Testing For Center Effects In Multicenter Survival Studies: A Monte Carlo Comparison Of Fixed And Random Effects Tests

Per Kragh Andersen
University of Copenhagen and Danish Epidemiology Centre
John P. Klein and Mei-Jie Zhang
Medical College of Wisconsin

Technical Report 23

January 1997

Division of Biostatistics
Medical College of Wisconsin
8701 Watertown Plank Road
Milwaukee WI 53226
Phone: (414)456-8280

Testing For Center Effects In Multicenter Survival Studies: A Monte Carlo Comparison Of Fixed And Random Effects Tests

Per Kragh Andersen
University of Copenhagen and Danish Epidemiology Centre

And

John P. Klein and Mie-Jie Zhang
Medical College of Wisconsin

SUMMARY

The problem of testing for a center effect following a proportional hazards regression is considered. Two approaches to the problem can be used. One approach fits a proportional hazards model with a fixed covariate included for each center. The need for a center specific adjustment is evaluated using either a score, Wald or likelihood ratio test of the hypothesis that all the center specific covariates are equal to zero. An alternative approach is to introduce a random effect or frailty for each center into the model. Recently, Commenges and Andersen [1], have proposed a score test for this random effects model.

By a Monte Carlo study we compare the performance of these two approaches when either the fixed or random effects model holds true. The study shows that for moderate samples the fixed effects tests have nominal levels much higher than specified, but the random effect test performs as expected under the null hypothesis. Under the alternative hypothesis the random effect test has good power to detect relatively small fixed or random center effects. Also if the center effect is ignored the estimator of the main effect may be quite biased and the estimator is inconsistent. The tests are illustrated on a retrospective multicenter study of the recovery from bone marrow transplantation.

1. Introduction

A common question arising in multi-center prospective clinical trials and in retrospective studies from collaborative registry studies is whether some statistical adjustment is needed to account for effects specific to the individual centers contributing patients to the trial. Such an adjustment may be needed to account for factors, related to the outcome, which vary from center to center but are not adjusted for in the analysis. These factors may involve measurable quantities like a center's protocol for supportive therapy, the number of similar cases treated by the center, etc., or they be unmeasurable factors like the quality of the center's medical staff or differences in a center's catchment population.

In this paper we study two methods for testing the hypothesis of no center specific effect when the outcome measure is the time to some event. In such studies, typically, data is

analyzed using the Cox [2] proportional hazards regression model. The typical analysis includes covariates for the main effect of interest in the study as well as patient specific covariates which are related to the outcome of interest. The patient specific covariates are included in the final model in a partial attempt to make an adjustment for differences in patient demographics between institutions (See Klein and Moeschberger [3] for details on model building in this situation.)

The first method used to test for the presence of a center effect in such studies is the use of a fixed effect proportional hazards model. In this approach one institution is picked as a baseline institution and a set of indicator covariates are included for all other institutions. If we let $\boldsymbol{Z}$ denote the treatment and patient specific covariates and $X_i = \{1$ if the patient is from institution i; 0 otherwise$\}$, for $i = 1, \cdots, K$, where K is the number of institutions contributing to the study, then the hazard rate for the jth patient from institution i is

$$\lambda_{ij}(t|\boldsymbol{Z}_{ij}) = \lambda_0(t) \exp\{\beta' \boldsymbol{Z}_{ij} + \theta' \boldsymbol{X}\} \tag{1}$$

where $\boldsymbol{X} = (X_1, \cdots, X_{K-1})$. If there is no center specific effect in the study then $\theta_1 = \theta_2 = \cdots = \theta_{K-1} = 0$. To test the hypothesis of no center effect one can used a standard Wald, likelihood ratio or score test available in many statistical packages (See Andersen et al [4] or Klein and Moeschberger [3] for details).

