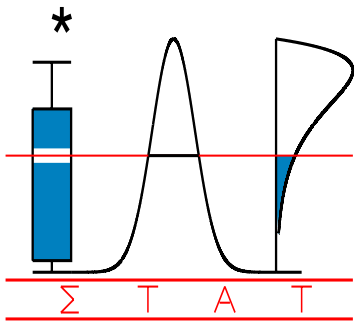
T E C H N I C A L

R E P O R T

0348

RESAMPLING PLANS FOR FRAILTY MODELS

Goele MASSONET, Tomasz BURZYKOWSKI and Paul JANSSEN



I A P S T A T I S T I C S

N E T W O R K

INTERUNIVERSITY ATTRACTION POLE

http://www.stat.ucl.ac.be/IAP

Resampling plans for frailty models

Goele Massonnet, Tomasz Burzykowski, Paul Janssen

Center for Statistics, Limburgs Universitair Centrum, Universitaire Campus, Diepenbeek, Belgium

Abstract

There are several estimation methods available for the estimation of parameters in shared frailty models. However, obtaining the standard error of the parameter estimates, especially for the variance of the frailty, is in general more difficult. A possible solution is to use bootstrap. In this paper we propose two model-based bootstrap algorithms. Furthermore, we compare them to a nonparametric algorithm proposed by Therneau and Grambsch (2000) by means of a simulation study. The results indicate that, under the correct model, one of the proposed algorithms provides relatively precise assessment of the empirical variability of the parameter estimates.

1 Introduction

The shared frailty model is used in order to model correlated survival times. The unobserved risk factor that is common for all the observations in the same cluster is called the frailty. A commonly used estimation procedure in frailty models is the EM algorithm (Klein, 1992). The EM algorithm provides estimates for the treatment effect and for the variance of the frailty density, but does not automatically provide estimates for the variances of these estimates. Klein and Moeschberger (1997, p.413) show how the standard errors of the estimates can be obtained from the inverse of the observed information matrix. This information matrix has rank equal to the number of distinct event times plus the number of covariates plus one (for the heterogeneity parameter). For large data sets, this procedure is not appropriate because of the high dimensionality.

For the gamma distributed frailty case, Therneau and Grambsch (2000, p.254) proved that the estimates obtained from the penalized partial likelihood maximization coincide with the

estimates obtained from the EM algorithm for any fixed value of the heterogeneity parameter. Hence we can use the fast algorithm for the penalized partial likelihood procedure available in S-Plus. However, the standard error estimates reported in S-Plus are computed under the assumption of fixed θ. Since θ needs to be estimated, the given standard errors are too small (Therneau and Grambsch, 2000, p.249).

Thus, the issue of estimating the standard errors of the parameter estimates requires further investigation. A useful tool might be the bootstrap. The results developed for resampling in linear mixed models show that resampling schemes need to be chosen in a careful way (Davison and Hinkley, 1997, p.100-102; Morris, 2002). Therneau and Grambsch (2000, p.249) already proposed a nonparametric bootstrap algorithm to obtain standard error estimates. If we assume a parametric model, however, we might prefer model-based resampling schemes above the nonparametric resampling plan. In this paper we propose two model-based resampling schemes that can be used to find variance estimates.

The frailty model is described in section 2. Some estimation methods for the frailty model are briefly reviewed in section 3. In section 4, two model-based resampling plans for frailty models are proposed. The main purpose of this paper is to compare model-based and nonparametric resampling plans. The comparison is based on a simulation study (section 5).

