Lifetime Data Anal (2006) 12:191–203
DOI 10.1007/s10985-006-9008-y

Effects of unmeasured heterogeneity in the linear transformation model for censored data

Bin Zhang · Yi Li · Rebecca A. Betensky

Received: 27 April 2005 / Accepted: 18 April 2006
Published online: 1 July 2006

Abstract We investigate the effect of unobserved heterogeneity in the context of the linear transformation model for censored survival data in the clinical trials setting. The unobserved heterogeneity is represented by a frailty term, with unknown distribution, in the linear transformation model. The bias of the estimate under the assumption of no unobserved heterogeneity when it truly is present is obtained. We also derive the asymptotic relative efficiency of the estimate of treatment effect under the incorrect assumption of no unobserved heterogeneity. Additionally we investigate the loss of power for clinical trials that are designed assuming the model without frailty when, in fact, the model with frailty is true. Numerical studies under a proportional odds model show that the loss of efficiency and the loss of power can be substantial when the heterogeneity, as embodied by a frailty, is ignored.

Keywords Omitted covariate · Frailty

1. Introduction

One major challenge in the study of brain tumors and many other cancers is the presence of unexplained heterogeneity, which can be understood as completely missing genetic or other prognostic information. Oncologists have long suspected that subsets of patients who benefit from specific therapies might be hidden in larger groups of resistant cases. The impact of unexplained

B. Zhang · Y. Li · R. A. Betensky (✉)
Department of Biostatistics, Harvard School of Public Health, Boston, MA 02115, USA
e-mail: betensky@hsph.harvard.edu

heterogeneity that confers different risks to patients is to undermine the power of a randomized trial to detect a truly beneficial therapy (Betensky et al. 2002). There is a large literature on the effects of unexplained heterogeneity in the context of the proportional hazards model. Lagakos and Schoenfeld (1984) investigated how misspecification of a proportional hazards regression model affects the resulting likelihood score test for comparing randomized treatment groups in the presence of covariate. Li et al. (2002) extended the results of Lagakos and Schoenfeld (1984) to examine the impact of unmeasured heterogeneity, as captured through a continuous frailty that interacts with treatments, on the efficiency and power of the simple log-rank test for treatment effect. Solomon (1984) examined the behavior of the MLE of the regression parameter if proportional hazards are assumed for analysis when the accelerated life model holds, and conversely when accelerated life model is assumed when the proportional hazards model holds. Bretagnolle and Huber-Carol (1988) examined the effect of omitting some of the independent explanatory variables in the Cox regression model for survival data with censoring, and predicted the sign of bias left under any number of omitted covariates. Schmoore and Schumacher (1997) investigated the properties of the MPLE of the treatment effect under two common misspecifications of the models when analyzing survival data in randomized clinical trials using Cox proportional hazards model (a) an important continuous prognostic factor is omitted from the analysis and (b) an important continuous prognostic factor is categorized. Kosorok et al. (2004) studied the *proportional hazards frailty regression model*, a class of semiparametric regression model for right-censored univariate failure times, which assumes the hazard given the covariates and a random frailty unique to each individual has the proportional hazards form multiplied by the frailty.

While the literature on the effects of unexplained heterogeneity is well-developed for the proportional hazards model, there are almost no papers that examine this problem for alternative semiparametric regression models. One exception is a parametric examination of the accelerated failure time model (Solomon 1984). Cheng et al. (1995) introduced a class of semi-parametric linear transformation models, including the proportional hazards and proportional odds models as special cases, for censored data, under which an unknown strictly increasing transformation of the survival time is linearly related to the covariates with some specified error distribution. Slud and Vonta (2004) studied the large sample consistency for non-parametric maximum likelihood estimators of an unknown baseline continuous cumulative-hazard-type function and parameter of group survival differences, based on right-censored two-sample survival data with marginal survival function assumed to follow a transformation model. Slud and Vonta (2005) developed theoretical results specialized to censored linear regression and to a class of semiparametric survival regression model including the proportional hazards models with unobserved random effect. In this paper, we will introduce a random error term (a frailty term) representing the missing prognostic information, to this

class of linear transformation models in the clinical trials setting. We then study the effects of ignoring the frailty term on the parameter estimates of fixed covariates and on the power of associated tests. We also compare the asymptotic relative efficiency (ARE) of the treatment effect based on assuming the simple linear transformation model to that from the frailty model.

There are two distinct types of asymptotic relative efficiency in the literature. One compares two different test statistics as applied to data generated from a single underlying model. One example is a comparison of the simple log-rank test to the optimally weighted log-rank test assuming the model of unobserved heterogeneity to be true (e.g. Lagakos and Schoenfeld, 1984; Li et al. 2002). This comparison measures the loss of efficiency due to use of the wrong test statistic. The second concept of ARE compares the same test statistic as applied to data generated from two different underlying models (Lagakos and Schoenfeld, 1984, Appendix 3; Morgan, 1986). This concept is especially useful for evaluating the loss of power when a study is designed assuming a model without the frailty term when in fact the model with frailty is true. We will study the second type of ARE and translate it into a comparison of planned power to actual power in the presence of the frailty. All future references of efficiency imply the second usage.