An alternate approach to testing for a center effect is to use a random effects or frailty model. Such models were introduced by Clayton [5] and Vaupel et al. [6] and further discussed by, among others, Klein [7], Nielsen et al [8], Andersen et al. [4] and Klein and Moeschberger [3]. Here one assumes that the center specific effect for the ith center is represented by a mean 0, variance 1, unobservable random variable, ϵ_i, which acts multiplicatively on the hazard rate for all individuals within the center. That is

$$\begin{aligned} \lambda_{ij}(t|\boldsymbol{Z}_{ij}) &= \lambda_0(t) \exp\{\beta' \boldsymbol{Z}_{ij} + \sigma \epsilon_i\} \\ &= \lambda_0(t) u_i \exp\{\beta' \boldsymbol{Z}_{ij}\} \end{aligned} \tag{2}$$

where $u_i = \exp\{\sigma \epsilon_i\}$. The ϵ_i's are an *i.i.d.* sample from the unknown frailty distribution. In this model, the test of no center effect reduces to a test of the hypothesis that σ is equal to 0. Commenges and Andersen [1] have recently developed a score test of this hypothesis that does not require specification of the unknown frailty distribution. Computational details of this test are given in the Appendix.

In this paper we examine the relative performance of these two procedures by a Monte Carlo study. Details of the study are given in Section 2. In Section 3 we examine the performance of the two approaches when the null hypothesis is true. In Section 4 we examine the power of the two approaches when either the fixed or random effect model is true. In Section 5 we illustrate the use of the two statistics on a data set of allogeneic bone marrow transplants based on data from a collaborative bone marrow transplant registry. Finally, in Section 6 we summarize our conclusions and make some suggestions of how to proceed when the hypothesis of no center effect is rejected.

2. The Monte Carlo Study

To study the two approaches to testing for a potential center effect a Monte Carlo study was performed. In the study a single fixed time covariate, Z, was used. The covariate Z was taken to be $+1$ for half of the patients at each center and -1 for the remaining half. The value of the regression coefficient was taken to be either zero or $\ln(2)$. The baseline hazard rate was assumed to be one for all t. A random censoring time was generated for each subject from an exponential population with hazard rate equal to either 1/9 or 3/7. This leads to appoximately 10% or 30% of the observations being censored, respectively.

To investigate the relationship between the number of centers and the number of observations per center on the power of the tests we generated data coming from 5, 10 or 20 centers with a total of 100, 200, or 400 observations in the total sample. Data was generated from one of five models for the center effect. For the first case all observations were independent and no center effect was generated. This corresponds to the null case. For the other four cases data was generated either from a model with fixed center effects (1) or from the random effects model (2) with either a gamma, positive stable or inverse Gaussian frailty model. To make the model comparable for the random effects models the parameters of the frailty model were chosen to give a Kendall's τ of either 0.1, 0.3, or 0.5 between individuals within a center. Note since the inverse Gaussian model has a τ of less than 0.5 only the $\tau = 0.1$ and 0.3 cases were available.

For the gamma frailty model the u_i were simulated from a gamma distribution with mean 1 and variance α using the IMSL routine rngam. This model has a value of $\tau = \alpha/(\alpha+1)$. For the inverse Gaussian distribution with probability density function $f(u) = (\eta\pi)^{-1/2}\exp\{2/\eta\}\exp\{-u/\eta - 1/(\eta u)\}$, the u_i were generated using the routine in Micheal et al [9]. For this model Kendall's τ is $0.5 - 2/\eta + (8/\eta^2)\exp\{4/\eta\}\int_{4/\eta}^{\infty}\exp(-u)/u du$. For the positive stable distribution with Laplace transform $\exp(-u^\rho)$, $0 \le \rho \le 1$, the u_i's were generated using results in Chambers et al [10]. Here Kendall's τ is $1-\rho$. For the fixed center effects model we model the center effect as $\theta_i = c(i-3)$, for $i = 1, \cdots, 5$ when $K = 5$ and as $\theta_i = c[-K-2+2i]/2$, $i = 1, \cdots, K/2$ and $\theta_i = c[i-K/2]$ for $i = K/2, \cdots, K$ when $K = 10$ or $K = 20$. To determine the value of c we treat the θ_i as arrising from a discrete distribution, E, with mass $1/K$ at each θ_i. Then the expected value of E is zero as is the expected value of ϵ_i in (2). To find c we match the variance of $\exp\{E\}$ with that of the variance of the gamma frailty distribution. This gives a "association" in the fixed effects model of roughly the same strength as in the gamma frailty model. Note that while we treated the center effect in a random manner to get a value of c, in the simulation the values of c are fixed. Table 1 summarizes the parameters used in our study.