2 The shared frailty model

Assume we have a total of n individuals that come from K different groups, group i having n_i individuals ($n = \sum_{i=1}^{K} n_i$). Each subject is observed from a time zero to a failure time T_{ij} or to a potential right censoring time C_{ij}. Let $T_{ij}^0 = \min(T_{ij}, C_{ij})$ be the observed time and δ_{ij} be the censoring indicator which is equal to 1 if $T_{ij}^0 = T_{ij}$ and 0 otherwise. Hence the observed data available for the ith individual in the jth group is $y_{ij} = (T_{ij}^0, \delta_{ij})$, with $j = 1, \ldots, n_i$ and $i = 1, \ldots, K$. The number of observed events in group i is $D_i = \sum_{j=1}^{n_i} \delta_{ij}$.

The frailty model is given by

$$h_{ij}(t) = h_0(t) \exp(x_{ij}^T \beta + w_i), \tag{1}$$

where $h_{ij}(t)$ is the hazard rate at time t for individual j from group i, $h_0(t)$ is the baseline hazard at time t, x_{ij} is the vector of p covariates recorded for the individual and w_i is the random effect for group i. In this model $h_0(t)$ can be left unspecified or it may be assumed

to have some specific parametric form. The w_i's, $i = 1, \ldots, K$, are a sample (independent and identically distributed) from a density $f_W(.)$.

Model (1) can be rewritten as:

$$h_{ij}(t) = h_0(t) u_i \exp(x_{ij}^T \beta).$$

The factor $u_i = \exp(w_i)$ is termed the frailty for the ith group. A typical choice for the frailty density is the one-parameter gamma density of form

$$f_U(u) = \frac{u^{(1/\theta)-1} \exp(-u/\theta)}{\theta^{1/\theta} \Gamma(1/\theta)}, \quad \theta > 0.$$

The corresponding density for W is

$$f_W(w) = \frac{\{\exp(w)\}^{1/\theta} \exp\{-\exp(w)/\theta\}}{\theta^{1/\theta} \Gamma(1/\theta)}.$$

For the gamma density E(U_i)= 1. Typically Var(U) = θ is used to describe heterogeneity.

3 Methods of estimation for the shared frailty model

Klein (1992) shows that the observable (marginal) likelihood is given by

$$\begin{aligned} l_{obs}(\beta, \theta, h_0(.)) &= \sum_{i=1}^{K} \Bigg[D_i \log\theta - \log\Gamma(\frac{1}{\theta}) + \log\Gamma(\frac{1}{\theta} + D_i) \\ &\quad - \left(\frac{1}{\theta} + D_i\right) \log\{1 + \theta \sum_{j=1}^{n_i} H_0(t_{ij}) \exp(x_{ij}^T \beta)\} \\ &\quad + \sum_{j=1}^{n_i} \delta_{ij} \{x_{ij}^T \beta + \log h_0(t)\} \Bigg], \end{aligned} \tag{2}$$

where $H_0(t) = \int_0^t h_0(u) du$ is the cumulative baseline hazard.

As noted in the previous section, the baseline hazard $h_0(t)$ in the frailty model can be specified explicitly or left unspecified. Under the parametric assumption, the parameters in the resulting model can be estimated using maximum likelihood estimation procedures. For example, for $h_0(t) \equiv h_0$ constant, the parameters β, θ and h_0 can be estimated by maximizing the observable log likelihood $l_{obs}(\beta, \theta, h_0)$. If $h_0(t)$ is left unspecified, the EM algorithm (Klein, 1992) and the penalized partial likelihood approach (Therneau and Grambsch, 2000) can be used to estimate the unknown parameters in (2).

The EM algorithm for the gamma frailty

To estimate $\zeta = (\theta, \beta)$, we would like to base the likelihood maximization on the observable log likelihood (2). However, this likelihood is difficult to maximize as it contains, apart from ζ, also the unspecified baseline hazard. We therefore rely on the EM algorithm to estimate ζ (for details see, e.g., Duchateau et al., 2002).

It is worth noting that Therneau and Grambsch (2000, p.254) have shown that for any fixed θ, the EM algorithm and the penalized partial likelihood maximization have the same solution for the gamma frailty case. Since S-Plus contains a fast algorithm for the penalized partial likelihood approach, this property is very important from a practical point of view.