We introduce notation and the models in Section 2, and evaluate the bias of the estimate based on ignoring the frailty term in Section 3. In Section 4, we obtain an approximation for the ARE, and we derive an analytic approximation for the actual power of a study designed under the simple linear transformation model when the frailty model is true. We conclude in Section 5 with numerical studies of these results.

2. Notation and models

Suppose that n patients are randomly assigned with probability p to treatment group 1 and $1 - p$ to treatment group 2. Let Z_i $(i = 1, \ldots, n)$ be the treatment indicator for the ith subject with $Z_i = 1$ if subject i received treatment 1 and $Z_i = 0$ if subject i received treatment 2. Let T_i be the failure time and C_i the censoring time associated with subject i. For each subject, a bivariate vector (X_i, Δ_i) is observed, where $X_i = \min(T_i, C_i)$, $\Delta_i = 1$ when $X_i = T_i$, and $\Delta_i = 0$ when $X_i = C_i$. Censoring and the failure times are assumed to be independent. We denote by $G_Z(\cdot)$ the stratified survival function of the censoring variable C_i based on covariate Z. Finally, associated with each subject i is a frailty b_i modeling unobserved heterogeneity, where $b_i \sim F_b(\theta)$ and θ parameterizes the frailty distribution, F_b.

If there is no unobserved heterogeneity, the treatment effect can be estimated by the method proposed by Cheng et al. (1995). Let

$$h(T) = -Z\beta + \varepsilon, \tag{1}$$

where $h(\cdot)$ is an unspecified strictly increasing function of the failure time and ε is a random error with known distribution F. Cheng et al. (1995) proposed the following estimating equations for β,

$$U_0(\beta) = \sum_{i=1}^{n}\sum_{j=1}^{n} \omega(Z_{ij}\beta)Z_{ij}\left\{\frac{\Delta_j I(X_i \geq X_j)}{\widehat{G}_{Z_i}(X_j)\widehat{G}_{Z_j}(X_j)} - \xi_0(Z_{ij}\beta)\right\} = 0, \qquad (2)$$

where $Z_{ij} = Z_i - Z_j$, $\omega(\cdot)$ is a weight function, $\widehat{G}_{Z_i}$ is the Kaplan–Meier estimator for the survival function, G, of the censoring variable, associated with covariate Z_i, and $\xi_0(\cdot)$ is the survival function of $\varepsilon_i - \varepsilon_j$. This class of estimating equations can be generalized when Z is multi-dimensional.

Model (1) represents a large class of survival models indexed by the distribution of ε, F. For example, when F is the extreme value distribution, (1) is the proportional hazards model and when F is the standard logistic distribution, (1) is the proportional odds model. The estimating Eq. (2) yields a unique and consistent estimate for β when Z is a binary variable, such as treatment assignment in a clinical trial (Cheng et al. 1995).

In the presence of unobserved heterogeneity, we extend model (1) through inclusion of a frailty term b, leading to

$$h(T) = -Z\beta + b + \varepsilon. \qquad (3)$$

If the distribution of frailty b were known, this model would simply be a member of the class of model (1). In this case, the approach proposed by Cheng et al. (1995) could be used to estimate β in model (3) by correctly replacing ξ_0 in the estimating Eq. (2), the survival function of $\varepsilon_i - \varepsilon_j$, with ξ_θ, the survival function of $b_i - b_j + \varepsilon_i - \varepsilon_j$. Thus β should be consistently estimated by solving

$$U_\theta(\beta) = \sum_{i=1}^{n}\sum_{j=1}^{n} \omega(Z_{ij}\beta)Z_{ij}\left\{\frac{\Delta_j I(X_i \geq X_j)}{\widehat{G}_{Z_i}(X_j)\widehat{G}_{Z_j}(X_j)} - \xi_\theta(Z_{ij}\beta)\right\} = 0. \qquad (4)$$

Hence, directly using (2), when the heterogeneity does exist (i.e. $\theta \neq 0$), will yield foreseeable biases, which will be investigated in the next section.

Finally, we remark that, when the covariate Z is not a degenerate random variable, the identifiability of β and the frailty parameter, θ, has been detailed in Section 5 of Kosorok et al. (2004), where a set of sufficient identifiability conditions were stated.

3. Estimation and bias

We are now in a position to investigate the effect of ignoring the frailty in the estimation of the treatment effect, β. We assume that the weight function $\omega(\cdot)$ is non-random and positive, and assume for any fixed θ function

$\xi_\theta(\cdot)$ is invertible and the first derivative $\xi_\theta'(\cdot)$ exists. The results are summarized in the following theorem, whose proof is deferred to the Appendix.

