Biometrika (1988), **75**, 1, *pp.* 57–64
Printed in Great Britain

Tests for no treatment effect in randomized clinical trials

BY M. H. GAIL

Biostatistics Branch of the Division of Cancer Etiology, National Cancer Institute, Bethesda, Maryland 20892, U.S.A.

W. Y. TAN

Department of Mathematical Sciences, Memphis State University, Memphis, Tennessee 38152, U.S.A.

AND S. PIANTADOSI

Biometry Branch of the Division of Cancer Prevention and Control, National Cancer Institute, Bethesda, Maryland 20892, U.S.A.

SUMMARY

We propose a test of the null hypothesis of no treatment effect in a randomized clinical trial that is based on the randomization distribution of residuals. These residuals result from regressing the response on covariates, but not treatment. In contrast to model-based score tests, this procedure maintains nominal size when the model is misspecified, and, in particular, when relevant covariates are omitted from the regression. The efficiency of the procedure is evaluated for regressions with some, but not all, required covariates. For many generalized linear models and survival models, conventional model-based score tests are shown to have supranominal size when relevant covariates are omitted, but logistic regression and the proportional hazards model are robust.

Some key words: Analysis of covariance; Nonlinear regression; Randomization; Randomized clinical trial; Robust test.

1. INTRODUCTION

Fisher's (1935) theory of randomization yields a valid test of the null hypothesis of no treatment effect in a randomized clinical trial without recourse to any probability model and without regard to any covariates that may or may not be present, provided censoring of responses acts equally on both treatment groups. Nonetheless, most clinical trials analysts ignore the randomization distribution and rely on models that incorporate covariates. These models yield more precise estimates of treatment effect, as discussed by Fisher (1925), Cochran (1957) and Cox & McCullagh (1982), and may be applied even if each treatment group is subjected to a different noninformative censoring mechanism (Kalbfleisch & Prentice, 1980, § 5·2). These models are popular also because many analysts are not content to state the conclusions of a randomized clinical trial in terms of the long run significance level under hypothetical reallocations of responses to treatment. Instead, many analysts assess treatment imbalances for important prognostic factors and 'adjust' the analysis for known imbalances. The adjustment is based either on models for analysis of covariance (Cornfield, 1971) or on 'post-stratification' (Peto et al., 1977), which may be justified by the population model assumption that treatments are exchangeable within strata, under the null hypothesis.

If the regression model for covariates is misspecified, these adjustment procedures may lead to tests for no treatment effect with supranominal size. We propose, instead, a test based on the randomization distribution of residuals computed from the regression on covariates, but not on treatment. This procedure captures the efficiencies of analysis of covariance, yet retains nominal size if the regression is misspecified. We apply these ideas to generalized linear models and to certain survival models.

2. NOTATION AND ANALYSIS FOR THE EXPONENTIAL FAMILY

Following Gail, Wieand & Piantadosi (1984), suppose we have a sample of n patients, each with a set of important covariates X. Without loss of generality, we assume $E(X)=0$, and we denote the covariance of X by Ω. We assign treatment $T=1$ or $T=-1$ to each patient, independently of X and of previous assignments, so that $E(T)=0$ and var $(T)=1$. This method of simple randomization is practical for trials in which patients accrue gradually and yields the same large-sample variance estimates as the classical model that allocates exactly $\frac{1}{2}n$ patients to each treatment. We observe the response variable Y.

Following Nelder & Wedderburn (1972), suppose the log likelihood of Y, conditional on X and T, is

$$l=\kappa(\theta)[Y\gamma(\eta)-g\{\gamma(\eta)\}+R(Y)]+\psi(Y,\theta), \tag{2·1}$$

where $\kappa(\theta)>0$ is a known scale factor, γ is a function that links η with the natural parameter of the exponential family, and $\eta=\mu+T\alpha+X\beta$. The regression is thus $E(Y|T,X)\equiv h(\eta)=g'\{\gamma(\eta)\}$. We assume that the correct regression is

$$h(\eta)=h(\mu+T\alpha+X\beta)\equiv h(\mu+T\alpha+X_1\beta_1+X_2\beta_2), \tag{2·2}$$

where the $p\times 1$ vector β has been partitioned into components β_1 and β_2 associated respectively with covariate row vectors X_1 and X_2 and dimensions p_1 and p_2, with $p=p_1+p_2$.