Table 1
Parameters used in The Monte Carlo Study

Center Effect		τ=0.1	$\tau = 0.3$	$\tau = 0.5$
Gamma		$\alpha = 2/9$	$\alpha = 6/7$	$\alpha = 2$
Positive Stable		$\rho = 0.9$	$\rho = 0.7$	$\rho = 0.5$
Inverse Gaussian		$\eta = 0.551$	$\eta = 4.070$	Not Possible
Constant	K=5	c=0.311	c=0.534	c=0.709
	K=10	c=0.132	c=0.230	c=0.305
	K=20	c=0.007	c=0.122	c=0.161

For each sample we compute the Wald, likelihood ratio and score test for the fixed effects model, the score test for the random effects model and the estimate of the β based on a proportional hazards model which does not adjust for center effects and for the model which makes a fixed effect adjustment for the center effects. This is done in each run for 5,000 samples. We estimate the power of the four test of center effects at a 0.05 significance level and the bias and mean squared error of the two estimates of β.

3. Signficance levels of the tests

Table 2 shows the estimated null power of the likelihood ratio fixed effects test and the random effects score test, at a 0.05 significance level, based on 5,000 replicates for each combination of β, K and total sample size. Here we have reported only the likelihood ratio test for the fixed effects model since its performance was in all cases the best of the three possible fixed effects test statistics. From this table we first see that the test based on a fixed center effects model requires a very large sample size before it achieves the desired level. When the number of subjects at each center is small the test is anti-conservative. This fact appears to be true even when there are ten or more groups with 400 total observations and the results suggest that unless the number of subjects in each group is very large the fixed effect test should not be used because it rejects the hypothesis of no center effect too often when the null hypothesis is true.

For the random effects score test, with only a few exceptions, the nominal level of the test is achieved. When $K = 5$ and the total sample is 100 the test may be slightly anti-conservative, but the estimated power achieved is closer to 0.05 than for any of the fixed effects tests.

4. Behavior When There Is A Group Effect

As seen in the previous section the fixed effects test for a group effect tends to reject the null hypothesis of no group effect too often when the number of subjects per group is small. The random effects test does, however, appear to maintain the correct significance level for these small sample cases. In our examination of the power of these tests we found that the power of the fixed effects test was higher in all cases than the random effects test. However,

due to the problem with the fixed effects test when the null hypothesis is true these higher powers give a false impression that this test is performing better than the random effects test. Higher power is to be expected since the nominal significance levels of the fixed effects tests are higher than those of the random effects test.

To examine the power of the random effects tests we report in Table 3 the estimated power of the random effects tests for $\tau = 0.1$ and 0.3 for the gamma, inverse Gaussian, and positive stable frailty models and the fixed effects models. When $\tau = 0.5$ for the gamma, positive stable random effects models and for the fixed effects model, all tests essentially have a power of 1. From this table we see that the random effects test has good power to detect fixed group effects. The power is quite high for all types of group effects for small associations between individuals within a group when the total sample size is large or the number per group is large. For a given number of groups and a given total sample size the power decreases as the censoring fraction increases.

While the random effects test for group effects has reasonable power to detect these effects a natural question is whether the presence of a group effect has an effect on the estimate of treatment efficacy. To examine this question we studied the relative excess bias in estimating β in a model that ignores the center effect when such an effect exists. We computed for each combination of the total sample size N, number of groups K, the degree of association, τ, and group effect β the quantity

$$r = \frac{B(0) - B(\tau)}{|B(0)|}, \tag{3}$$

where $B(\tau)$ is the estimator of the bias of the estimator of β based on a model which ignores the center effect. Here $B(0)$ is from data simulated from a model with no center effect.

