j'}|X_{i',j'})}\right)^{\delta_{ij}} \\ &= \prod_{i=1}^{N}\prod_{j=1}^{n}\left(\frac{e^{\beta^* X_{ij}}}{\sum_{v_{i',j'}\geq v_{ij}} e^{\beta^* X_{i',j'}}}\right)^{\delta_{ij}} \end{aligned} \tag{16}$$

Note that this is not a proper partial likelihood, since it ignores the correlations due to clustering. However, Lee et al. (1992) showed that the estimator based on (16) is consistent for β^* provided that the marginal distributions are correctly specified. They also derived an sandwich estimator to correct the standard errors for the correlations due to clustering.

2.2.3 The marginal Weibull model

Parametric marginal modelling only requires a specification of the marginal survival times along with information of cluster membership. The marginal survival times can also be modeled as following some parametric distribution, most often exponential, Weibull, normal or lognormal. A marginal parametric model may be derived from a parametric frailty model by taking the expectation over the conditional survival functions:

$$\begin{aligned} &S(t|X) \\ &= E_{Z|X}\left[S(t|X,Z)\right] \\ &= E_Z\left[S(t|X,Z)\right] \\ &= E_Z\left[exp\left\{-ZH_0(t)e^{\beta X}\right\}\right] \\ &= L_Z\left\{H_0(t)e^{\beta X}\right\} \end{aligned}$$

Then the marginal hazard functions are

$$h(t|X) = \frac{\partial}{\partial t}\left[-log\{S(t|X)\}\right]$$

Note that the marginal hazard functions derived from a parametric frailty model need not be proportional. An exception is the stable-Weibull model, where the marginal survival functions are

$$exp\left(-\lambda_0^{\alpha} e^{\alpha\beta X} t^{\alpha c}\right)$$

with marginal hazards

$$h(t|X) = \lambda_0^{\alpha} e^{\alpha\beta X} \alpha c t^{\alpha c - 1}$$

This corresponds to Weibull distributed survival times with scale parameters $\lambda_0^{\alpha} e^{\alpha\beta X}$ and shape parameter αc. The log hazard ratio is $\beta^* = \alpha\beta$, which is the same parameter as the one targeted in the marginal Cox model.

We will now derive the likelihood for individual j in cluster i, based on the marginal density under the Weibull model. We have that

$$\begin{aligned}
&p(V_{ij} = v_{ij}, \Delta_{ij} = \delta_{ij}|X_{ij}) \\
&= \{p(T_{ij} = v_{ij}, C_{ij} > v_{ij}|X_{ij})\}^{\delta_{ij}} \{p(T_{ij} > v_{ij}, C_{ij} = v_{ij}|X_{ij})\}^{1-\delta_{ij}} \\
&= \{p(T_{ij} = v_{ij}|X_{ij})\}^{\delta_{ij}} \{p(T_{ij} > v_{ij}|X_{ij})\}^{1-\delta_{ij}} \\
&\qquad \{p(C_{ij} > v_{ij}|X_{ij})\}^{\delta_{ij}} \{p(C_{ij} = v_{ij}|X_{ij})\}^{1-\delta_{ij}}
\end{aligned}$$

where the last equality follows from the assumption (14) of random censoring. Multiplying over all clusters and individuals, and ignoring the terms involving C_{ij}, gives

$$\begin{aligned}
L(\beta) &= \prod_{i=1}^{N} \prod_{j=1}^{n} \{p(T_{ij} = v_{ij}|X_{ij})\}^{\delta_{ij}} \{p(T_{ij} > v_{ij}|X_{ij})\}^{1-\delta_{ij}} \\
&= \prod_{i=1}^{N} \prod_{j=1}^{n} f(v_{ij}|X_{ij})^{\delta_{ij}} S(v_{ij}|X_{ij})^{1-\delta_{ij}} \\
&= \prod_{i=1}^{N} \prod_{j=1}^{n} h(v_{ij}|X_{ij})^{\delta_{ij}} S(v_{ij}|X_{ij}) \\
&= \prod_{i=1}^{N} \prod_{j=1}^{n} \left(\lambda_0^{\alpha} e^{\alpha\beta X_{ij}} \alpha c t_{ij}^{\alpha c - 1}\right)^{\delta_{ij}} exp\left(-\lambda_0^{\alpha} e^{\alpha\beta X_{ij}} t_{ij}^{\alpha c}\right) \qquad (17)
\end{aligned}$$

The parameters α, β, λ_0 and c are not identifiable by maximizing this expression. To see this, suppose that α, β, λ_0 and c maximizes the expression. Then, for any $p > 0$, the parameter combination $\alpha_p = p\alpha$, $\beta_p = \beta/p$, $\lambda_p = (\lambda_0)^{(1/p)}$ and $c_p = c/p$ will also maximize the expression. Letting $\beta^* = \alpha\beta$, $\lambda^* = \lambda_0^\alpha$ and $c^* = \alpha c$, we can rewrite the the product (17) as

$$\prod_{i=1}^{N}\prod_{j=1}^{n}\left(\lambda^* e^{\beta^* X_{ij}} c^* t_{ij}^{c^*-1}\right)^{\delta_{ij}} exp\left(-\lambda^* e^{\beta^* X_{ij}} t_{ij}^{c^*}\right)$$

which is a likelihood for independent Weibull distributed survival times. From this product, the parameters β^*, λ^* and c^* are identifiable. The unidentifiability of α from the marginal distributions was a motivation behind the stable-Weibull model. If the dependence parameter is identifiable from the marginal distributions (as in other common frailty models), then it measures something more besides dependence, which is an undesirable property (Hougaard, 1986).

