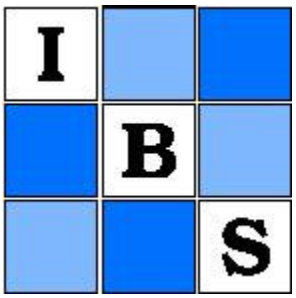

BIOMETRICS 45, 145–156
March 1989

Modelling Paired Survival Data with Covariates

William J. Huster

Biostatistics Department, Merrell-Dow Research Institute,
Merrell-Dow Pharmaceuticals, Cincinnati, Ohio 45215, U.S.A.

Ron Brookmeyer

Department of Biostatistics, The Johns Hopkins School of Hygiene and Public Health,
Baltimore, Maryland 21205, U.S.A.

and

Steven G. Self

Division of Public Health Sciences, The Fred Hutchinson Cancer Research Center,
Seattle, Washington 98104, U.S.A.

SUMMARY

The objective of this paper is to consider the parametric analysis of paired censored survival data when additional covariate information is available, as in the Diabetic Retinopathy Study, which assessed the effectiveness of laser photocoagulation in delaying loss of visual acuity. Our first approach is to extend the fully parametric model of Clayton (1978, *Biometrika* **65,** 141–151) to incorporate covariate information. Our second approach is to obtain parameter estimates from an independence working model together with robust variance estimates. The approaches are compared in terms of efficiency and computational considerations. A fundamental consideration in choosing a strategy for the analysis of paired survival data is whether the correlation within a pair is a nuisance parameter or a parameter of intrinsic scientific interest. The approaches are illustrated with the Diabetic Retinopathy Study.

1. Introduction

This paper focuses on modelling paired, censored, survival data when information on covariates is available. Examples are studies of familial tendency or of "naturally" paired systems, such as eyes. One approach is to extend Cox's proportional hazards regression model to the paired-data case (Holt and Prentice, 1974). In this semiparametric model, a separate nuisance hazard function is assumed for each pair and the covariate acts proportionally on that hazard for each pair. This formulation treats the association within a pair as a nuisance.

In contrast, Clayton (1978) and Oakes (1982) present a fully parametric model for paired survival data. In this formulation, the covariate is assumed to act proportionally on the marginal hazard and an additional parameter describes the association within a pair.

In this paper, we extend the Clayton–Oakes model to allow censoring and the incorporation of covariates (§2) and propose a simple alternative to this approach based on an independence working model together with a robust variance estimate (§3). Section 4 compares the efficiency of these two estimators. The techniques are illustrated in Section 5 using the Diabetic Retinopathy Study.

Key words: Censoring; Covariates; Paired survival data; Parametric modelling.

2. Parametric Modelling of Bivariate Survival Data

Let $S_1(t_1; \boldsymbol{\beta}) = \Pr(T_1 > t_1; \boldsymbol{\beta})$ and $S_2(t_2; \boldsymbol{\beta}) = \Pr(T_2 > t_2; \boldsymbol{\beta})$ denote the marginal survivor functions for each member of a pair, where (T_1, T_2) are the survival times of the members of a pair. The joint survivor function for the Clayton–Oakes model is

$$S(t_1, t_2; \boldsymbol{\beta}, \theta) = [S_1(t_1, \boldsymbol{\beta})^{1-\theta} + S_2(t_2, \boldsymbol{\beta})^{1-\theta} - 1]^{[-1/(\theta-1)]},$$

where θ (> 1) is the parameter measuring positive association between the survival times and $\boldsymbol{\beta}$ represents the vector of marginal parameters. The vector $\boldsymbol{\beta}$ may include parameters unique to each margin if the pairs are ordered. If the pairs are unordered, the marginal distributions will be identical. When $\theta = 1$, the survival times are independent.

The attractive features of this family are twofold: (1) any form of marginal distribution may be used, and (2) θ is separate from the parameters of the marginal distributions and has the following interpretation: If $h_a = h(t_1 \mid T_2 = t_2)$ is the hazard function of T_1 conditional on $T_2 = t_2$ and $h_b = h(t_1 \mid T_2 > t_2)$ is the hazard function of T_1 conditional on $T_2 > t_2$, then for all t_1 and t_2, $\theta = h_a/h_b$ (Clayton, 1978). Thus, θ can be interpreted as a relative risk—the risk of failing for member one if the second member fails relative to the risk for member one if the second member does not fail. The parameter θ can also be given a "random effects interpretation" (Clayton, 1978; Cox and Oakes, 1984, pp. 158–161).

Covariates can be introduced into the model by allowing the marginal distributions to be functions of the covariates. As an example, for an individual with the vector of covariates $\mathbf{X}$, Aitkin and Clayton (1980) parametrize the marginal Weibull survivor function as

$$S(t; \alpha, \boldsymbol{\beta}, \mathbf{X}) = \exp(-t^{\alpha} e^{\boldsymbol{\beta}'\mathbf{X}}),$$

where α is the shape parameter and $\boldsymbol{\beta}$ is the vector of parameters associated with the covariates. This formulation corresponds to modelling the effects of the covariates in a proportional hazards framework.