The penalized partial likelihood for shared frailty models.

An alternative proposal for the likelihood to use for the estimation of $\zeta = (\theta, \beta)$ is the penalized partial likelihood

$$l_{ppl}(\zeta, w) = l_{part}(\zeta, w) - l_{pen}(\zeta, w),$$

where

$$l_{part}(\zeta, w) = \sum_{l=1}^{r} \left[\sum_{t_{ij}=t_{(l)}} \eta_{ij} - N_{(l)} \log\{ \sum_{t_{qs} \geq t_{(l)}} \exp(\eta_{qs}) \} \right],$$

with $\eta_{ij} = x_{ij}^T \beta + w_i$, r denoting the number of different event times, $t_{(1)} \leq \ldots \leq t_{(r)}$ being the ordered event times, $N_{(l)}$ denoting the number of events at time $t_{(l)}$, $l = 1, \ldots, r$ and

$$l_{pen}(\theta, w) = -\sum_{i=1}^{K} \log f_W(w_i).$$

For random effects w_i, $i = 1, \ldots, K$, with corresponding one-parameter gamma density for the frailties, we have

$$l_{pen}(\theta, w) = -\sum_{i=1}^{K} \left\{ \frac{w_i - \exp(w_i)}{\theta} \right\} - K \left\{ \frac{\log \theta}{\theta} - \log \Gamma \left(\frac{1}{\theta} \right) \right\}.$$

The maximization of the penalized log likelihood consists of an inner and an outer loop. In the inner loop the Newton-Raphson procedure is used to maximize, for a provisional value of θ, $l_{ppl}(\zeta, w)$ for β and w. In the outer loop, a likelihood similar to (2) is maximized for θ as in the case of the EM algorithm. The process is iterated until convergence (for details see, e.g., Duchateau et al., 2002).

4 Bootstrap : Resampling schemes

The EM algorithm does not provide estimates for the variances of the estimates in the frailty model. Klein and Moeschberger (1997) determine the standard errors of the estimates of β and θ from the inverse of the observed information matrix of the observable likelihood. The information matrix is a square matrix of size $r + p + 1$. For large data sets, this approach is not appropriate because of the high dimensionality. On the other hand, the standard error estimates reported by S-Plus are computed under the assumption of θ known (Therneau and Grambsch, 2000, p.249). In many cases, this assumption is not correct and the estimated standard errors are too small. An alternative approach for finding variance estimates might be provided by the bootstrap.

Therneau and Grambsch (2000, p.249) proposed the following nonparametric bootstrap technique to obtain standard error estimates:

1. Choose K groups by sampling with replacement from the K groups in the study.
2. The bootstrap sample contains the subjects from the selected groups.
3. Fit a gamma frailty model with covariates to this bootstrap sample.

This procedure is repeated a number of times. The estimates of the coefficients $\hat{\beta}^*$ and the estimates of the heterogeneity parameter $\hat{\theta}^*$ are stored for each bootstrap sample. The standard errors of the estimated parameters $\hat{\beta}$ and $\hat{\theta}$ are calculated based on the variability of $\hat{\beta}^*$ and $\hat{\theta}^*$.

If a parametric model is appropriate, we might prefer model-based resampling techniques above the nonparametric resampling plan. We therefore propose two model-based resampling schemes. We rely on a resampling plan for a simple random effects model with a balanced design, proposed by Davison and Hinkley (1997, p.102). A random effects model can be written as

$$y_{ij} = x_i + z_{ij}, \quad j = 1, \ldots, n_i = n, \quad i = 1, \ldots, K,$$

where K is the number of groups, $n_i = n$ is the number of subjects per group, the x_i's are randomly sampled from F_x and independent of the z_{ij}'s, which are randomly sampled from F_z with $E(Z) = 0$ to force uniqueness of the model.