We note that (17) is not a proper likelihood for the data, since it ignores the clustering. However, as Huster et al. (1989) have demonstrated, maximizing this expression with respect to $(\lambda^*, \beta^*, c^*)$ will yield a consistent estimate of β^*.

3 Matched data and general clustered data

For data that is matched on the exposure, the exposure vector is fixed over the clusters. Therefore there can be no dependence between the cluster effect Z and the exposure vector $\boldsymbol{X}$, meaning that assumption (8) is always fulfilled for matched data. For general data, there may be dependence between the frailty Z and the exposure $\boldsymbol{X}$. Therefore, the association between the exposure and survival is confounded by Z. This will have different consequences for different models.

For instance, the parametric frailty models are misspecified when (8) is violated, since the equality (12) does not hold. Ignoring violations of (8) may therefore lead to biased effect estimates when applying the frailty model (Sjölander et al., 2013). A possible workaround to the problem might be to use a within-between decomposition of the covariate effect, in analogy with a method for generalized linear mixed models proposed by Neuhaus and Kalbfleisch (1998). For generalized linear mixed models, they decomposed βX_j into the components $\beta_{between}\bar{X}$ and $\beta_{within}(\bar{X} - X_j)$, where $\bar{X} =$

$\frac{1}{n}\sum_{j=1}^{n} X_j$. They showed that the estimate $\widehat{\beta_{within}}$ consistently estimates the conditional effect β, even when there was dependence between the cluster effect Z and the covariates $X_1, ..., X_n$. However, few attempts have been made to extend this method to models for survival data.

For marginal methods, violation of (8) will still lead to valid estimates in the sense that the estimate measures the population averaged association. However, one may no longer be able to give a causal interpretation of this association, since it is confounded by Z.

The stratified Cox model does not rely on assumption (8). In the stratified Cox model, the effect β is calculated conditionally on Z. Since Z is canceled out in the conditional likelihood, the stratified Cox model will still estimate an association that is unconfounded by Z (Sjölander et al., 2013). However, the efficiency gains of using general data instead of matched data with the stratified Cox model may be small. For twin data with one binary exposure and no other covariates, pairs concordant on the exposure will contribute to the partial likelihood by a constant. Therefore, the estimate and the variance will be the same as when only matched pair data are used. With other covariates in the model, some increase in efficiency might be achieved.

4 An important special case of parametric frailty models

As the distribution of Z becomes degenerate at some point $c > 0$, that is when $Z = c$ for some constant c, the multivariate survival function becomes

$$S(t_1, ..., t_n|X_1, ..., X_n, Z = c) = exp\left(-c\sum_{j=1}^{n} H(t_j|X_j)\right)$$

which factorizes, implying marginal independence between the lifetimes. In general, this corresponds to the case when the frailty variance approaches 0. The positive stable distribution used here have no finite mean or variance, but is degenerate for $\alpha = 1$ at 1 (Hougaard, 2000), meaning that $c = 1$. This also means that the marginal hazards

$$h(t|X) = ch_0(t)e^{\beta X}$$

are proportional. From the form of the marginal hazard function we can also see that the the conditional and marginal associations under the stable-

Weibull model coincide, that is $\beta = \beta^*$.

5 Simulations

Four estimation methods were compared: A fully parametric stable-Weibull model, a stratified Cox model, a marginal Cox model and a marginal Weibull model. These methods were compared in terms of statistical power (section 5.1) and performance under model misspecification (section 5.2).

The simulations were performed with varying amount dependence between individuals in the same cluster. The dependence was measured with Kendall's τ, defined as

$$\tau = E[sign\{(T_{ij} - T_{i'j})(T_{ij'} - T_{i'j'})\}]$$

where $i \neq i'$ and $j \neq j'$. This dependence measure is similar to the Pearson correlation in that it has the range $[-1, 1]$ and that $T_{ij} \perp T_{ij'}$ for $j \neq j'$ implies that $\tau = 0$. However, Kendall's τ makes less assumptions about the dependence structure than the Pearson correlation. Like the Cox regression methods, this measure is based on the ranks and not on the actual values of the observations. In frailty models, τ can be calculated from the parameter of the frailty distribution. As we will only look at positive dependence, Kendall's τ will be non-negative.

Uniform censoring was used in all censoring, with parameter chosen to achieve the desired amount of censoring. For example, suppose the marginal univariate survival function is $S(t|X)$. If we want to achieve a censored proportion of p_c when generating clusters with 1 exposed and k unexposed with $C \sim U(0, c_{max})$, then we can find the parameter c_{max} of the censoring distribution by solving the equation

$$p_m = P(C < T) = \int_0^{c_{max}} \frac{1}{c_{max}} \left\{ \frac{1}{k} S(t|X = 1) + \frac{k-1}{k} S(t|X = 0) \right\} dt$$

for c_{max}

In all simulations, data were generated under the key assumptions (1), (2), (3), (4), (8) and (9).