We analysed these data using ANOVA techniques as in Andersen et al. [11]. Separate analyses were made for $\beta = 0$ and $\beta = \ln(2)$ and we included the factors $N * K$, τ, percent CENSoring, and DISTribution of the center effects since inclusion of more interactions did not improve the fit of the model. That is, the model used for both values of β was

$$E(r) = \alpha + \beta_{N*K} + \gamma_\tau + \delta_{\text{CENS}} + \varepsilon_{\text{DIST}}.$$

For $\beta = 0$ none of these factors had any significant effect on r which, as one would expect, was small in all cases. For $\beta = \ln(2)$, $E(r)$ was everywhere larger than for $\beta = 0$. That is because under the random effects models and apparantly under the constant effects models as well the estimates computed without adjustment for center effects tend to shrink towards zero. Furthermore, $E(r)$ increased in absolute value as the strength of association, τ, increases. Thus the averages over the other factors in the model were -17.2, -45.2, and -70.9 for τ =0.1, 0.3, and 0.5, respectively. The amount of censoring had no effect and the type of distribution and the number of groups, K, had little effect on r whereas $E(r)$ increased in absolute value when the total sample size, N, increases, the averages over the other factors in the model being -19.3, -41.7, and -72.7 for $N = 100$, 200 and 400, respectively.

This suggests that the estimators computed by ignoring either a fixed or random effect are inconsistent. It implies that the so called marginal approach of Lee et al [12] or Wei et al [13] which computes the estimate of β under an independent working model and uses a robust variance estimator is not appropriate in this problem.

5. Example

To illustrate the tests we consider a sample of 609 Acute Myelogenous Leukemia (AML) patients reported to the International Bone Marrow Transplant registry (IBMTR). All patients were given an HLA-identical sibling transplant for the leukemia which was in their first complete remission at the time of transplant. The IBMTR is an international cooperative group which collects data on allogenic transplants conducted world wide. The sample here consists of data reported by the 60 largest reporting centers over the period 1988-1994. Each center contributed at least 5 transplants to the study and had at least one patient relapsing or dying. Table 4 shows the distribution of the number of cases per center.

The goal of the study was to model the relationship between the patient's age (dichotomized as $\leq$30 versus >30) and Karnofski score (<90 versus $\geq$90) at the time of transplant and treatment failure. The treatment is said to fail if the patient dies or relapses. Ignoring any possible center effects the estimates of the risk coefficients were 0.26 (se=0.13, p=0.05) for the effect of being over thirty at transplant and 0.32 (se=0.17, p=0.07) for having a Karnofski score under 90. The four tests for a possible group effect give the following results:

Fixed Effects	
Likelihood ratio Test	p=0.228
Score Test	p=0.008
Wald Test	p=0.104
Random Effects	p=0.996

Note that the fixed effects score test suggests the presence of a center effect while the other tests do not show evidence of a center effect. In light of our simulation results which show that the score test rejects too often we conclude that their is no need here to adjust for a center effect. Note that if we had chosen to adjusts for a fixed center effect then the estimates for the risk coefficients would be 0.33 (se=0.15, p=0.024) for age and 0.25 (se=0.22, p=0.249) for Karnofski score which would lead to somewhat different conclusions than the model without a group effect.

6. Discussion

Our Monte Carlo study has shown that the use of a fixed effects model to test for a center effect in a small to moderate size multi-center trial tells us too often that an adjustment for such an effect is needed when in fact there is no such effect. This test requires a large number

of subjects in each center to give significance levels close to the nominal level. The sample sizes needed in each center are much larger than what is commonly encountered in practice. The random effects test of Commenges and Andersen [1] seems to behave quite well under the null hypothesis of center effect even when the number of observations in each group is fairly small and it seems to have reasonable power to detect either a fixed or random group effect.

The random effects test has a few additional advantages over the fixed effects test. First, the estimates of the center effect in the fixed effects proportional regression model requires at least one event for each center. When this does not hold the estimates do not exist. This restriction is not required for the random effects model. Second, when all the events in one center occur before (or after) all the events at an other center then the estimates of that center's fixed effect is at minus infinity (or plus infinity). Again this is not a problem for the random effects test. Finally, the Wald and likelihood ratio tests for fixed effects test requires the maximization of a log likelihood which is a function of $p + (K - 1)$ parameters, where p is the number of patient specific covariates. When there is a large number of centers this may be a large number of parameters and numerical problems may occur if good starting values are not used. Note that the random effects test requires maximization with respect to only p covariates.