Theorem 1 *The Eq.* (2) *has, asymptotically, a unique solution* $\widehat{\beta}$ *when the weight function* $\omega(\cdot)$ *is positive. Moreover, if the frailty model* (3) is true, $\widehat{\beta}$ *obtained from Eq.* (2) *is a consistent estimator of* $\xi_0^{-1}\{\xi_\theta(\beta_0)\}$.

As a special case, this theorem implies when the weight function $\omega(\cdot) \equiv 1$, the solution to (2) is unique. In the ensuing development we evaluate the bias and ARE under $\omega(\cdot) = 1$ for simplicity, and the results for the general weight functions will follow immediately. Indeed, the estimating procedure with a constant weight has been shown to work well for the proportional odds model and other models in the absence of frailties (Cheng et al. 1995).

Under $\omega(\cdot) = 1$, Eq. (2) reduces to

$$\begin{aligned} U_0(\beta) &= \sum_{i=1}^{n}\sum_{j=1}^{n} Z_{ij}\left\{\frac{\Delta_j I(X_i \geq X_j)}{\widehat{G}_{Z_i}(X_j)\widehat{G}_{Z_j}(X_j)} - \xi_0(Z_{ij}\beta)\right\} \\ &= \sum_{i,j:Z_i\neq Z_j} Z_{ij}\frac{\Delta_j I(X_i \geq X_j)}{\widehat{G}_{Z_i}(X_j)\widehat{G}_{Z_j}(X_j)} + \frac{1}{2}\sum_{i,j} I(Z_i \neq Z_j) - \xi_0(\beta)\sum_{i,j} I(Z_i \neq Z_j). \end{aligned}$$

The estimating equation $U_0(\beta) = 0$ has a unique solution given by

$$\widehat{\beta} = \xi_0^{-1}\left(\frac{\sum_{i,j:Z_i\neq Z_j} Z_{ij}\{\Delta_j I(X_i \geq X_j)/\widehat{G}_{Z_i}(X_j)\widehat{G}_{Z_j}(X_j)\}}{\sum_{i,j} I(Z_i \neq Z_j)} + \frac{1}{2}\right). \tag{5}$$

If model (1) is true, Cheng et al. (1995) proved that $\widehat{\beta}$ is an unbiased and consistent estimator of the true treatment effect β_0. However, $\widehat{\beta}$ is a biased estimator when the frailty model (3) is true.

For the asymptotic bias of $\widehat{\beta}$, note that

$$E\left\{\frac{\Delta_j I(X_i \geq X_j)}{G_{Z_i}(X_j)G_{Z_j}(X_j)}|Z_i, Z_j\right\} = E[I\{h(T_i) \geq h(T_j)\}|Z_i, Z_j] = \xi_\theta(Z_{ij}\beta_0),$$

and

$$n^{-2}\sum_{i,j} I(Z_i \neq Z_j) \xrightarrow{\text{a.s.}} 2p(1-p).$$

$$\frac{\sum_{i,j:Z_i\neq Z_j} Z_{ij}\{\Delta_j I(X_i \geq X_j)/G_{Z_i}(X_j)G_{Z_j}(X_j)\}}{\sum_{i,j} I(Z_i \neq Z_j)} + \frac{1}{2} \xrightarrow{\text{a.s.}} \xi_\theta(\beta_0).$$

Again, replacing G with its consistent estimator $\widehat{G}$ results in $\xi_0(\widehat{\beta}) \xrightarrow{P} \xi_\theta(\beta_0)$, and $\widehat{\beta} \xrightarrow{P} \xi_0^{-1}(\xi_\theta(\beta_0))$. Under the assumption of uniform integrability of $\widehat{\beta}$, $E(\widehat{\beta}) \longrightarrow \xi_0^{-1}\{\xi_\theta(\beta_0)\}$, and the asymptotic bias of $\widehat{\beta}$ is $E(\widehat{\beta}) - \beta_0 \longrightarrow \xi_0^{-1}\{\xi_\theta(\beta_0)\} - \beta_0$.

4. ARE and power

In this section, we compare the asymptotic efficiency of $\widehat{\beta}$, the estimate of treatment effect derived using estimating Eq. (2), when the data arise from the simple model (1) versus the frailty model (3).

We begin by deriving the asymptotic distribution of $\widehat{\beta}$, the estimator obtained using (2), under each of the two models (1) and (3). When model (1) is true, it follows that

$$\begin{aligned} & n^{-\frac{3}{2}} \sum_{i,j} I(Z_i \neq Z_j)[\xi_0(\widehat{\beta}) - \xi_0(\beta)] \\ & \quad = n^{-\frac{3}{2}} \left\{ \sum_{i,j:Z_i \neq Z_j} \frac{Z_{ij}\{\Delta_j I(X_i \geq X_j)\}}{\widehat{G}_{Z_i}(X_j)\widehat{G}_{Z_j}(X_j)} + \frac{1}{2}\sum_{i,j} I(Z_i \neq Z_j) - \xi_0(\beta) \sum_{i,j} I(Z_i \neq Z_j) \right\} \\ & \quad = n^{-\frac{3}{2}} U_0(\beta). \end{aligned}$$