The software package `R` version 2.15 was used to simulate the data. For estimation from the stable-Weibull method, package `parfm` version 2.02 was

used. For the stratified Cox, the marginal Cox and the marginal Weibull methods the package `survival` version 2.36 was used. The standard errors for the marginal Weibull estimates was calculated using a 'sandwich' estimator. For the marginal Cox method, standard error was calculated using a grouped jackknife procedure, which is asymptotically equivalent (Lipsitz and Parzen, 1996) to the sandwich estimator proposed by Lee et al. (1992).

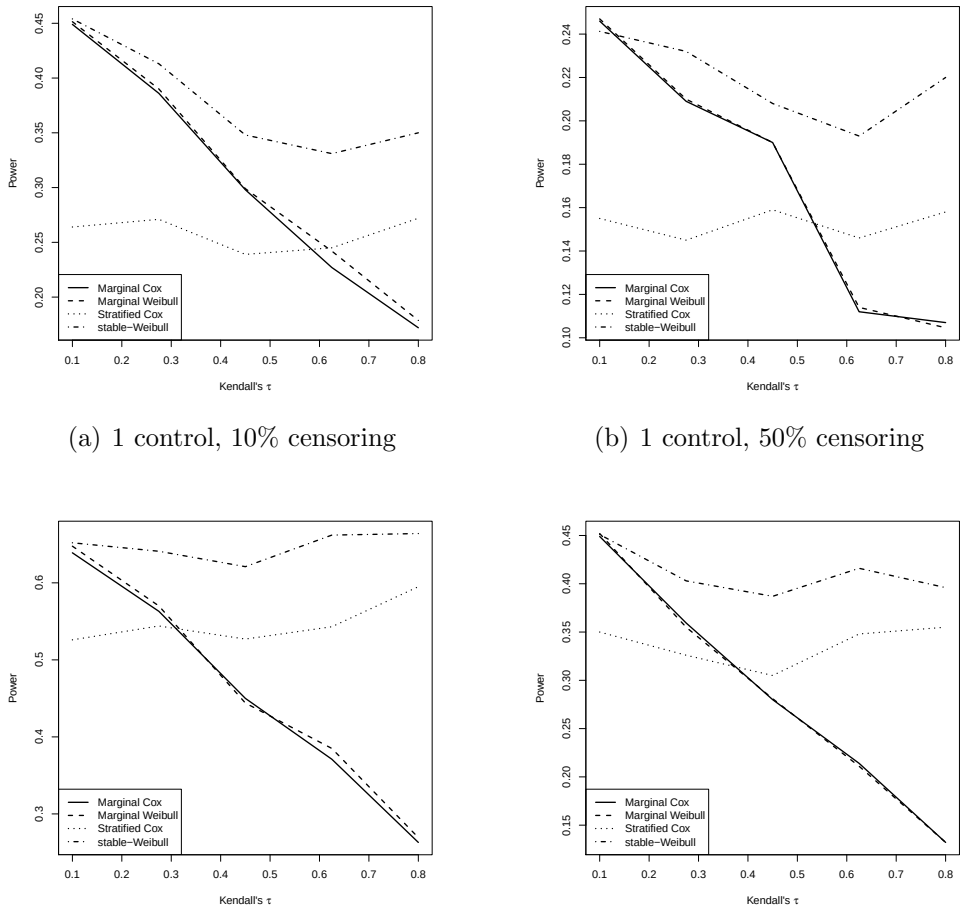
5.1 Statistical power

As both conditional and marginal models are correctly specified under the stable-Weibull model, we simulated data from this model to compare the models in terms of statistical power. Although the methods target different parameters (β and β^*), they are comparable in terms of power when testing the hypothesis $H_0 : \beta = 0$ since $\beta = 0$ iff $\beta^* = 0$.

For values of τ between 0.1 and 0.8, lifetimes of 200 clusters were generated 1000 times from to the model

$$\begin{aligned} \alpha &= 1 - \tau \\ Z_i &\sim Ps(\alpha_\tau) \\ T_{ij}|X_{ij}, Z_i &\sim \text{ Weibull}(Z_i \lambda_0 e^{\beta X_{ij}}, c) \end{aligned}$$

The baseline hazard scale and shape were set to $\lambda_0 = 0.75$ and $c = 2$ respectively. To study the impact of different cluster sizes, both 1:1 and 1:5 matched cohort data were generated. Censoring was either 10 % or 50 %. In all simulations β was set to 0.2. For each value of τ, and each combination of cluster size and censoring fraction, the empirical power was computed; i.e. the probability (over 1000 samples) to reject the hypothesis $H_0 : \beta = 0$, using a Wald test at 5 % significance level. The results are shown in Figure 1.