When the presence of a center effect is detected then the natural question arrises as to how adjust for this effect. As noted earlier some adjustment is needed since the presence of a center effect, either fixed or random, makes the estimators of the risk coefficients computed under an assumption of no center effect inconsistent. The suggestion of Liang et al. [14] to use an independence working model in this case and a robust estimator of the variance of the estimator is not appropriate since the estimators do not seem to be consistent in these cases.

Some model which incorporates the center effect is needed. One possibility is to use the fixed effects model for this adjustment. This model can be fit using standard statistical software. We looked at the relative excess bias (3) of this main effect adjusted for a fixed center effect as compared to the bias under the independence model (data not shown) and in this case, as opposed to the unadjusted relative bias studied above, the relative bias decreased as the sample size increases. This was true, not only when the fixed effects model is correct, but is also true when the random effect model is true. This suggests that this model may provide a quick means of making a crude adjustment for a center effect when the sample sizes are large. A second possibility would be to estimate the treatment effect in a Cox regression model *stratified* by center but then centers with no events would contribute no information to the estimate.

An alternative to using fixed effect models to adjust for a center effect would be to use a frailty model. The technology for fitting a proportional hazards model with a fixed effect can be found in Nielsen et al [8], Klein [6] and Andersen et al [15], for the gamma frailty model and Klein et al [16] for the inverse Gaussian model and Wang et al [17] for the positive stable model.

Acknowledgements

This research was supported by Grant 5 R01 CA54706-05 from the National Cancer Institute. The activities of the Danish Epidemiology Society are supported by a grant from the Danish National Research Foundation.

REFERENCES

1. Commenges, D. and Andersen, P. K. (1995). Score Test Of Homogeneity For Survival Data. *Lifetime Data Analysis,* **1**, 145-160.

2. Cox, D. R. (1972). Regression Models And Life-Tables (With Discussion). *Journal Of The Royal Statistical Society B*, **34**, 187-220.

3. Klein, J. P. and Moeschberger, M. L. (1997) *Survival Analysis: Techniques For Censored And Truncated Data,* Springer-Verlag, New York.

4. Andersen, P. K., Borgan, Ø., and Gill, R. D. and Keiding, N. (1993). *Statistical Models Based On Counting Processes.* Springer-Verlag, New York.

5. Clayton, D. G. (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.

6. Vaupel, J. W. , Manton, K.G. and Stallard, E. (1979) The Impact Of Hetrogeneneity In Individual Frailty And The Demographics Of Mortality. *Demography* , **16**, 439-454.

7. Klein, J. P. (1992). Semiparametric Estimation Of Random Effects Using The Cox Model Based On The Em Algorithm. *Biometrics,* **48**, 795-806.

8. Nielsen, G. G., Gill, R. D., Andersen, P. K., and Sørensen, T. I. A. (1992). A Counting Process Approach To Maximum Likelihood Estimation In Frailty Models. *Scandinavian Journal Of Statistics* , **19**, 25-43.

9. Micheal, J.R., Schucany, W.R. and Hass, R.W. (1976) Generating Random Variates Using Transformations With Multiple Roots, *The American Statistician* , **30**, 88-90.

10. Chambers, J.M., Mallows, C. L. and Stuck, B.W. (1976) A Method For Simulating Positive Stable Random Variates. *Journal of the American Statistical Association* ,**71**, 340-350.

11. Andersen, P.K., Bentzon, M.W., and Klein, J.P. (1996) Estimating The Survival Function In the Proportional Hazards Regression Model: A Study Of The Small Sample Size Properties. *Scandinavian Journal Of Statistics*, **23**, 1-12.

