

**Testin For Center Effects In
ulticenter Sur i al Studies:
A Monte Carlo Comparison Of
Fixed And Rando Effects Tests**

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

Technical Report 23

January 1997

Division of Biostatistics
Medical College of Wisconsin
8701 Watertown Plain 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 $\mu\alpha c$

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 \mathbf{Z} denote the treatment and patient specific covariates and $X_i = \{1 \text{ if the patient is from institution } i; 0 \text{ otherwise}\}$ (and) $X_j = \{1 \text{ if the patient is from institution } j; 0 \text{ otherwise}\}$

$$\lambda(t) = \lambda_0(t) \exp(\beta' \mathbf{Z} + \alpha_1 X_1 + \alpha_2 X_2 + \dots + \alpha_k X_k)$$

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 . 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 approximately 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 λ_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 $f(x) = (\eta\pi)^{-1/2} \exp\{-2/\eta\} \exp\{-x/\eta - 1/(\eta x)\}$, the λ_i were generated using the routine in Micheal et al [9]. For this model Kendall's τ is $0.5 - 2/\eta + (8/\eta^2) \exp\{4/\eta\} \int_{4/\eta}^{\infty} \exp(-x) / dx$. For the positive stable distribution with Laplace transform $\exp(-\rho)$, $0 \leq \rho \leq 1$, the λ_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 follows

Table 1
Parameters used in The Monte Carlo Study

Center Effect		$\tau=0.1$	$\tau = 0.3$	$\tau = 0.5$
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. Significance 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 , τ 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 $\tau = 5$ 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

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 allogeneic 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 ≤ 30 versus > 30) and Karnofsky score (< 90 versus ≥ 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 Karnofsky score under 90. The four tests for a possible group

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 arises 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.

Acknowledgemen

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 Co '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

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

Total Sample Size	Number of Groups	Percent Deaths	β	Li elihood 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

N	β	τ	Constant		Gamma		Inverse Gaussian		Positive Stable		
			70% Dead	90% Dead	70% Dead	90% Dead	70% Dead	90% Dead	70% Dead	90% 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 j th subject $j = 1, \dots, S_i$ in the i th group $i = 1, \dots, n$ 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

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

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 which has mean 0 and variance 1.

Let $\bar{N} = \sum_{i,j} N_{ij}$, $S^{(0)}(\beta, s) = \sum_{i,j} Y_{ij}(s) e^{-p(\beta'z_{ij})}$, and $\hat{\beta}$ 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

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

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

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

where $\hat{\Lambda}_{ij}(\beta, t) = e^{-p(\beta'z_{ij})} \hat{\Lambda}_0(t)$.

Let $p_{ij}(\beta, s) = Y_{ij}(s) e^{-p(\beta'z_{ij})} / S^{(0)}(\beta, s)$ 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

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

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

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

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

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

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

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