(a) 1 control, 10% censoring

(b) 1 control, 50% censoring

(c) 5 controls, 10% censoring

(d) 5 controls, 50% censoring

Figure 1: Power rejecting the hypothesis $H_0 : \beta = 0$ when $\beta = 0.2$ under the stable-Weibull model

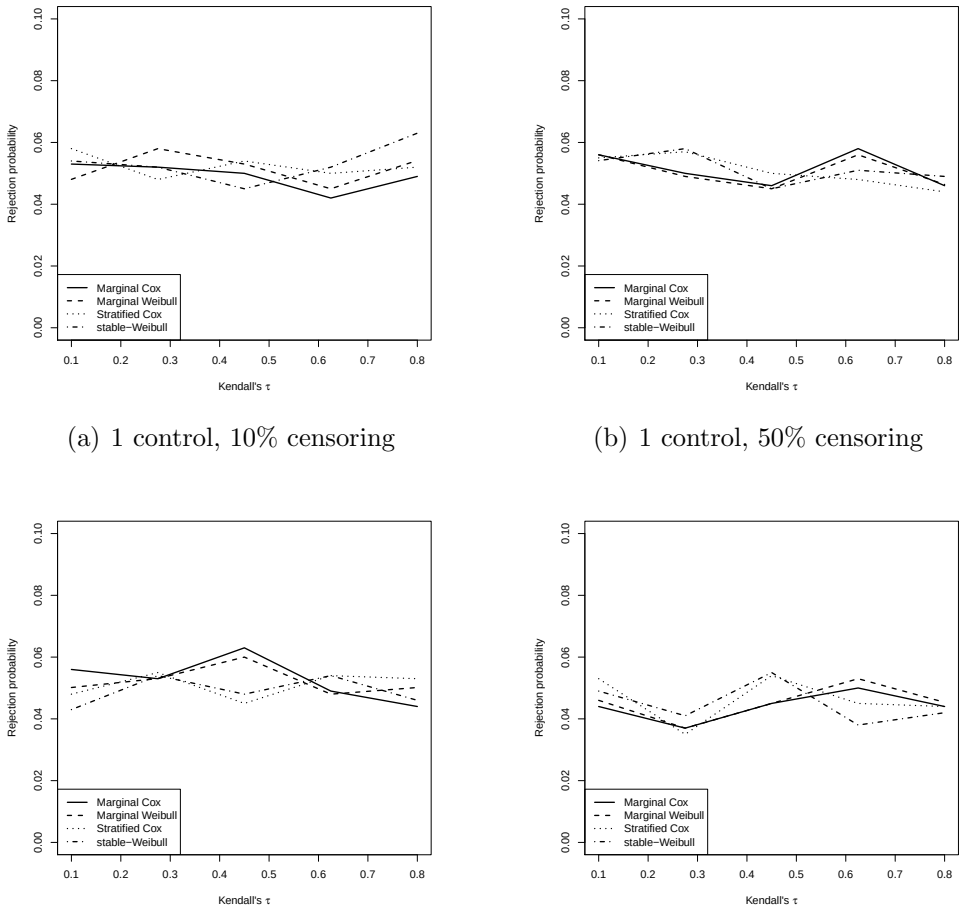
The power of the stable-Weibull model was the highest for all models. This was expected, since this is the fully parametric model for data. The power of stratified Cox model was substantially smaller in all scenarios. The power of the stratified Cox model was stable over τ in all four scenarios (a)-(b). As we saw in section 2.1.2, neither the estimate $\widehat{\beta}$ nor the variance $Var(\widehat{\beta})$ for the stratified Cox model depend on the amount of dependence due to the clustering when we have uncensored matched pairs data. Figure 1 suggests that both the estimate and its variance is also insensitive to this

dependence when clusters a larger and when data are censored.

In contrast, the power of the marginal Cox model was very similar to the fully parametric marginal Weibull model. As the dependence approaches 0, the marginal methods approached the same power as the stable-Weibull model. This was expected from theory, as discussed in section 4. As the dependence increased, the power of the marginal models decreased substantially.

In all scenarios, the power of the marginal models was smaller compared to the stratified Cox model for higher values of τ with lines intersecting close to some value of τ. This value of τ was less influenced by the amount of censoring than by the cluster size. It appears that the stratified Cox model gains relatively more power when the cluster size is increased.

The same simulations were repeated with $\beta = 0$. By construction of the statistical test, the (asymptotic) "power" should in this case equal the significance level, i.e. 5 %. As seen in Figure 2, all models have a power close to 5 % for all values of τ in all scenarios.