12. Lee, E. W., Wei, L. J., and Amato, D. A. (1992). Cox-Type Regression Analysis For Large Numbers Of Small Groups Of Correlated Failure Time Observations. In J. P. Klein and P. Goel, Editors, *Survival Analysis: State Of The Art*, pp. 237-248. Kluwer Academic Publishers, Boston.

13. Wei, L. J., Lin, D. Y., and Weissfeld, L. (1989). Regression Analysis Of Multivariate Incomplete Failure Time Data By Modeling Marginal Distributions. *Journal Of The American Statistical Association,* **84**, 1065-1073.

14. Liang, K-Y. Self, S.G. and Chang, Y-C (1993) Modelling Marginal Hazards in Multivariate Failure Time Data. *Journal of the Royal Statistical Society* ,**B55**, 441-453.
15. Andersen, P.K., Klein, J.P., Kunudnsen, K.M. and Tabanera y Palacios, R. Estimation Of Variance In Cox's Regression Model With Gamma Frailties. Technical Report #10, Division of Biostatistics, The Medical College of Wisconsin, 1995.
16. Klein, J. P., Moeschberger, M. L., Li, Y. H. and Wang S.T. (1992). Estimating Random Effects In The Framingham Heart Study. In J. P. Klein and P. Goel, Editors, *Survival Analysis: State Of The Art*, pp. 99-120. Kluwer Academic Publishers, Boston.
17. Wang, S.T., Klein, J. P. and Moeschberger, M. L. (1995). Semi-Parametric Estimation of Covariate Effects Using The Positive Stable Frailty Model. *Applied Stochastic Models and Data Analysis*, **11**, 121-133.

Table 2. Estimated Null Power Of The Fixed and Random Effects Tests

Total Sample Size	Number of Groups	Percent Deaths	β	Likelihood Ratio	Random Effects Test
100	5	0.7	0.00	0.0674**	0.0588*
100	5	0.7	0.69	0.0540	0.0538
100	5	0.9	0.00	0.0670**	0.0626**
100	5	0.9	0.69	0.0590*	0.0582*
100	10	0.7	0.00	0.0802**	0.0610**
100	10	0.7	0.69	0.0774**	0.0532
100	10	0.9	0.00	0.0860**	0.0592*
100	10	0.9	0.69	0.0816**	0.0498
100	20	0.7	0.00	0.1592**	0.0590*
100	20	0.7	0.69	0.1486**	0.0526
100	20	0.9	0.00	0.1996**	0.0556
100	20	0.9	0.69	0.1494**	0.0552
200	5	0.7	0.00	0.0590*	0.0608**
200	5	0.7	0.69	0.0566*	0.0564*
200	5	0.9	0.00	0.0640**	0.0556
200	5	0.9	0.69	0.0520	0.0530
200	10	0.7	0.00	0.0704**	0.0572*
200	10	0.7	0.69	0.0600**	0.0568*
200	10	0.9	0.00	0.0690**	0.0508
200	10	0.9	0.69	0.0652**	0.0564*
200	20	0.7	0.00	0.0948**	0.0556
200	20	0.7	0.69	0.0956**	0.0544
200	20	0.9	0.00	0.1032**	0.0578*
200	20	0.9	0.69	0.0926**	0.0508
400	5	0.7	0.00	0.0550	0.0528
400	5	0.7	0.69	0.0508	0.0566*
400	5	0.9	0.00	0.0564*	0.0566*
400	5	0.9	0.69	0.0556	0.0626**
400	10	0.7	0.00	0.0542	0.0556
400	10	0.7	0.69	0.0604**	0.0580*
400	10	0.9	0.00	0.0584*	0.0490
400	10	0.9	0.69	0.0562*	0.0500
400	20	0.7	0.00	0.0670**	0.0462
400	20	0.7	0.69	0.0690**	0.0534
400	20	0.9	0.00	0.0736**	0.0506
400	20	0.9	0.69	0.0668**	0.0470

$**$-more than 3 SE larger than the nominal level

$*$- 2-3 SE larger than the nominal level.