It also follows from Cheng et al. (1995) that the distribution of $n^{-\frac{3}{2}}U_0(\beta)$ can be approximated by $N(0,\Gamma_0)$, where

$$\begin{aligned} \Gamma_0 = \lim_{n\to\infty} \Bigg[& \frac{1}{n^3} \sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{k=1,k\neq j}^{n} \{e_{0,ij}(\beta) - e_{0,ji}(\beta)\} \\ & \times \{e_{0,ik}(\beta) - e_{0,ki}(\beta)\} Z_{ij} Z_{ik} - 4 \int_0^\infty \frac{q(t)q(t)'}{\pi(t)} d\Lambda_G(t) \Bigg], \end{aligned}$$

$$e_{0,ij}(\beta) = \frac{\Delta_j I(X_i \geq X_j)}{\widehat{G}_{Z_i}(X_j)\widehat{G}_{Z_j}(X_j)} - \xi_0(Z_{ij}\beta), \qquad \pi(t) = \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} I(X_i \geq t),$$

$$q(t) = \lim_{n\to\infty} \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} Z_{ij} \frac{\Delta_j I(X_i \geq X_j)}{\widehat{G}_{Z_i}(X_j)\widehat{G}_{Z_j}(X_j)} I(X_j \geq t),$$

and Λ_G is the common cumulative hazard function of the censoring variable, C.

Combining these two results through an application of the Slutsky theorem and the delta method yields the asymptotic distribution of $\widehat{\beta}$ as

$$\sqrt{n}(\widehat{\beta} - \beta) \overset{\text{distribution}}{\longrightarrow} N\left(0, \frac{\Gamma_0}{4p^2(1-p)^2\{\xi_0'(\beta)\}^2}\right). \tag{6}$$

Similarly, under the frailty model (3), it follows that

$$
\begin{aligned}
& n^{-\frac{3}{2}} \sum_{i,j} I(Z_i \neq Z_j)[\xi_0(\widehat{\beta}) - \xi_\theta(\beta)] \\
& \quad = n^{-\frac{3}{2}} \Bigg\{ \sum_{i,j:Z_i \neq Z_j} Z_{ij} \frac{\Delta_j I(X_i \geq X_j)}{\widehat{G}_{Z_i}(X_j)\widehat{G}_{Z_j}(X_j)} \\
& \qquad + \frac{1}{2} \sum_{i,j} I(Z_i \neq Z_j) - \xi_\theta(\beta) \sum_{i,j} I(Z_i \neq Z_j) \Bigg\} \\
& \quad = n^{-\frac{3}{2}} U_\theta(\beta).
\end{aligned}
$$

Also as in Cheng et al. (1995), $n^{-\frac{3}{2}} U_\theta(\beta)$ is approximately $N(0, \Gamma_\theta)$, where

$$
\begin{aligned}
\Gamma_\theta = \lim_{n\to\infty} \Bigg[& \frac{1}{n^3} \sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{k=1, k\neq j}^{n} \{e_{\theta,ij}(\beta) - e_{\theta,ji}(\beta)\} \\
& \times \{e_{\theta,ik}(\beta) - e_{\theta,ki}(\beta)\} Z_{ij} Z_{ik} - 4 \int_0^\infty \frac{q_\theta(t) q_\theta(t)'}{\pi_\theta(t)} d\Lambda_G(t) \Bigg], \\
e_{\theta,ij}(\beta) = & \frac{\Delta_j I(X_i \geq X_j)}{\widehat{G}_{Z_i}(X_j)\widehat{G}_{Z_j}(X_j)} - \xi_\theta(Z_{ij}\beta),
\end{aligned}
$$

$\pi_\theta(t)$, $q_\theta(t)$ and Λ_G are defined in a similar fashion as before.

Another application of the Slutsky theorem and delta method yields the asymptotic distribution of $\widehat{\beta}$ under model (3) as

$$
\sqrt{n}(\widehat{\beta} - \xi_0^{-1}[\xi_\theta(\beta)]) \overset{\text{distribution}}{\rightarrow} N\left(0, \frac{\Gamma_\theta}{4p^2(1-p)^2\{\xi_\theta'(\beta)\}^2}\right). \tag{7}
$$

To investigate the behavior of the ARE, we consider a sequence of alternatives that converge to the null hypothesis at the appropriate rate as sample size goes to infinity, which allow the estimators to be asymptotically unbiased under both models. The result is stated in the following theorem and its proof can be seen in the Appendix.