(a) 1 control, 10% censoring

(b) 1 control, 50% censoring

(c) 5 controls, 10% censoring

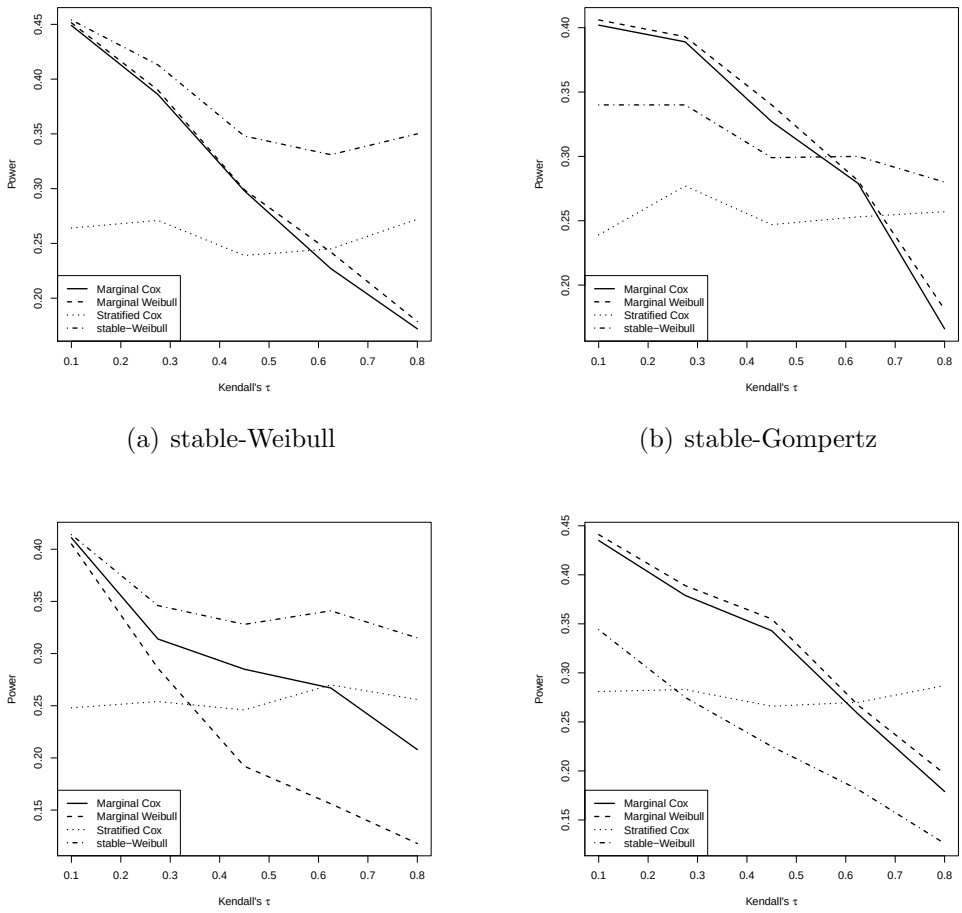
(d) 5 controls, 50% censoring

Figure 2: Probability rejecting the hypothesis $H_0 : \beta = 0$ when $\beta = 0$ under the stable-Weibull model

5.2 Performance under model misspecification

To compare the stable-Weibull model, the stratified Cox model, the marginal Cox model and the Weibull model in terms of performance under model misspecification, data were generated from four different models - a stable-Weibull model (included as a comparison), a stable-Gompertz model, a gamma-Weibull model and a gamma-Gompertz model (see the appendix for

a description of these models). For the Weibull baseline hazard, the scale and shape was $\lambda_0 = 2$ and $c = 2$ respectively. For the Gompertz baseline hazards, the scale and shape was set to $b = 0.75$ and $\eta = 2$ respectively. β was set to 0.2 for all models. For each model, Kendall's τ was varied between 0.1 and 0.8. For each value of τ, 1000 samples of 200 1:1 matched pairs were simulated with 10 % uniform censoring. As in section 5.1 the empirical power was calculated. The results are shown in Figure 3.



(a) stable-Weibull

(b) stable-Gompertz

(c) gamma-Weibull

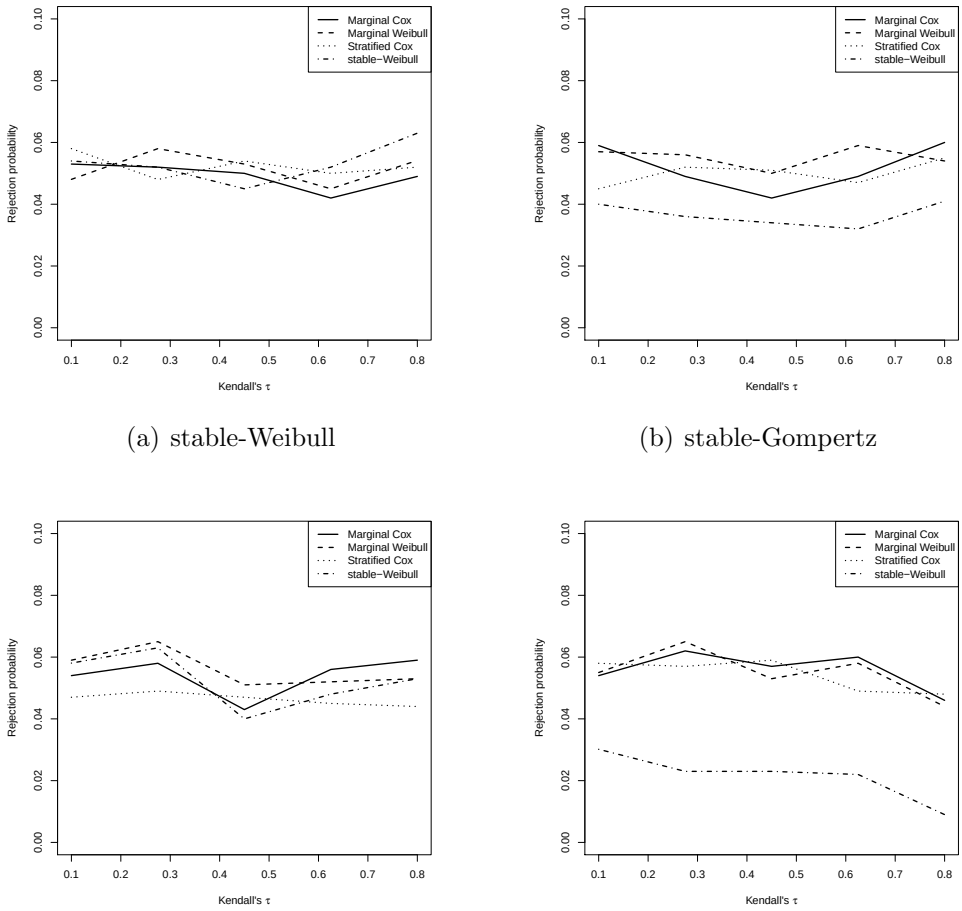
(d) gamma-Gompertz

Figure 3: Power rejecting the hypothesis $H_0 : \beta = 0$ when $\beta = 0.2$ under different frailty-models

Under the stable-Weibull model all models were correctly specified. When data were generated from other frailty models, only the stratified Cox model was correctly specified. In spite of this, the power for the stratified Cox model was relatively low in all scenarios. The stable-Weibull performed well when only the frailty was misspecified under the gamma-Weibull model, but appeared to be sensitive to misspecification of the baseline hazard in the sense that the power was low when when the baseline hazard was misspecified.