To compute the score test for the null hypothesis H_0: $\alpha=0$, we let $\hat{\mu}_0$ and $\hat{\beta}_0$ represent solutions to

$$\sum\gamma'(\hat{\eta}_0)\{Y-h(\hat{\eta}_0)\}=0,\quad \sum\gamma'(\hat{\eta}_0)X\{Y-h(\hat{\eta}_0)\}=0, \tag{2·3}$$

where sums are over the n independent observations (Y,T,X). The notation $\eta_0=\mu+X\beta$ and $\hat{\eta}_0=\hat{\mu}_0+X\hat{\beta}_0$ will be used. Under H_0, $\hat{\mu}_0$ and $\hat{\beta}_0$ converge to μ and β. The score for testing $\alpha=0$ is

$$\partial l/\partial\alpha\equiv U\equiv\kappa(\theta)\sum T\gamma'(\hat{\eta}_0)\{Y-h(\hat{\eta}_0)\}. \tag{2·4}$$

At $\alpha=0$, the expected values of $\partial^2 l/\partial\alpha\partial\mu$ and $\partial^2 l/\partial\alpha\partial\beta$ are zero for each component of β. The model-based estimate of var $(n^{-\frac{1}{2}}U)$ is therefore

$$\hat{V}=-n^{-1}\sum\frac{\partial^2 l}{\partial\alpha^2}=\{\kappa(\theta)/n\}\sum\{\gamma'(\hat{\eta}_0)\}h'(\hat{\eta}_0). \tag{2·5}$$

In most cases, $\kappa(\theta)$ is known from the model. A model-based score test of H_0: $\alpha=0$ is obtained from the standardized deviate $Z=U(n\hat{V})^{-\frac{1}{2}}$, and one or two-sided rejection regions of the form $Z>C_\alpha$ or $|Z|>C_{\frac{1}{2}\alpha}$ may be used. Here C_α and $C_{\frac{1}{2}\alpha}$ are $1-\alpha$ and $1-\frac{1}{2}\alpha$ level quantiles of the standard normal distribution.

An alternative variance estimate is based on the randomization distribution of the n residuals

$$r = \kappa(\theta)\gamma'(\hat{\eta}_0)\{Y - h(\hat{\eta}_0)\}. \tag{2·6}$$

These residuals are functions of Y and X only, and, under H_0, these residuals are independent of T, because T is independent of both Y and X. Moreover, $\Sigma r = 0$ from (2·3). We regard the n residuals as fixed constants and compute the variance of $n^{-\frac{1}{2}}U = n^{-\frac{1}{2}}\Sigma rT$ under repeated randomizations as

$$V_r = \text{var}\,(n^{-\frac{1}{2}}U) = n^{-1}\sum r^2. \tag{2·7}$$

A test of H_0 is based on the standardized deviate $Z_r = U(nV_r)^{-\frac{1}{2}}$.

There follow some miscellaneous comments on this procedure.

The variance estimate V_r is valid whether or not the model (2·1)-(2·2) holds. In § 3, we show that omitting X_2 from (2·2) usually increases the size of the test based on Z to supranominal levels, whereas the test based on Z_r retains nominal size.

The statistic Z_r retains full efficiency under (2·1)-(2·2) because V_r is consistent for var $(n^{-\frac{1}{2}}U)$ under that model.

Other randomization schemes lead to slightly different variance estimates, V_r. For example, if n subjects had been randomly divided into the two groups $T = 1$ and $T = -1$ with exactly $\frac{1}{2}n$ subjects each, then n would be replaced by $n - 1$ in (2·7), as follows from standard arguments for sampling $\frac{1}{2}n$ residuals at random without replacement from the finite population of n residuals.