Theorem 2 *Consider a sequence of local alternatives* $H_{A,n} : \beta_n = \frac{c}{\sqrt{n}}$, *where* c *is a constant, and further assume that* ξ_0 *and* ξ_θ *are survival functions with continuous second derivatives in a small neighborhood of* 0. *Then, under the linear transformation model* (1), *the ARE of* $\widehat{\beta}$ *under the frailty model* (3) *versus* $\widehat{\beta}$ *under the linear transformation model* (1) *is*

$$
\left\{ \frac{\xi_\theta'(0)}{\xi_0'(0)} \right\}^4 \frac{\Gamma_0}{\Gamma_\theta}. \tag{8}
$$

We remark that, in the particular setting this paper concerns, (8) will not exceed one. Indeed, since $\Gamma_0 \leq \Gamma_\theta$ by definition, we only need to show $\xi_\theta'(0) \leq \xi_0'(0)$. To see this, recall that $\xi_0'(\cdot)$ is the density function of $\varepsilon_i - \varepsilon_j$ at point 0 and $\xi_\theta'(\cdot)$ is the density function of $b_i - b_j + \varepsilon_i - \varepsilon_j$ at point 0. Consider two generic independent random variables X and Y. Let f be the probability density function and F the corresponding CDF of X, and g be the probability density function and G the corresponding CDF of Y. It follows that $f_{X+Y}(t) = \int_{-\infty}^{\infty} f(t-y)dG(y)$. For symmetric density function of $X, f_{X+Y}(0) = \int_{-\infty}^{\infty} f(y)dG(y) = E(f_X(Y))$. Hence, $\xi_\theta'(0) = E\{\xi_0'(b_i - b_j)\}$. On the other hand,

$$E\{\xi_0'(b_i - b_j)\} \leq \xi_0'(0),$$

holds when $\xi_0'(x)$ attains its maximum at 0. This is satisfied by the common random error variable ε in linear transformation models, including extreme value distribution, standard logistic distribution.

We are now ready to study the actual power of a clinical trial designed assuming the simple linear transformation model (1) when, in fact, the frailty model (3) is true. For a sequence of local alternatives $\beta_n = c\, n^{-1/2}$, the distribution of $\widehat{\beta}$ under the linear transformation model (1) is approximated by

$$N\left(\beta_n, \frac{\Gamma_0}{4np^2(1-p)^2\{\xi_0'(0)\}^2}\right), \tag{9}$$

and the distribution of $\widehat{\beta}$ under the frailty model (3) is approximated by

$$N\left(\beta_n \frac{\xi_\theta'(0)}{\xi_0'(0)}, \frac{\Gamma_\theta}{4np^2(1-p)^2\{\xi_\theta'(0)\}^2}\right). \tag{10}$$

Suppose that we design a clinical trial based on the wrong model (1). For simplicity we assume that the censoring distributions for the two treatment groups are the same, and there is equal number of subjects in the two treatment groups. It then follows that the approximate sample size needed to detect H_a: $\beta = \beta^* > 0$ vs H_0: $\beta = 0$ under the wrong model (1) with power δ and one-sided type I error α is

$$n_0 = \frac{(Z_{1-\alpha} + Z_\delta)^2 \Gamma_0}{\beta^{*2} 4p^2(1-p)^2\{\xi_0'(0)\}^2}. \tag{11}$$

Hence, for a study design based on the incorrect simple linear transformation model (1) when in fact the model with frailty (3) holds, the actual power is

$$1 - \Phi\left(\frac{\xi_\theta'(0)\Gamma_0^{\frac{1}{2}}}{\xi_0'(0)\Gamma_\theta^{\frac{1}{2}}}\left\{Z_{1-\alpha} - \frac{\xi_\theta'(0)}{\xi_0'(0)}(Z_{1-\alpha} + Z_\delta)\right\}\right), \tag{12}$$

where Φ is the cumulative distribution function of a standard normal distribution. Note that the power does not depend on β^* because the relative power depends only on the ratio of the variances and the non-centrality parameters of $\widehat{\beta}$ under the two different models.

5. Numerical studies

In this section, we numerically evaluate, under several parameter configurations, the ARE of the simple (i.e. via (2)) treatment estimate based on the simple linear transformation model to that based on the frailty model, and the power of the test statistics for a study designed assuming model (1). Specifically, we study the special case of the proportional odds model, a popular alternative to the proportional hazards model, for which these investigations have not been previously carried out. We examine the ARE and power under three distributions for the frailty b: 1) $b \sim N(\mu,\ \theta)$, 2) $\exp(b) \sim gamma(\theta^{-1}, \theta^{-1})$, 3) $b \sim$ Inverse Gaussian$(\mu = 1, \lambda = \theta^{-1})$. In particular, we note when b is inverse Gaussian with mean μ and precision parameter λ, its variance is given by μ^3/λ. All the simulations are conducted using R (http://www.r-project.org/).