The marginal models performed very similarly under the stable-Gompertz model and the gamma-Gompertz model. Under the gamma-Weibull model the power for the marginal Weibull model decreased rapidly with increasing τ. Compared to the stable-Weibull model, the marginal methods performed well when the baseline hazard was misspecified for the stable-Weibull model.

Next, the simulations were repeated with $\beta = 0$ and the empirical "power" of testing $H_0 : \beta = 0$ was computed as in section 5.1. The simulation results are shown in Figure 4.



(a) stable-Weibull

(b) stable-Gompertz

(c) gamma-Weibull

(d) gamma-Gompertz

Figure 4: Probability rejecting the hypothesis $H_0 : \beta = 0$ when $\beta = 0$ under different frailty models

The simulation under the stable-Weibull model from Figure 2(a), was included here as a comparison. Under the other frailty models, only the semiparametric models are correctly specified, since both conditional and marginal models are proportional. In spite of this, the parametric Weibull model worked well in all scenarios (power around 5 %). For models with Gompertz baseline hazards, the power for the stable-Weibull method was smaller than 5 % from moderate and large values of τ.

6 Discussion

During the last two decades, models for clustered survival data have generated a lot of interest. However, few authors have focused on the consequences of model choice on causal inference. In this report, we have compared the stratified Cox model, parametric frailty models, the marginal Cox model and marginal parametric models in terms of their underlying assumptions. In particular, we demonstrated that parametric frailty models requires stricter assumptions than the stratified Cox model regarding independence between the frailty and exposure. A previous simulation study have shown that the gamma-Weibull frailty model gives biased estimates of β when (8) is violated compared to estimates from stratified Cox (Sjölander et al., 2013). There are no reasons to assume that this is unique for the gamma-Weibull model, but also extends to other frailty models. The stratified Cox model does not require this assumption, which may often be violated for real data. Further, the stratified Cox model makes less strict assumptions regarding the independence between censoring and the frailty. The stratified Cox model is thus more robust than the parametric frailty model for general data, and may therefore be preferable for general unmatched data.

With matched data, the assumption (8) is automatically fulfilled, so a causal interpretation of the frailty estimate $\widehat{\beta}$ is more plausible. By choosing a frailty model (parametric of semi-parametric) over a stratified Cox model, gains in power can be expected. As Wild (1983) noted, the stratified Cox model can be quite inefficient when compared to the gamma-Weibull frailty model, since it does not utilize between-cluster information. The results in section 5.1 showed considerable loss of power when the stratified Cox was compared to the stable-Weibull model.

In applications, the marginal Cox model still dominates. Marginal models does not require the assumption (8), and the estimated log hazard ratios can be interpreted as population averaged association. As discussed in section 3, violations of (8) will have consequences for causal interpretation of the association. If the cluster membership is both associated with exposure and the outcome, then the association is confounded and cannot be given a causal interpretation. However, when data are matched on all confounders, the marginal association can have a causal interpretation. In section 5.1 we saw that marginal models have good power relative to the fully parametric stable-Weibull model when the dependence is small, but performs badly when the dependence is larger.

As Efron (1977) have demonstrated, the Cox regression estimate is asymptotically fully efficient for independent Weibull distributed data. The results in section 5.1 suggest that this result also translates to clustered data, if we by "fully" efficient mean the efficiency that can be achieved by the right specification of the marginal distribution.

Previous authors have suggested that, since a frailty is often present in real life data, marginal models are more often misspecified which might lead to substantial bias for the estimates (Henderson and Oman, 1999). On the other hand, inference from frailty models requires more assumptions regarding the dependence structure. In frailty models we assume a parametric form for the frailty and we also assume that it acts multiplicatively on a baseline hazard. The semiparametric gamma frailty have been found to be robust (Glidden and Vittinghoff, 2004). In section 5.2 we found that the stable-Weibull model performed well (in terms of high power) when the frailty distribution was misspecified. However, the performance was sensitive to misspecification of the baseline hazard. This suggests that a semi-parametric frailty model, with positive stable distributed frailty might perform better in this scenario.

In this report we have focused on measures of association, that is the marginal and the conditional log hazard ratios. The frailty parameter have therefore been treated as a nuisance parameter. If the intra-cluster dependence is of interest, frailty models gives a measure of this dependence through the frailty parameter. In this situation, the positive stable distribution may be preferable, since it is the only frailty distribution the parameter of which is not identifiable from the marginal distribution and thus does not measure anything else besides dependence (Hougaard, 1986).