In the "naive" version of their algorithm, Davison and Hinkley (1997, p.102) define

$$\hat{x}_i = \bar{y}_i \quad \text{and} \quad \hat{z}_{ij} = y_{ij} - \bar{y}_i \quad .$$

The resampled data set is then obtained in the following way

1. Choose $x_1^*, \ldots, x_K^*$ by randomly sampling with replacement from $\hat{x}_1, \ldots, \hat{x}_K$;
2. Choose $z_{i1}^*, \ldots, z_{in}^*$ randomly with replacement from one group of residuals $\hat{z}_{k1}^*, \ldots, \hat{z}_{kn}^*$, either from a randomly selected group or the group corresponding to x_i^*;
3. Set $y_{ij}^* = x_i^* + z_{ij}^*, \quad j = 1, \ldots, n, \quad i = 1, \ldots, K.$

To construct a resampling plan for frailty models, we can argue that sampling from the means of the groups in the case of the random effects model is like sampling from the frailty estimates in the case of the frailty model. However, in the situation of frailty models, we do not have any residuals to resample from. Therefore, we will adapt a resampling scheme for proportional hazards regression, proposed by Hjort (1985) (see also Davison and Hinkley, 1997, p.351). This resampling scheme can be applied if the survival times are assumed to be independent. We combine both ideas to obtain a model-based resampling plan for the frailty model.

Model-based bootstrap, algorithm 1:

For $j = 1, \ldots, n_i$, $i = 1, \ldots, K$,

1. Fit the model; obtain the estimate $\hat{\beta}$ and the predictions $\hat{u}_1, \ldots, \hat{u}_K$.
2. Choose $u_1^*, \ldots, u_K^*$ by sampling with replacement from $\hat{u}_1, \ldots, \hat{u}_K$.
3. Generate the true failure time T_{ij}^* from the estimated failure time survivor function $\hat{S}_{ij}(t) = \{\hat{S}_0(t)\}^{u_i^* \exp(x_{ij}^T \hat{\beta})}$.
4. If $\delta_{ij} = 0$, set $C_{ij}^* = T_{ij}^0$, and if $\delta_{ij} = 1$, generate C_{ij}^* from the conditional censoring distribution given that $C_{ij} > T_{ij}^0$, namely

$$\frac{\hat{G}(t) - \hat{G}(T_{ij}^0)}{1 - \hat{G}(T_{ij}^0)},$$

   where $\hat{G}$ is an estimate (e.g. Kaplan-Meier) of the censoring distribution G. Assume that G is independent of the covariates.
5. Set $T_{ij}^{0*} = \min(T_{ij}^*, C_{ij}^*)$, with $\delta_{ij}^* = 1$ if $T_{ij}^{0*} = T_{ij}^*$ and zero otherwise.

Steps 3, 4 and 5 are the adaption of the algorithm proposed by Hjort (1985). For mixed models it has been demonstrated (Morris, 2002) that the variances of the BLUP's are biased downwards as estimators of the variance components. Due to this bias, bootstrapping BLUPs results in underestimation of the variation in the data, causing standard error estimates biased downwards. The above-mentioned model-based resampling algorithm may suffer from this problem. Therefore, we propose a second resampling scheme, where resampled frailty parameters are obtained by sampling from a gamma distribution with parameter $\hat{\theta}$. We again assume that censoring is independent of the covariates.