For a proportional odds model, ε has a standard logistic distribution, $\xi_0(x) = 1 - F_\varepsilon(x) = \frac{e^{-x}}{1+e^{-x}}$. To obtain the density functions of $b_i - b_j + \varepsilon_i - \varepsilon_j$ and $\varepsilon_i - \varepsilon_j$, which usually do not have closed form solutions, we first calculate their characteristic functions and then use the inverse formula to transform the characteristic functions back into density functions. Letting $\phi_b(t)$ denote the characteristic function of b and $\phi_\varepsilon(t)$ denote that of ε, the characteristic functions of $b_i - b_j + \varepsilon_i - \varepsilon_j$ and $\varepsilon_i - \varepsilon_j$ are simply $\phi_{b,\varepsilon}(t) = \phi_b(t)\phi_b(-t)\phi_\varepsilon(t)\phi_\varepsilon(-t)$ and $\phi_{0,\varepsilon}(t) = \phi_\varepsilon(t)\phi_\varepsilon(-t)$. These are real valued functions because the distributions of $b_i - b_j + \varepsilon_i - \varepsilon_j$ and $\varepsilon_i - \varepsilon_j$ are symmetric about 0. Then the density functions $\xi_\theta'(x)$ and $\xi_0'(x)$ are obtained by the inverse formula $\xi'(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-itx}\phi(t)dt$. The values of $\xi_\theta'(0)$ and $\xi_0'(0)$ can be readily obtained by numerical integration for virtually any distributions of b and ε.

We fix $\beta_0 = 0.3$ and generate 500 data sets with $n = 100$ subjects and 500 data sets with $n = 200$. We examine values of θ ranging from 0.1 to 2 in increments of 0.2. We then generate values of b from the three frailty distributions. We generate uniform censoring times $U(0,C_j)$, ($j = 1,2,3$), and C_j is so chosen that censoring rate of 10%, 20% and 50% for each parameter configuration are achieved, i.e., the proportion of $T_{ij} < C_{ij}, i = 1, \ldots, n$ is 10%, 20% and 50% for $j = 1,2,3$. For the ARE formula (8) and power calculation (12), the values of $\frac{\xi_\theta'(0)\Gamma_0^{1/2}}{\xi_0'(0)\Gamma_\theta^{1/2}}$ are difficult to compute numerically. Since they are equal to the square root of the ratio of the variances in Eqs. (9) and (10), we estimate them using the ratio of the empirical variances of the estimated value of β from the simulations.

Figure 1 displays the lowess smoothed ARE and power as a function of θ for the three frailty models when $n = 100$. The graphs for $n = 200$ (not shown) display

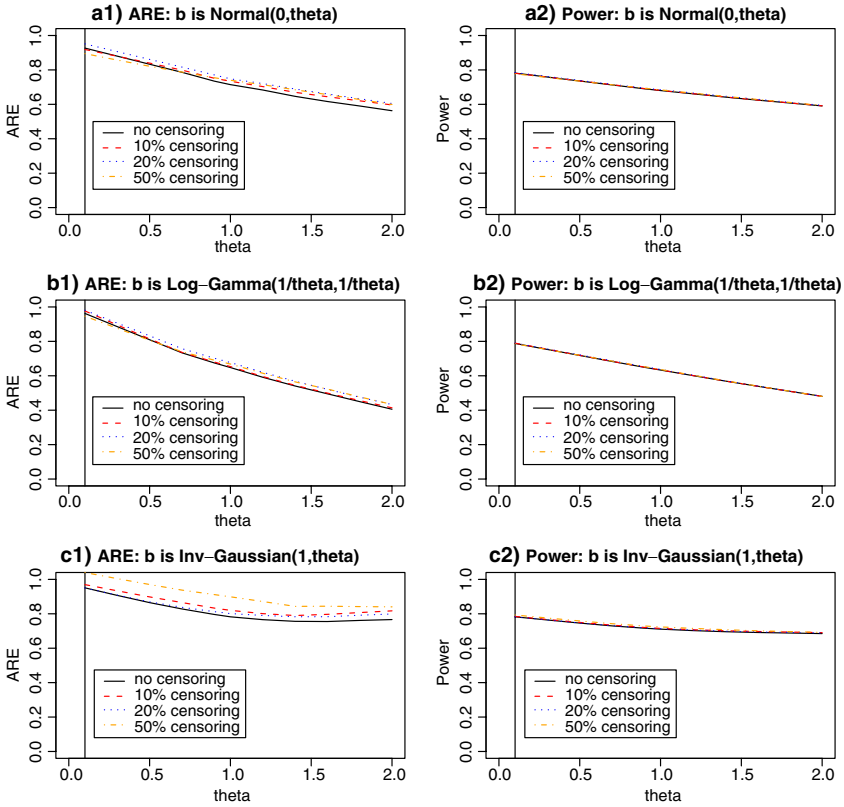

Fig. 1 (**a1**) ARE of normal frailty model, (**a2**)Power of normal frailty model, (**b1**) ARE of log-gamma frailty model, (**b2**) Power of log-gamma frailty model, (**c1**) ARE of inverse-Gaussian frailty model, (**c2**) Power of inverse-Gaussian frailty model

similar patterns. Figures *a*1–*c*2 display the ARE and power comparisons based on the normal frailty model, the log–gamma frailty model and the inverse Gaussian frailty model, respectively, all of which demonstrate that the ARE and power decrease as θ, the variance of the frailty term, increases. Additionally, Fig. 2 indicates that the bias of the estimates increases as θ increases. The bias associated with heavy censoring, e.g., 50%, is obviously larger than those with moderate censorings. Also of note is that the effect of censoring on the ARE and actual power is negligible. This is not surprising as the ARE and power are related to the ratio of the variances of $\widehat{\beta}$ between the simple model (1) and the frailty model (3), and it is likely that the censoring affects both variances in a similar magnitude.