The randomization estimate V_r may also be justified as a 'robust' variance estimate, $n^{-1}\Sigma(\partial l/\partial\alpha)^2$, under a population model in which the n triplets (X, T, Y) are independent and identically distributed (Huber, 1967; Kent, 1982; Royall, 1986).

3. Effects of omitting needed covariates

3·1. *The size of tests*

Tests based on permutational variance estimates have valid size regardless of whether the model (2·1)-(2·2) is correctly specified, whereas tests based on model-based variance estimates often have supranominal size if the false model

$$h(\eta^*) = h(\mu^* + T^*\alpha + X_1\beta_1^*) \tag{3·1}$$

is used in place of (2·2). Under this model, equations like (2·3) with X_1 in place of X lead to estimates $\hat{\mu}_0^*$ and $\hat{\beta}_{10}^*$, the score U^* is obtained from (2·4) with $\hat{\eta}_0^* = \hat{\mu}_0^* + X_1\hat{\beta}_{10}^*$ in place of $\hat{\eta}_0$, and the model-based variance estimate, $\hat{V}^*$, is given by (2·5) with $\hat{\eta}_0^*$ in place of $\hat{\eta}_0$. The corresponding deviate $Z^* = U^*(n\hat{V}^*)^{-\frac{1}{2}}$ is not properly standardized. Indeed, under H_0: $\alpha = \alpha^* = 0$, the variance estimate $\hat{V}^*$ converges to

$$\{\kappa(\theta)\}E[\{\gamma'(\eta_0^*)\}h'(\eta_0^*)], \tag{3·2}$$

whereas the correct limiting null variance of $n^{-\frac{1}{2}}U^*$ is

$$\{\kappa(\theta)\}^2E[\{\gamma'(\eta_0^*)\}^2\{\kappa(\theta)\gamma'(\eta_0)\}^{-1}h'(\eta_0)] + \{\kappa(\theta)\}^2E[\{\gamma'(\eta_0^*)\}^2\{h(\eta_0) - h(\eta_0^*)\}^2]. \tag{3·3}$$

The quantities μ_0^* and β_{10}^* in η_0^* are constants to which $\hat{\mu}_0^*$ and $\hat{\beta}_{10}^*$ converge, and μ_0^* and β_{10}^* do not equal μ and β_1 in general. The expectations in (3·2) and (3·3) are over the distribution of X. We let V_r^* denote the permutational variance from (2·7) of the

residuals (2·6) computed under (3·1) with $\hat{\eta}_0^*$ in place of $\hat{\eta}_0$, and we define the properly standardized statistic $Z_r^* = U^*(nV_r^*)^{-\frac{1}{2}}$.

A special case is given in Table 1 where the true model has argument $\eta = \mu + X_2\beta_2$, and the false model has argument $\eta^* = \mu^*$. Here $\text{pr}(X_2 = 1) = \text{pr}(X_2 = -1) = 0{\cdot}5$. The terms additive, multiplicative, reciprocal and logistic in Table 1 describe the regression function $h(\eta)$. Simulations reveal that the size of the test based on Z^* may exceed the nominal 0·05 level substantially, especially for Poisson and exponential models. The ratio of (3·3) to (3·2), called the variance ratio, is a good guide to the size of the test based on Z^*.

For the normal model, the size of the test based on Z^* exceeds nominal levels when $\kappa(\theta) = \sigma^{-2} = 1$ is taken as known. When σ^2 is estimated from the residual sum of squares, a test equivalent to Z_r^* results. This procedure is discussed further in § 5.