7 Conclusions

The stratified Cox model makes few model assumptions and is thus very robust. However, compared to other methods, the stratified Cox have relatively poor performance in terms of power, especially when the cluster size is small and when the dependence is small.

The stable-Weibull model makes many assumptions and may give biased results when there is an association between exposure and dependence. When data are matched, the stable-Weibull performs well in terms of power. It is robust against model misspecification, although it gives lower than expected

power when the baseline hazard is misspecified.

No substantial improvements in terms of power were seen for the marginal Weibull model when it was compared to the marginal Cox model. The marginal Cox model performed well for small dependence, but the power decreased with increasing dependence. The marginal Cox model performed similarly under model misspecification.

References

Clayton, D. (1978). A model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence. *Biometrika* **65,** 141–151.

Cox, D. (1972). Regression models and life-tables. *Journal of the Royal Statistical Society. Series B (Methodological)* pages 187–220.

Duchateau, L. and Janssen, P. (2008). *The frailty model*. Springer Verlag.

Efron, B. (1977). The efficiency of cox's likelihood function for censored data. *Journal of the American statistical Association* **72,** 557–565.

Glidden, D. and Vittinghoff, E. (2004). Modelling clustered survival data from multicentre clinical trials. *Statistics in Medicine* **23,** 369–388.

Greenland, S. (1996). Absence of confounding does not correspond to collapsibility of the rate ratio or rate difference. *Epidemiology* pages 498–501.

Greenland, S., Robins, J., and Pearl, J. (1999). Confounding and collapsibility in causal inference. *Statistical Science* **14,** 29–46.

Henderson, R. and Oman, P. (1999). Effect of frailty on marginal regression estimates in survival analysis. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **61,** 367–379.

Holt, J. and Prentice, R. (1974). Survival analyses in twin studies and matched pair experiments. *Biometrika* **61,** 17–30.

Hougaard, P. (1986). A class of multivanate failure time distributions. *Biometrika* **73,** 671–678.

Hougaard, P. (2000). *Analysis of Multivariate Survival Data*. Springer, New York.

Huster, W., Brookmeyer, R., and Self, S. (1989). Modelling paired survival data with covariates. *Biometrics* pages 145–156.

Kendall, M. (1938). A new measure of rank correlation. *Biometrika* **30,** 81–93.

Klein, J. (1992). Semiparametric estimation of random effects using the cox model based on the em algorithm. *Biometrics* **48,** 795–806.

Lee, E., Wei, L., and Amato, D. (1992). Cox-type regression analysis for large numbers of small groups of correlated failure time observations. *Survival analysis: State of the art* **211,** 237–247.

Lipsitz, S. and Parzen, M. (1996). A jackknife estimator of variance for cox regression for correlated survival data. *Biometrics* pages 291–298.

Lorino, T., Sanaa, M., Robin, S., and Daudin, J. (2004). Comparison of semiparametric regression models for correlated survival data using simulations. *Communications in Statistics-Theory and Methods* **33,** 1975–1991.

Manatunga, A. and Oakes, D. (1999). Parametric analysis for matched pair survival data. *Lifetime Data Analysis* **5,** 371–387.

Neuhaus, J. and Kalbfleisch, J. (1998). Between-and within-cluster covariate effects in the analysis of clustered data. *Biometrics* **54,** 638–645.

Oakes, D. (1982). A model for association in bivariate survival data. *Journal of the Royal Statistical Society. Series B (Methodological)* pages 414–422.

Oakes, D. (1986). Semiparametric inference in a model for association in bivanate survival data. *Biometrika* **73,** 353–361.

Sjölander, A. and Greenland, S. (2013). Ignoring the matching variables in cohort studies - when is it valid, and why? *Statistics in medicine* In revision.

Sjölander, A., Johansson, A., Lundholm, C., Altman, D., Almqvist, C., and Pawitan, Y. (2012). Analysis of 1: 1 matched cohort studies and twin studies, with binary exposures and binary outcomes. *Statistical Science* **27,** 395–411.

Sjölander, A., Larsson, H., Lichtenstein, P., and Pawitan, Y. (2013). Between-within models for survival analyses. *Statistics in medicine* Accepted for publication.

Tsiatis, A. (1975). A nonidentifiability aspect of the problem of competing risks. *Proceedings of the National Academy of Sciences* **72,** 20–22.

Wild, C. (1983). Failure time models with matched data. *Biometrika* **70,** 633–641.

A Distributions

A.1 The positive stable distribution

Positive stable distributions are a family of distributions with the property that for n i.i.d positively stable distributions $Z_1, ..., Z_n$, there exists a function $c(n)$ such that $c(n)Z_1$ has the same distribution as $Z_1 + ... + Z_n$. The function $c(n)$ has the form $n^{1/\alpha}$. The distribution has two parameters:α and δ with Laplace transform $L(s) = exp\{-\delta s^\alpha/\alpha\}$ In this report, we will set $\delta = \alpha$ and let $0 < \alpha \leq 1$. For $\alpha = 1$, the distribution is degenerate. For $0 < \alpha < 1$ it has density

$$f(y) = -\frac{1}{\pi y}\sum_{k=1}^{\infty}\frac{\Gamma(k\alpha+1)}{k!}(-y^{-\alpha})^k sin(k\alpha\pi)$$

Even thought this distribution is a proper density for $0 < \alpha < 1$, it does not have a finite mean. This distribution will be denoted Posstab(α) and has Laplace transform

$$L(s) = e^{-s^\alpha} \tag{18}$$

For positive stable frailties with parameter α, Kendall's τ is $1 - \alpha$.