Model-based bootstrap, algorithm 2:

For $j = 1, \ldots, n_i$, $i = 1, \ldots, K$,

1. Fit the model; obtain the estimates $\hat{\beta},\ \hat{\theta}$.
2. Sample $u_1^*, \ldots, u_K^*$ from a gamma distribution with mean 1 and variance $\hat{\theta}$.
3. Generate the true failure time T_{ij}^* from the estimated failure time survivor function $\hat{S}_{ij}(t) = \{\hat{S}_0(t)\}^{u_i^* \exp(x_{ij}^T \hat{\beta})}$.
4. If $\delta_{ij} = 0$, set $C_{ij}^* = T_{ij}^0$, and if $\delta_{ij} = 1$, generate C_{ij}^* from the conditional censoring distribution given that $C_{ij} > T_{ij}^0$, namely
$$\frac{\hat{G}(t) - \hat{G}(T_{ij}^0)}{1 - \hat{G}(T_{ij}^0)}.$$
5. Set $T_{ij}^{0*} = \min(T_{ij}^*, C_{ij}^*)$, with $\delta_{ij}^* = 1$ if $T_{ij}^{0*} = T_{ij}^*$ and zero otherwise.

For a semi-parametric model, the true failure times are generated from the estimated failure time survival function
$$\hat{S}_{ij}(t) = \{\hat{S}_0(t)\}^{u_i^* \exp(x_{ij}^T \hat{\beta})},$$
where $\hat{S}_0(t) = \exp(-\hat{H}_0(t))$ is the estimated baseline survival function, with
$$\hat{H}_0(t) = \sum_{t_{(l)} \leq t} \hat{h}_{l0},$$
where $\hat{H}_0(t)$ is the estimated baseline cumulative hazard at time t and
$$\hat{h}_{l0} = \frac{N_{(l)}}{\sum_{t_{qs} \geq t_{(l)}} u_s^*\ \exp(x_{qs}^T \hat{\beta})}.$$
For a parametric model, the true failure times are generated under the parametric assumption.

5 Simulations

In this section, the two model-based resampling plans are compared to the nonparametric resampling plan by simulation. We will consider the setting of a multicenter clinical trial as the model for the simulations. The size of a multicenter clinical trial is a function of the number of centers, K, and the number of patients per center, n_i, $i = 1, \ldots, K$. In the simulations we study the effect of the number of patients per center on the precision of the variance estimation. Additionally, we investigate the effect of the event rate $h_0(t)$ (assumed constant over time: $h_0(t) = h_0$), the size of the true heterogeneity parameter θ and the treatment effect β.

We assume 15 centers, with 20 or 40 patients per center. The parameter values h_0, β and θ are chosen in such a way that a different magnitude of spread in the median time to event from center to center is induced. This was determined by computing the density function of the median time to event over the centers (Figure 1). It can be shown that this density function $f_{T_M}(t)$ is given by

$$f_{T_M}(t) = \left(\frac{\log(2)}{\theta h_0 \exp(\beta)}\right)^{\frac{1}{\theta}} \frac{1}{\Gamma(1/\theta)} \left(\frac{1}{t}\right)^{1+1/\theta} \exp\left(-\frac{\log(2)}{\theta t h_0 \exp(\beta)}\right).$$

For the treatment effect, we use $\beta = 0.25$. As true values for the event rate, we take $h_0 = 0.1$ and $h_0 = 0.5$. The heterogeneity parameter is set at $\theta = 0.1$ and $\theta = 0.6$.

For setting 1 ($\theta = 0.6, h_0 = 0.5$) and setting 2 ($\theta = 0.6, h_0 = 0.1$), there is little spread in the median time to event over the centers, with a bigger spread for $\theta = 0.6$. For setting 3 ($\theta = 0.1, h_0 = 0.5$) and setting 4 ($\theta = 0.1, h_0 = 0.1$), there is much spread in the median time to event over the centers . Again $\theta = 0.6$ induces more spread.

For each parameter setting $(K, n, h_0, \theta, \beta)$, 100 data sets are generated. Given a particular parameter setting, the observations for each data set are generated in the following way. First, K frailty parameters $u_1, \ldots, u_K$ are generated from a gamma distribution with mean 1 and variance θ. The time to event outcome for the j th patient from center i is randomly generated from an exponential distribution with parameter $h_{ij} = h_0 u_i \exp(x_{ij}^T \beta)$, where x_{ij} is generated from a Bernoulli distribution. The censoring time for each patient is randomly generated from a uniform distribution so that approximately 30% censoring is obtained.