One can prove that (3·2) equals (3·3) for all Bernoulli models, regardless of the form of $h(\eta)$, for the special case $\eta = \mu + X_2\beta_2$ and $\eta^* = \mu^*$. Thus, standard computer programs yield valid tests of no treatment effect under the model $\eta^* = \mu^* + T\alpha^*$ which omits all covariates. If partial adjustment is attempted via $\eta^* = \mu^* + T\alpha^* + X_1\beta_1^*$, when the full model (2·2) is needed, the variance ratio can exceed 1·0 very slightly, even for Bernoulli

Table 1. *Effects of omitting covariates on the size of the critical region* $|Z^*| > 1{\cdot}96$ *and on the efficiency of* Z_r^*

Model*	$\gamma(\eta)$	$h(\gamma)$	Efficiency of Z_r^* relative to Z_r	Variance ratio	Size
Poisson multiplicative	η	e^{η}			
$\mu = 0$			0·81	1·24	0·081‡
$\mu = 2$			0·36	2·78	0·231‡
Poisson additive	$\log \eta$	η			
$\mu = 1$			0·60	1·25	0·085‡
$\mu = 100$			1·00	1·00	0·052
Normal†	η	η			
$\mu = 0$			0·80	1·25	0·087‡
Exponential multiplicative†	$-e^{\eta}$	$e^{-\eta}$			
$\mu = 0$			0·70	1·43	0·130‡
Exponential reciprocal	$-\eta$	η^{-1}			
$\mu = 1$			0·83	1·50	0·104‡
$\mu = 100$			1·00	1·00	0·059
Exponential additive	$-\eta^{-1}$	η			
$\mu = 1$			0·30	1·50	0·117‡
$\mu = 100$			1·00	1·00	0·050
Bernoulli logistic	η	$e^{\eta}(1+e^{\eta})^{-1}$			
$\mu = 0$			0·94	1·00	0·052
$\mu = 2$			0·97	1·00	0·056
Bernoulli additive	$\log\{\eta/(1-\eta)\}$	η			
$\mu = 0{\cdot}5, \beta_2 = 0{\cdot}25$			0·75	1·00	0·050

* The quantities $\alpha = \beta_1 = 0$ in all models. Except for Bernoulli additive model, $\beta_2 = 0{\cdot}5$. Variance ratio is ratio of true variance (3·3) to false variance (3·2). Asymptotic relative efficiency is ratio of (3·5) to (3·4).

† Only $\mu = 0$ presented because results independent of μ. For normal model, variance $\sigma^2 = \kappa(\theta)^{-1} = 1$ assumed known.

‡ Size is proportion of times $|Z^*| > 1{\cdot}96$ in 1000 Monte Carlo replications. ‡ indicates values outside the interval (0·036, 0·064), which should contain 95% of estimates if size is 0·05.

models. For example, suppose (X_1, X_2) have joint probabilities 0·15, 0·05, 0·15, 0·05, 0·2 and 0·4 respectively for the outcomes $(-3, -1)$, $(-3, 1)$, $(0, -1)$, $(0, 1)$, $(1, -1)$ and $(1, 1)$. For the Bernoulli logistic model with $\eta = 1 + X_1 + X_2$, partial adjustment with $\eta^* = \mu^* + X_1\beta_1^*$ yields a variance ratio 1·0036. Such examples suggest that, for practical purposes, logistic regression yields tests of H_0 with nominal size, whether no covariates for a few covariates are included in the model.

3·2. *Efficiency*

Under the model (2·1)–(2·2) with α near 0, both Z^2 and Z_r^2 have noncentrality

$$n\kappa(\theta)\alpha^2[E\{\gamma'(\eta_0)h'(\eta_0)\}]^2[E\{\gamma'(\eta_0)\}h'(\eta_0)]^{-1}, \tag{3·4}$$

where expectations are over the distribution of X. If covariates X_2 are omitted, the statistic Z_r^* is asymptotically standard normal under H_0. The noncentrality of Z_r^{*2} for small α is

$$n\kappa(\theta)\alpha^2[E\{\gamma'(\eta_0^*)h'(\eta_0)\}]^2\{E\{\gamma'(\eta_0^*)\}^2[h'(\eta_0)\{\gamma'(\eta_0)\}^{-1} + \kappa(\theta)\{h(\eta_0) - h(\eta_0^*)\}^2]\}^{-1}, \tag{3·5}$$

and the asymptotic relative efficiency of a test of H_0 based on Z_r^* compared to one based on Z_r from the full model is given by the ratio of (3·5) to (3·4).