6. Summary

We have considered unobserved heterogeneity in the estimation of the linear transformation model with censored data in the clinical trials setting.

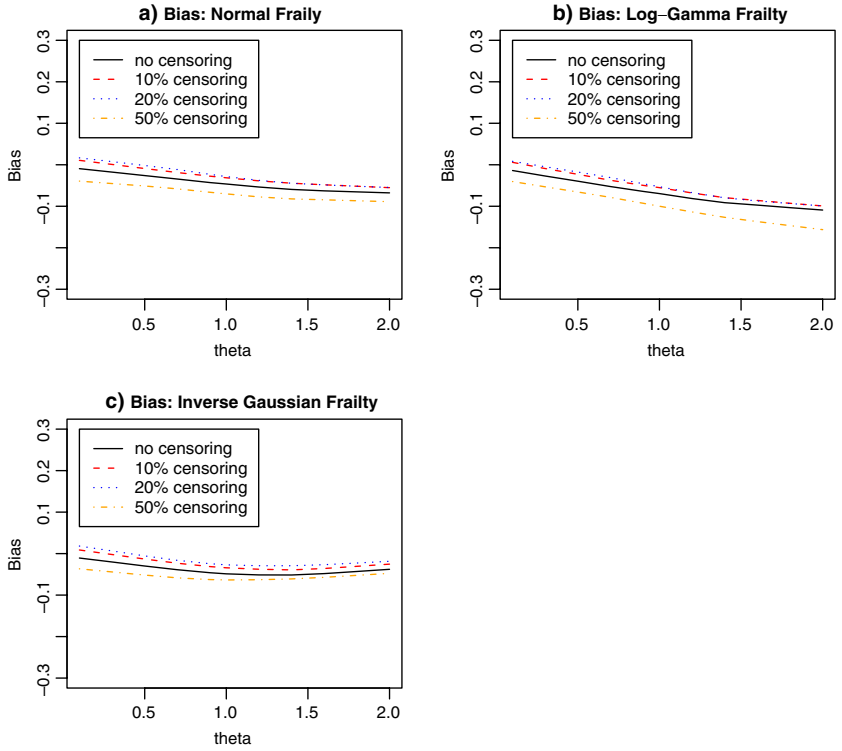

Fig. 2 (**a**) Bias of normal frailty model, (**b**) Bias of log-gamma frailty model, (**c**) Bias of inverse-Gaussian frailty model

The unobserved heterogeneity is represented by a frailty term in the model. An estimate of the treatment effect is obtained when such unobserved heterogeneity is ignored. We obtained the bias of the estimate when the frailty distribution is known, and we investigated the asymptotic relative efficiency of the estimate when data arises from the linear transformation model for censored data without the frailty as compared to data from the model with frailty. We further investigated the loss of power when a clinical trial is designed assuming the model without frailty when in fact the model with frailty is true. For a proportional odds model, it is shown through numerical studies that the loss of efficiency and the loss of power are substantial when the frailty model is true but the simple model is assumed. This is a real concern for many clinical trials that are designed assuming homogeneity among patients when, in fact, they are heterogeneous. If there are suggestions of heterogeneity in the patient population, an adaptive design might be implemented to maintain proper power of the test.

Acknowledgements We thank the editor, an associate editor and two referees for their insightful comments.

Appendix 1: Proof to Theorem 1

First note

$$Z_{ij} = Z_i - Z_j = \begin{cases} 0 & \text{with probability} \quad (1-p)^2 + p^2 \\ 1 & \text{with probability} \quad p(1-p) \\ -1 & \text{with probability} \quad p(1-p), \end{cases}$$

using standard asymptotic theory of multivariate U-statistics, it can be shown that

$$n^{-2}\sum_{i,j}\omega(Z_{ij}\beta)Z_{ij}\frac{\Delta_j I(X_i \geq X_j)}{G_{Z_i}(X_j)G_{Z_j}(X_j)} \xrightarrow{\text{a.s.}} p(1-p)\{[\omega(\beta)+\omega(-\beta)]\xi_\theta(\beta_0)-\omega(-\beta_0)\}.$$

It follows that

$$\begin{aligned} & n^{-2}\sum_{i=1}^{n}\sum_{j=1}^{n}\omega(Z_{ij}\beta)Z_{ij}\left\{\frac{\Delta_j I(X_i \geq X_j)}{G_{Z_i}(X_j)G_{Z_j}(X_j)} - \xi_0(Z_{ij}\beta)\right\} \\ & \quad = n^{-2}\sum_{Z_{ij}=1}\omega(\beta)\left\{\frac{\Delta_j I(X_i \geq X_j)}{G_{Z_i}(X_j)G_{Z_j}(X_j)} - \xi_0(\beta)\right\} \\ & \qquad - n^{-2}\sum_{Z_{ij}=-1}\omega(-\beta)\left\{\frac{\Delta_j I(X_i \geq X_j)}{G_{Z_i}(X_j)G_{Z_j}(X_j)} - \xi_0(-\beta)\right\}, \\ & \quad \xrightarrow{\text{a.s.}} p(1-p)\{\omega(\beta)+\omega(-\beta)\}\{\xi_\theta(\beta_0)-\xi_0(\beta)\}. \end{aligned}$$