For each simulated data set, $R = 100$ bootstrap samples are taken by using one of the three resampling algorithms. In the nonparametric resampling scheme, the penalized partial likelihood approach is used to estimate the treatment effect and the heterogeneity parameter (Therneau

and Grambsch, 2000). In the model-based resampling schemes, we consider both a parametric frailty model with a constant baseline hazard and a semi-parametric model to compute the estimates of the treatment effect and the heterogeneity parameter. For the parametric model, the model-based resampling schemes assume that the time to event follows an exponential distribution with parameter h_{ij}. Under this assumption, the parameters β, θ and h_0 can be estimated by maximizing the observable log likelihood $l_{obs}(\beta, \theta, h_0)$, given in (2), using the Newton-Raphson method. For the semi-parametric model, the estimates of the treatment effect and the heterogeneity parameter are obtained from the penalized partial likelihood approach.

5.1 Results of simulations

(a) Parametric frailty model with constant baseline hazard:

We will concentrate on the estimated standard error of the heterogeneity parameter (Table 1). For sake of completeness, however, the standard error of the treatment effect is also added to the results (Table 2).

By performing the bootstrap, we obtain for each data set a bootstrap estimate of the standard error of the heterogeneity parameter. The mean of these 100 estimated standard errors is denoted by mean(SE^B). The values of mean(SE^B) for each resampling scheme are compared to the empirical standard error of $\hat{\theta}$, denoted by SE^E. Moreover, since the simulated data sets have moderate size, an estimate of the standard error of the heterogeneity parameter can be determined, for each data set, based on the inverse of the observed information matrix, as explained by Klein and Moeschberger (1997). The mean of these 100 estimated standard errors will be denoted as mean(SE^M).

In all settings studied, the estimated standard error of the heterogeneity parameter obtained by the first model-based resampling plan underestimates the standard error, as compared to SE^E (Figure 2). This underestimation is more clearly visible when the number of patients per center is decreased from 40 to 20. The standard errors obtained by the nonparametric and the second model-based resampling plan are very close to each other. When there are 40 patients per center, the standard error from the nonparametric resampling plan is for each parameter setting between SE^E and mean(SE^M). This is also true for the second model-based resampling plan, except for the setting $\theta = 0.6$, $h_0 = 0.5$. Decreasing the number of patients per center to 20, increases the variability of the parameter estimates and, for all three bootstrap algorithms,

leads to standard errors smaller than SE^E and mean(SE^M).

(b) Semi-parametric frailty model:

As in the parametric case, we obtain for each data set a bootstrap estimate of the standard error of the heterogeneity parameter SE^B (Table 1). For each resampling scheme, the mean value of SE^B is computed. This value is compared to the empirical standard error of $\hat{\theta}$, denoted by SE^E. There is no value of mean (SE^M) provided.
In the case of 40 patients per center and $\theta = 0.6$, the first model-based resampling scheme gives the standard error of $\hat{\theta}$ which is nearest the empirical standard error, whereas the second model-based resampling scheme gives a larger standard error (Figure 2). When $\theta = 0.1$, the bootstrap standard error obtained by each of the three bootstrap algorithms is close to the empirical standard error, with the first model-based resampling scheme giving the smallest value. When the number of patients per center is decreased from 40 to 20, the estimated standard error for $\hat{\theta}$ obtained by the second model-based resampling scheme is nearest the empirical standard error for each setting studied. When $\theta = 0.1$, the values obtained by the nonparametric resampling scheme are very close to the values obtained by the second model-based resampling scheme.