Table 3. Power Of The Random Effects Test For Group Effects

				Constant		Gamma		Inverse Gaussian		Positive Stable	
				70%	90%	70%	90%	70%	90%	70%	90%
N	K	β	τ	Dead	Dead	Dead	Dead	Dead	Dead	Dead	Dead
100	5	0.00	0.1	0.659	0.784	0.538	0.618	0.555	0.644	0.446	0.486
100	5	0.00	0.3	0.995	1.000	0.909	0.942	0.903	0.952	0.891	0.921
100	5	0.69	0.1	0.638	0.767	0.541	0.620	0.542	0.630	0.444	0.477
100	5	0.69	0.3	0.994	1.000	0.903	0.944	0.902	0.953	0.890	0.911
100	10	0.00	0.1	0.495	0.619	0.502	0.619	0.485	0.608	0.454	0.466
100	10	0.00	0.3	0.979	0.996	0.944	0.984	0.948	0.986	0.947	0.963
100	10	0.69	0.1	0.470	0.612	0.495	0.608	0.481	0.598	0.440	0.454
100	10	0.69	0.3	0.972	0.996	0.939	0.979	0.944	0.983	0.943	0.964
100	20	0.00	0.1	0.278	0.386	0.301	0.501	0.297	0.451	0.289	0.303
100	20	0.00	0.3	0.850	0.937	0.839	0.978	0.871	0.977	0.923	0.948
100	20	0.69	0.1	0.271	0.381	0.313	0.492	0.299	0.431	0.272	0.278
100	20	0.69	0.3	0.841	0.938	0.861	0.980	0.881	0.978	0.924	0.942
200	5	0.00	0.1	0.967	0.994	0.784	0.846	0.790	0.853	0.624	0.680
200	5	0.00	0.3	1.000	1.000	0.975	0.986	0.969	0.986	0.960	0.977
200	5	0.69	0.1	0.954	0.987	0.777	0.842	0.778	0.845	0.610	0.661
200	5	0.69	0.3	1.000	1.000	0.971	0.979	0.971	0.987	0.955	0.974
200	10	0.00	0.1	0.902	0.973	0.827	0.894	0.824	0.901	0.689	0.730
200	10	0.00	0.3	1.000	1.000	0.996	0.999	0.995	0.999	0.990	0.997
200	10	0.69	0.1	0.897	0.964	0.814	0.886	0.826	0.895	0.672	0.717
200	10	0.69	0.3	1.000	1.000	0.995	0.998	0.997	0.999	0.991	0.995
200	20	0.00	0.1	0.767	0.875	0.767	0.877	0.776	0.876	0.659	0.696
200	20	0.00	0.3	1.000	1.000	0.997	1.000	0.999	1.000	0.998	0.999
200	20	0.69	0.1	0.741	0.866	0.757	0.871	0.746	0.867	0.651	0.671
200	20	0.69	0.3	1.000	1.000	0.998	1.000	0.998	1.000	0.997	1.000
400	5	0.00	0.1	1.000	1.000	0.920	0.942	0.927	0.952	0.766	0.814
400	5	0.00	0.3	1.000	1.000	0.993	0.996	0.992	0.995	0.987	0.993
400	5	0.69	0.1	1.000	1.000	0.924	0.947	0.920	0.949	0.765	0.809
400	5	0.69	0.3	1.000	1.000	0.993	0.997	0.991	0.996	0.985	0.993
400	10	0.00	0.1	0.999	1.000	0.974	0.990	0.974	0.990	0.861	0.891
400	10	0.00	0.3	1.000	1.000	1.000	1.000	1.000	1.000	0.999	1.000
400	10	0.69	0.1	0.999	1.000	0.964	0.983	0.972	0.988	0.854	0.895
400	10	0.69	0.3	1.000	1.000	1.000	1.000	1.000	1.000	0.999	1.000
400	20	0.00	0.1	0.995	1.000	0.983	0.992	0.983	0.994	0.900	0.924
400	20	0.00	0.3	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
400	20	0.69	0.1	0.993	0.999	0.979	0.992	0.979	0.993	0.895	0.921
400	20	0.69	0.3	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