Replacing G with its consistent estimate $\widehat{G}$ and

$$n^{-2}U_0(\beta) \xrightarrow{P} p(1-p)\{\omega(\beta)+\omega(-\beta)\}\{\xi_\theta(\beta_0)-\xi_0(\beta)\}.$$

When $\omega(\cdot) > 0$,

$$p(1-p)\{\omega(\beta+\omega(-\beta)\}\{\xi_\theta(\beta_0)-\xi_0(\beta)\} = 0$$

iff

$$\xi_\theta(\beta_0) - \xi_0(\beta) = 0.$$

Hence the unique asymptotic solution to $U_0(\beta) = 0$ is $\beta = \xi_0^{-1}\{\xi_\theta(\beta_0)\}$. This implies that $\widehat{\beta}$ obtained by solving equation (2) is a unique consistent estimate of $\xi_0^{-1}\{\xi_\theta(\beta_0)\}$.

Appendix 2: Proof to Theorem 2

For the local alternatives $H_{A,n} : \beta_n = \frac{c}{\sqrt{n}}$, when the linear transformation model (1) is true, (6) shows that

$$\sqrt{n}\left(\widehat{\beta}-\beta_n\right)=\sqrt{n}\widehat{\beta}-c \overset{\text{approx.}}{\sim} N\left(0,\frac{\Gamma_0}{4p^2(1-p)^2\{\xi_0'(0)\}^2}\right).$$

A Taylor expansion of $\xi_0^{-1}(\xi_\theta(\beta_n))$ yields

$$\xi_0^{-1}(\xi_\theta(\beta_n))=\xi_0^{-1}\left(\xi_\theta\left(\frac{c}{\sqrt{n}}\right)\right)=\frac{c}{\sqrt{n}}\frac{\xi_\theta'(0)}{\xi_0'(0)}+o(n^{-\frac{1}{2}}).$$

Under the frailty model (3), (7) shows that

$$\sqrt{n}\left\{\widehat{\beta}-\xi_0^{-1}[\xi_\theta(\beta_n)]\right\}=\sqrt{n}\widehat{\beta}-c\frac{\xi_\theta'(0)}{\xi_0'(0)}+o(1) \overset{\text{approx.}}{\sim} N\left(0,\frac{\Gamma_\theta}{4p^2(1-p)^2\{\xi_\theta'(0)\}^2}\right).$$

Hence the ARE of $\widehat{\beta}$ under the frailty model (3) versus $\widehat{\beta}$ under the linear transformation model (1) is simply the ratio of the two non-centralities,

$$\left\{\frac{\xi_\theta'(0)}{\xi_0'(0)}\right\}^2\frac{\Gamma_0}{\Gamma_\theta}\frac{\{\xi_\theta'(0)\}^2}{\{\xi_0'(0)\}^2}=\left\{\frac{\xi_\theta'(0)}{\xi_0'(0)}\right\}^4\frac{\Gamma_0}{\Gamma_\theta}.$$

References

Betensky RA, Louis DN, Cairncross JG (2002) The influence of unrecognized molecular heterogeneity on randomized clinical trials. J Clin Oncol 20:2495–2499

Bretagnolle J, Huber-Carol C (1988) Effects of omitting covariates in Cox's model for survival data. Scand J Statist 15:125–138

Cheng SC, Wei LJ, Ying Z (1995) Analysis of transformation models with censored data. Biometrika 82:835–845

Kosorok MR, Lee BL, Fine JP (2004) Robust inference for proportional hazards univariate frailty regression models. Ann Statist 32:1448–1491

Morgan TM (1986) Omitting covariates from the proportional hazard models. Biometrics 42:993–995

Lagakos SW, Schoenfeld DA (1984) Properties of Proportional-Hazard Score Tests Under Misspecified Regression Models. Biometrics 42:993–995

Li Y, Betensky RA, Louis DN, Cairncross JG (2002) The use of frailty hazard models for unrecognized heterogeneity that interacts with treatment: consideration of efficiency and power. Biometrics 58:232–236

Slud EV, Vonta F (2005) Efficient semiparametric estimators via modified profile likelihood. J Statist Plann Infer 129:339–367

Slud EV, Vonta F (2004) Consistency of the NPML estimator in the right-censored transformation model. Scand J Statist 31:21–41

Schmoor C, Schumacher M (1997) Effects of covariate omission and categorization when analysing randomized trials with the Cox model. Statist Med 16:225–237

Solomon PJ (1984) Effect of misspecification of regression models in the analysis of survival data. Biometrika 71:291–198