6 Conclusions

In this paper, the use of bootstrap for the estimation of the standard errors of the parameter estimates in a frailty model is proposed. To complement the existing nonparametric resampling plan, we propose two model-based bootstrap algorithms. The comparison between the nonparametric and model-based resampling schemes was studied by simulation. The results indicate that, in most cases, the first model-based resampling plan underestimates the empirical variability of the parameter estimates, whereas the second model-based resampling scheme provides relatively precise estimates. In general, the nonparametric and the second model-based resampling plans give similar results.
The results indicate that the proposed resampling schemes may offer a useful approach to variance extimation. Further investigation of their properties will be necessary, though. For instance, in the model-based resampling schemes we have made the assumption that censoring is independent of the covariates. In principle, it should be possible to extend the schemes to

the more general situation where the censoring distribution depends on the covariates, using the approach developed by Davison and Hinkley (1997, p.351). Furthermore, it would be also of interest to consider frailty densities other than gamma. Finally, the presented simulation results were obtained assuming that the model is correct. The question rises how the bootstrap algorithms behave if the model is misspecified. These are important topics for further research.

Acknowledgement

Financial support from the IAP research network nr. P5/24 of the Belgian Government (Belgian Science Policy) is gratefully acknowledged.

References

A.C. Davison and D.V. Hinkley, *Bootstrap Methods and their Application*, Cambridge University Press, 1997.

L. Duchateau, P. Janssen, P. Lindsey, C. Legrand, R. Ngutti and R. Sylvester, "The shared frailty model and the power for heterogeneity tests in multicenter clinical trials," *Computational Statistics and Data Analysis* vol. 40 pp. 603-620, 2002.

N. L. Hjort, "Bootstrapping Cox's regression model." Technical Report NSF-241, Department of Statistics, Stanford University, 1985.

J.P. Klein, "Semiparametric estimation of random effects using the Cox model-based on the EM algorithm," *Biometrics* vol. 48, pp. 795-806, 1992.

J.P. Klein, M.L. Moeschberger, *Survival Analysis, Techniques for Censored and Truncated Data*, Springer: New York, 1997.

J.S. Morris, "The BLUPs are not "best" when it comes to bootstrapping," *Statistics & Probability Letters* vol. 56, pp. 425-430, 2002.

T.M. Therneau, P.M. Grambsch, *Modeling Survival Data, Extending the Cox Model*, Springer: New York, 2000.

Table 1: Estimated standard error for heterogeneity parameter estimate

Setting (θ, h_0)	SE^E	mean(SE^M)	non-par. mean(SE^B)	model-based(1) mean(SE^B)	model-based(2) mean(SE^B)
Parametric 40					
(0.6, 0.5)	0.1850	0.1909	0.1870	0.1663	0.1829
(0.6, 0.1)	0.1811	0.1994	0.1887	0.1703	0.1958
(0.1, 0.5)	0.0367	0.0482	0.0405	0.0306	0.0396
(0.1, 0.1)	0.0324	0.0467	0.0383	0.0281	0.0371
Parametric 20					
(0.6, 0.5)	0.2266	0.2103	0.2031	0.1802	0.2036
(0.6, 0.1)	0.2351	0.2153	0.2139	0.1838	0.2119
(0.1, 0.5)	0.0531	0.0622	0.0527	0.0373	0.0515
(0.1, 0.1)	0.0527	0.0615	0.0496	0.0370	0.0502
Semi-parametric 40					
(0.6, 0.5)	0.1922		0.1975	0.1960	0.2194
(0.6, 0.1)	0.1862		0.2022	0.1927	0.2162
(0.1, 0.5)	0.0502		0.0418	0.0357	0.0467
(0.1, 0.1)	0.0420		0.0401	0.0348	0.0456
Semi-parametric 20					
(0.6, 0.5)	0.2288		0.1916	0.1798	0.2112
(0.6, 0.1)	0.2292		0.2082	0.2088	0.2426
(0.1, 0.5)	0.0579		0.0482	0.0394	0.0528
(0.1, 0.1)	0.0617		0.0503	0.0393	0.0559

Table 2: Estimated standard error for estimate of treatment effect.