A.2 The gamma distribution

The Gamma distribution with shape parameter α and rate parameter θ, has density

$$f_Z(y) = \theta^\alpha y^{\alpha-1}e^{-\theta y}/\Gamma(\alpha)$$

with Laplace transform

$$L(s) = \left(\frac{\theta}{\theta+s}\right)^\alpha$$

In this report, we will set $\theta = \alpha$, so both shape and rate is α and density is

$$f_Z(y) = \alpha^\alpha y^{\alpha-1} e^{-\alpha y}/\Gamma(\alpha)$$

The Laplace transform then becomes

$$L(s) = \left(\frac{\alpha}{\alpha+s}\right)^\alpha = \left(1+\frac{s}{\alpha}\right)^{-\alpha} \tag{19}$$

For gamma distributed frailties with shape and rate α, Kendall's τ is $\frac{1}{1+2\alpha}$.

A.3 The Weibull distribution

The cumulative hazard function for a Weibull distribution with shape parameter c and scale parameter λ has the form

$$H(t) = \lambda t^c$$

The cumulative hazard function

A.4 The Gompertz distribution

The Gompertz distribution with shape η and scale b have the hazard function

$$h(t) = \eta e^{\frac{t}{b}}$$

and cumulative hazard function

$$H(t) = b\eta(e^{\frac{t}{b}} - 1)$$

B Parametric frailty models

If both the baseline distribution and the frailty distribution have scale parameters, all parameters are not indentifiable. That is why whe have to restrict the parameters of the positive stable and gamma distributions.

B.1 The stable-Weibull model

The conditional survival function can then be written

$$S(t_1, ..., t_n|X_1, ..., X_n, Z) = exp\left(-Z\sum_{j=1}^{n}\lambda e^{\beta X_i}t_i^c\right)$$

Since the Laplace transform of the positive stable distribution is $L(s) = e^{-s^\alpha}$ and the cumulative hazard functions for the Weibull distribution is $M(t) = \lambda t^c$, the unconditional survival function can be written as

$$S(t_1, ..., t_n|X_1, ..., X_n) = exp\left\{-\left(\sum_{j=1}^{n}\lambda e^{\beta X_j}t_j^c\right)^\alpha\right\}$$

The marginal lifetimes are Weibull distributed:

$$S(t_j|X_j) = exp\left(-\lambda^\alpha e^{\alpha\beta X_j}t_j^{\alpha c}\right)$$

In the stable-Weibull model both the conditional and the marginal hazards are proportional. The fact the dependency parameter α cannot be identified from the marginal lifetimes is also unique for this model.

B.2 The gamma-Weibull model

The unconditional survival function can be written as

$$S(t_1, ..., t_n|X_1, ..., X_n) = \left(1 + \frac{\sum_{j=1}^{n}\lambda e^{\beta X_j}t_j^c}{\alpha}\right)^{-\alpha}$$

This distribution has marginal distributions that are Burr distributed

$$S(t_j|X_j) = \left(1 + \frac{\lambda e^{\beta X_j}t_j^c}{\alpha}\right)^{-\alpha}$$

B.3 The stable-Gompertz model

Conditional on X and Z the cumulative hazard function is defined as

$$H(t|X, Z) = Ze^{\beta X}b\eta(e^{\frac{t}{b}} - 1)$$

Since the Laplace transform of the positive stable distribution is $L(s) = e^{-s^\alpha}$ and the cumulative hazard functions for the Weibull distribution is $M(t) = \lambda t^c$, the unconditional survival function can be written as

$$S(t_1, ..., t_n | X_1, ..., X_n) = exp\left[- \left\{ \sum_{j=1}^{n} e^{\beta X_j} b\eta \left(e^{\frac{t_j}{b}} - 1 \right) \right\}^\alpha \right]$$

The marginal survival functions are:

$$S(t_j | X_j) = exp\left\{ -e^{\alpha \beta X_j} b^\alpha \eta^\alpha \left(e^{\frac{t_j}{b}} - 1 \right)^\alpha \right\}$$

B.4 The gamma-Gompertz model

The unconditional survival function can be written as

$$S(t_1, ..., t_n | X_1, ..., X_n) = \left\{ 1 + \frac{\sum_{j=1}^{n} e^{\beta X_j} b\eta \left(e^{\frac{t_j}{b}} - 1 \right)}{\alpha} \right\}^{-\alpha}$$

This distribution has the marginal survival functions:

$$S(t_j | X_j) = \left\{ 1 + \frac{e^{\beta X_j} b\eta \left(e^{\frac{t_j}{b}} - 1 \right)}{\alpha} \right\}^{-\alpha}$$

C Kendall's τ

Kendall's τ (Kendall, 1938) is defined as

$$\tau = E[sign\{(T_{i,j} - T_{i',j})(T_{i,j'} - T_{i',j'})\}]$$

where $i \neq i'$ and $j \neq j'$. The first indices i and i' numbers clusters and j and j' numbers individuals within clusters.

Let

$$p = P[(T_{i,j} - T_{i',j})(T_{i,j'} - T_{i',j'}) > 0]$$

Then, for continous distributions, $\tau = 2p - 1$