For the normal linear model with $h' = \gamma' = 1$ and $\text{var}(Y \mid X, T) = \sigma^2 = \{\kappa(\theta)\}^{-1}$, the ratio of (3·5) to (3·4) reduces to $\sigma^2(\sigma^2 + \beta_2^{\mathrm{T}}\Omega_{22.1}\beta_2)^{-1}$. For the special case in which X_1 is null, this is $1 - R^2$, where R^2 is the multiple correlation of Y on X_2. This result was discussed by Cochran (1957) and Cox & McCullagh (1982), who also made allowance for loss in degrees of freedom from estimating parameters.

Efficiency losses can be severe, as in Table 1, but logistic regression retains high efficiency.

These asymptotic relative efficiency calculations are in good agreement with efficiencies estimated empirically for samples of practical size. Unreported simulations with samples containing 25, 100 and 225 observations confirmed this fact for the Poisson multiplicative, exponential multiplicative and Bernoulli logistic models. Empirical efficiency estimates were obtained as did Lininger et al. (1979).

4. Proportional hazards analysis of survival data

Lagakos & Schoenfeld (1984) showed that the size of the test of H_0 based on the model of Cox (1972) retained nominal size when all covariates were omitted, and they mentioned simulations that suggested that near nominal size was also retained with partial adjustment. Substantial power losses were demonstrated with omitted covariates, however (Morgan, 1986).

Following Gail et al. (1984), we consider parametric models with known nuisance hazard, $\lambda_0(w)$ and known cumulative hazard $\Lambda_0(w)$. For a person with treatment T and covariates X, the hazard is assumed to satisfy $\lambda(w \mid T, X) = \lambda_0(w)\exp\{\gamma(\eta)\}$. We observe $Y = \min(W, C)$, where W is the survival time and C is a censoring time independent of W. Under the full model, the log likelihood is $l = D\gamma(\eta) - \Lambda_0(Y)\exp\{\gamma(\eta)\}$, where D is an indicator variable with value 1 if $W \leqslant C$ and 0 otherwise. The score statistic for testing $\alpha = 0$ is

$$U = \partial l/\partial\alpha = \sum T\gamma'(\hat{\eta}_0)[D - \Lambda_0(Y)\exp\{\gamma(\hat{\eta}_0)\}],$$

where $\hat{\eta}_0$ is obtained from solution to the score equations with $\alpha = 0$. Under H_0, the expectations of $\partial^2 l/\partial\alpha\partial\mu$ and $\partial^2 l/\partial\alpha\partial\beta$ are zero, and

$$E\left(-\frac{\partial^2 l}{\partial\alpha^2}\right) = E(-\gamma''(\eta_0)[D - \Lambda_0(Y)\exp\{\gamma(\eta_0)\}]) + E[\{\gamma'(\eta_0)\}^2\Lambda_0(Y)\exp\{\gamma(\eta_0)\}]$$
$$= E[D\{\gamma'(\eta_0)\}^2], \tag{4·1}$$

because $E(D) = E[\Lambda_0(Y)\exp\{\gamma(\eta)\}]$ (Breslow, 1978). A robust variance estimate may be obtained by applying (2·7) to the residuals $r = \gamma'(\hat{\eta}_0)[D - \Lambda_0(Y)\exp\{\gamma(\hat{\eta}_0)\}]$, provided the same censoring process acts on both treatment groups.