Setting	SE^E	mean(SE^M)	non-par. mean(SE^B)	model-based(1) mean(SE^B)	model-based(2) mean(SE^B)
Parametric 40					
(0.6, 0.5)	0.0934	0.0962	0.0953	0.1017	0.0993
(0.6, 0.1)	0.1085	0.0963	0.0963	0.1008	0.0992
(0.1, 0.5)	0.0894	0.0887	0.0956	0.0999	0.0972
(0.1, 0.1)	0.1102	0.0858	0.0940	0.1009	0.0976
Parametric 20					
(0.6, 0.5)	0.1400	0.1369	0.1393	0.1443	0.1456
(0.6, 0.1)	0.1341	0.1345	0.1330	0.1463	0.1440
(0.1, 0.5)	0.1436	0.1206	0.1362	0.1420	0.1431
(0.1, 0.1)	0.1323	0.1156	0.1431	0.1432	0.1417
Semi-parametric 40					
(0.6, 0.5)	0.0929		0.0969	0.1069	0.1059
(0.6, 0.1)	0.1023		0.0968	0.1049	0.1074
(0.1, 0.5)	0.1005		0.0948	0.0993	0.1014
(0.1, 0.1)	0.0943		0.0934	0.0997	0.1013
Semi-parametric 20					
(0.6, 0.5)	0.1512		0.1375	0.1521	0.1531
(0.6, 0.1)	0.1550		0.1440	0.1562	0.1567
(0.1, 0.5)	0.1588		0.1342	0.1433	0.1433
(0.1, 0.1)	0.1326		0.1300	0.1402	0.1435

Figure 1: Density function of the median time to event from center to center .

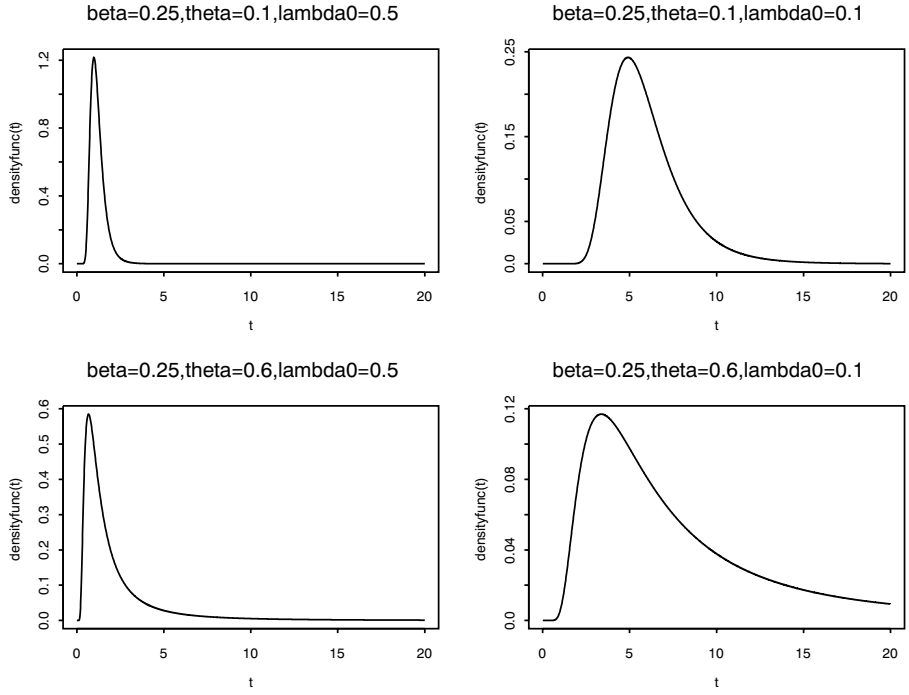


Figure 2: Mean estimated standard errors for the heterogeneity parameter estimate.