Table 4. The Distribution Of The Number Of Cases Per Center

Number Of Cases	Number Of Centers
5	11
6	13
7	7
8	6
10	2
11	5
12	2
13	1
14	1
15	2
17	1
18	1
19	2
20	2
22	1
26	1
28	1
34	1

Appendix

For the jth subject $j = 1, \ldots, S_i$ in the ith group $i = 1, \ldots, n$ let T_{ij} be the observation time of subject and $D_{ij} = 1$ if subject died and 0 otherwise. The frailty model (2) proposed in Section 1 can be specified as a counting process $N_{ij} = I(T_{ij} \leq t, D_{ij} = 1)$ with

$$dN_{ij}(s) = dM_{ij}(s) + Y_{ij}(s) \exp\{\sigma\varepsilon_i \ + \ \beta^{'} z_{ij}\} \ \lambda_0(s) \ ds,$$

where $Y_{ij}(s) = I(T_{ij} > s)$, $M_{ij}(\cdot)$ is a martingale, ε_i's are iid random variables with an unspecified distribution G which has mean 0 and variance 1.

Let $\bar{N} = \sum_{i,j} N_{ij}$, $S^{(0)}(\beta, s) = \sum_{i,j} Y_{ij}(s) \exp(\beta^{'} z_{ij})$, and $\widehat{\beta}$ be the maximum partial likelihood estimate of β under the null hypothesis of $\sigma = 0$. The cumulative baseline hazard function $\Lambda_0(t) = \int_0^t \lambda_0(s)ds$ can be estimated by

$$\widehat{\Lambda}_0(t) = \int_0^t \frac{d\bar{N}(s)}{S^{(0)}(\widehat{\beta}, s)}.$$

Then the martingale $M_{ij}(t)$ can be estimated as

$$\widehat{M}_{ij}(t) = N_{ij}(t) - \widehat{\Lambda}_{ij}(\widehat{\beta}, t),$$

where $\widehat{\Lambda}_{ij}(\beta, t) = \exp(\beta' z_{ij}) \ \widehat{\Lambda}_0(t)$.

Let $p_{ij}(\beta, s) = Y_{ij}(s) \exp(\beta' z_{ij}) / S^{(0)}(\beta, s)$ and $p_i(\beta, s) = \sum_j p_{ij}(\beta, s)$. To test the hypothesis of homogeneity of $\sigma = 0$, the score test statistic is given by

$$T(\widehat{\beta}) = \sum_{i=1}^{n} \left(\sum_{j=1}^{S_i} \widehat{M}_{ij}(t) \right)^2 - \bar{N}(\infty) + \int_0^\infty \sum_{i=1}^{n} p_i^2(\widehat{\beta}, s) d\bar{N}(s).$$

Let $H_i(\beta, s) = 2 \left\{ \widehat{M}_i(s) - \sum_{l=1}^{n} \widehat{M}_l(s-) p_l(\beta, s) - p_i(\beta, s) + \sum_{l=1}^{n} p_l^2(\beta, s) \right\}$, where $\widehat{M}_i(s) = \sum_j \widehat{M}_{ij}(s)$. The variance of $T(\widehat{\beta})$ can be consistently estimated by

$$\widehat{I}_c = \widehat{I}(\widehat{\beta}) - \widehat{J}(\widehat{\beta}) \, I_{\widehat{\beta}}^{-1} \, \widehat{J}(\widehat{\beta})',$$

where $I_{\widehat{\beta}}^{-1}$is the information matrix relative to $\widehat{\beta}$,

$$\widehat{I}(\beta) = \sum_{i=1}^{n} \int_0^\infty H_i^2(\beta, s) \, p_i(\beta, s) \ d\bar{N}(s), \text{ and}$$

$$\widehat{J}(\beta) = \sum_{i=1}^{n} \int_0^\infty H_i(\beta, s) \sum_{j=1}^{S_i} z_{ij} \, p_{ij}(\beta, s) \ d\bar{N}(s).$$

Then the test statistic for homogeneity is $H = T(\widehat{\beta}) / \sqrt{\widehat{I}_c}$ which has an asymptotic standard normal distribution under the null hypothesis.