If X_2 is mistakenly omitted from the argument η_0, the variance estimate $D\{\gamma'(\hat{\eta}_0^*)\}^2$ will be too small, and the resulting statistic $Z^* = U^* D^{-\frac{1}{2}}\{\gamma'(\eta_0^*)\}^{-1}$ will produce tests with supranominal size. Theoretical calculations of the variance ratio, defined as in § 3, are available from the authors. Such variance ratios were calculated for exponential survival distributions with two types of censorship; see Table 2. All patients enter at time $t = 0$ and the data are analysed at $t = \tau$ for type I censoring. For uniform entry, patients enter at random on the interval $[0, \tau]$, and analysis is at time τ. The effect of omitting covariates is more severe for diseases with moderate hazard rates than for diseases with low hazard rates. Censoring, and especially type I censoring, makes the size of conventional score tests fairly robust to omitted covariates. Nonetheless, a cancer trial with slow accrual might correspond to the variance ratio entry 1·22 in Table 2, and the size of the corresponding nominal 0·05 level test of H_0 would be approximately $2[1 - \Phi\{1{\cdot}96(1{\cdot}22)^{-\frac{1}{2}}\}] = 0{\cdot}076$.

Table 2. *Theoretical variance ratios for censored exponential data**

Moderate hazard rates ($\mu = 0$)					
$E(D)$	0·228	0·380	0·782	0·887	1·000
Variance ratios					
Uniform entry	1·00	1·02	1·22	1·33	1·43
Type I censoring	1·00	1·01	1·09	1·18	1·43
Low hazard rates ($\mu = -4$)					
$100E(D)$	0·514	1·024	4·955	9·521	100·00
Variance ratios					
Uniform entry	1·00	1·00	1·00	1·00	1·43
Type I censoring	1·00	1·00	1·00	1·00	1·43

* Calculations assume hazard $\exp(\mu + X_2\beta_2)$ with $\beta_2 = 0{\cdot}5$ and $\text{pr}(X_2 = 1) = \text{pr}(X_2 = -1) = 0{\cdot}5$. This is exponential multiplicative model in Table 1. In notation of § 4, $\gamma(\eta) = \eta$, so that (4·1) reduces to $E(D)$, probability of death. Uniform entry on $[0, \tau]$ is for $\tau = 0{\cdot}5$, 1, 5, 10, ∞. For type I censoring at τ, values of τ chosen to match $E(D)$. For $\mu = 0$, values of τ were 0·236, 0·448, 1·645, 2·568 and ∞, whereas, for $\mu = -4$, values were 0·249, 0·499, 2·474, 4·898 and ∞.

5. DISCUSSION

The randomization statistic Z_r captures the efficiencies of analysis of covariance and is robust to the omission of needed covariates and other types of model misspecification, whereas the size of conventional score tests usually exceeds nominal levels if needed covariates are omitted. Standard logistic regression and proportional hazards regression are, however, robust to omission of needed covariates.

Blocked randomization within strata may achieve perfect treatment balance within strata. Omission of perfectly balanced stratum factors has qualitatively different effects on the size of Z than does omission of these factors in simply randomized trials. Work submitted for publication shows that conventional model-based score tests have size slightly below nominal levels for Bernoulli models, slightly above nominal levels for exponential models and at nominal levels for Poisson models when perfectly balanced covariates are omitted. Robust variance estimates can be obtained from the blocked stratified randomization distribution of the residuals.

Omission of needed covariates may be regarded as a source of 'overdispersion' in simply randomized trials. One approach is to treat omitted covariates as random effects. However, such techniques are computationally difficult and are still under development for generalized linear models, and such methods are often based on the unreasonable assumption that the omitted covariates are uncorrelated with covariates that are included in η^*. A second approach would be to multiply the model-based variance estimate $\hat{V}^*$ by the scale factor

$$(n-d)^{-1}\sum\{Y-h(\hat{\eta}^*)\}^2\{h'(\hat{\eta}^*)\}^{-1}\gamma'(\hat{\eta}^*)\kappa(\theta), \tag{5·1}$$

as in equation (8.12) of McCullagh & Nelder (1983), where $d=1+p_1$ is the number of estimated parameters in $\hat{\eta}^*$. This scale factor converges to the required ratio, $V_r^*/\hat{V}$, in two important cases: (i) no covariate adjustment is attempted, namely $\eta^*=\mu^*$, and (ii) the classical analysis of covariance model with $\gamma'=h'=1$ and $\kappa(\theta)=\sigma^{-2}$. The latter case explains why tests of nominal size are obtained with classical analysis of covariance, even when covariates are omitted. If σ^2 is unknown, the mean squared error estimate of σ^2 is equal to $nV_r^*/(n-d)$, which converges to V_r^*. Outside these two cases, however, one can anticipate from equation (10) of Cox (1983) that the simple rescaling by (5·1) will be an imperfect correction for overdispersion caused by omitted covariates, and the use of V_r^* provides a simple, reliable alternative.

Acknowledgements

We thank Jennifer Donaldson for typing the manuscript and the reviewer and Professor D. R. Cox for numerous clarifications.

References

Breslow, N. (1978). The proportional hazards model: Applications in epidemiology. *Comm. Statist.* A **7**, 315–32.

Cochran, W. G. (1957). Analysis of covariance: its nature and uses. *Biometrics* **13**, 261–81.

Cornfield, J. (1971). The University Group Diabetes Program: A further statistical analysis of the mortality findings. *J. Am. Med. Assoc.* **217**, 1676–87.

Cox, D. R. (1972). Regression models and life tables (with discussion). *J.R. Statist. Soc.* B **34**, 187–220.

Cox, D. R. (1983). Some remarks on overdispersion. *Biometrika* **70**, 269–74.

Cox, D. R. & McCullagh, P. (1982). Some aspects of analysis of covariance. *Biometrics* **38**, 541–61.

Fisher, R. A. (1925). *Statistical Methods for Research Workers.* Edinburgh: Oliver & Boyd.

Fisher, R. A. (1935). *The Design of Experiments.* Edinburgh: Oliver & Boyd.

Gail, M. H., Wieand, S. & Piantadosi, S. (1984). Biased estimates of treatment effect in randomized experiments with nonlinear regressions and omitted covariates. *Biometrika* **71**, 431–44.

Huber, P. J. (1967). The behavior of maximum likelihood estimates under nonstandard conditions. *Proc. 5th Berkeley Symp.* **1**, 221–33.

Kalbfleisch, J. D. & Prentice, R. L. (1980). *The Statistical Analysis of Failure Time Data.* New York: Wiley.

Kent, J. T. (1982). Robust properties of likelihood ratio tests. *Biometrika* **69**, 19–27.

LAGAKOS, S. W. & SCHOENFELD, D. A. (1984). Properties of proportional hazards score tests under misspecified regression models. *Biometrics* **40**, 1037–48.

LININGER, L., GAIL, M. H., GREEN, S. B. & BYAR, D. P. (1979). Comparison of four tests for equality of survival curves in the presence of stratification and censoring. *Biometrika* **66**, 419–28.

MCCULLAGH, P. & NELDER, J. A. (1983). *Generalized Linear Models.* London: Chapman and Hall.

MORGAN, T. M. (1986). Omitting covariates from the proportional hazards model. *Biometrics* **42**, 993–5.

NELDER, J. A. & WEDDERBURN, R. W. M. (1972). Generalized linear models. *J.R. Statist. Soc.* A **135**, 370–84.

PETO, R., PIKE, M. C., ARMITAGE, P., BRESLOW, N. E., COX, D. R., HOWARD, S. V., MANTEL, N., MCPHERSON, K., PETO, J. & SMITH, P. G. (1977). Design and analysis of randomized clinical trials requiring prolonged observation of each patient. Part 2. Analysis and examples. *Br. J. Cancer* **35**, 1–39.

ROYALL, R. M. (1986). Model robust confidence intervals using maximum likelihood estimates. *Int. Statist. Rev.* **54**, 221–6.

[*Received April* 1985. *Revised July* 1987]