Likelihood-based methods of inference can be used to construct tests of hypotheses and to estimate the association and marginal parameters. Let $f_i(t_i; \boldsymbol{\beta})$, $S_i(t_i, \boldsymbol{\beta})$, $i = 1, 2$, be the marginal density and survival functions of the members of a pair, respectively; and let $f(t_1, t_2; \boldsymbol{\beta}, \theta)$, $S(t_1, t_2; \boldsymbol{\beta}, \theta)$ be the joint density and survivor functions, respectively. Let $g(C_1, C_2)$ be the joint density of C_1 and C_2, the censoring times for members of a pair. If $g(C_1, C_2)$ does not involve θ and $\boldsymbol{\beta}$, and the vector of the censoring times, (C_1, C_2), is independent of the vector of the survival times, (T_1, T_2), then the relevant contribution to the likelihood from one pair consists of one of four terms: If neither the first nor the second member failed, the contribution is $S(t_1, t_2)$; if both members failed, the contribution is $f(t_1, t_2)$; if the first member failed and the second member did not, the contribution is

$$\frac{S(t_1, t_2)^{\theta} f_1(t_1)}{S_1(t_1)}.$$

An analogous contribution is obtained if the second member failed and the first member did not. The relevant part of the full likelihood is just the product of the contributions from all pairs.

Estimates of the marginal and association parameters are found by maximizing the likelihood using constrained Newton–Raphson optimization techniques. Variance estimates are obtained from the observed information matrix.

3. Independence Working Model

If the purpose of the analysis is to make inferences about the marginal distributions while treating the dependence between pair members as a nuisance, a simple alternative to the complete parametric analysis described above is possible. We specify parametric models

for the marginal distributions but leave the nature of dependence between pair members completely unspecified. Parameters in the marginal models are then estimated by using the likelihood associated with the model presuming independence of the members (even though this assumption is incorrect). This likelihood is the product of marginal likelihoods over every individual in the data set. We call this the Independence Working Model (IWM).

If, in reality, there is dependence between members of the pair, one would be concerned that (1) the estimates of the marginal parameters would not be consistent for those of the bivariate distribution, and (2) the estimate of the variance of the estimates of the marginal parameters based on the second derivatives of the IWM likelihood would not be consistent.

Let T_i be the survival time random variables having parametric marginal survivor functions S_i, with parameter vector $\boldsymbol{\beta}$ ($i = 1, 2$ pair members). Assuming (incorrectly) independence of the pair members, the likelihood contribution for one pair is

$$L_{\mathrm{IWM}}(\boldsymbol{\beta}) = L_{\mathrm{IWM}}(t_1, t_2; \boldsymbol{\beta}) = f_1(t_1, \boldsymbol{\beta})^{\delta_1} S_1(t_1, \boldsymbol{\beta})^{1-\delta_1} f_2(t_2, \boldsymbol{\beta})^{\delta_2} S_2(t_2, \boldsymbol{\beta})^{1-\delta_2},$$

where f_i are the marginal densities, t_i are the observed survival times, and δ_i are the censoring indicators ($i = 1, 2$). The IWM likelihood is then the product of these terms over all pairs.

If the marginal distributions of the IWM likelihood are the same as those of the true bivariate model, then the IWM estimates, $\hat{\boldsymbol{\beta}}$, are consistent (see Appendix 1). This is true regardless of the nature of dependence between pair members.

Even though $\hat{\boldsymbol{\beta}}$ is consistent for any bivariate model, the usual likelihood-based estimates of the variance of $\hat{\boldsymbol{\beta}}$ are not generally consistent. Given a consistent estimator of $\hat{\boldsymbol{\beta}}$, Royall (1986) has presented an approach to variance estimation that is robust to model failure. Following his approach and with additional mild regularity conditions, one can easily show that $\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})$ converges in distribution as $n \to \infty$ to a Gaussian distribution with mean zero and covariance matrix given by the limiting value of

$$\boldsymbol{\Lambda}(\boldsymbol{\beta}) = n\boldsymbol{\Upsilon}(\boldsymbol{\beta})^{-1}\mathrm{E}\{\mathbf{U}(\boldsymbol{\beta})[\mathbf{U}(\boldsymbol{\beta})]^{\mathrm{T}}\}\boldsymbol{\Upsilon}(\boldsymbol{\beta})^{-1},$$

where $\mathbf{U}$ ($=\partial \ln L_{\mathrm{IWM}}(\boldsymbol{\beta})/\partial\boldsymbol{\beta}$) and $\boldsymbol{\Upsilon}$ ($=\mathrm{E}[-\partial^2 \ln L_{\mathrm{IWM}}(\boldsymbol{\beta})/\partial^2\boldsymbol{\beta}]$) are the score vector and expected information matrix, respectively. The expectations are taken with respect to the bivariate model.

In practice, one can estimate $\boldsymbol{\Lambda}(\boldsymbol{\beta})$ by

$$\hat{\boldsymbol{\Lambda}}(\hat{\boldsymbol{\beta}}) = n\mathbf{I}(\hat{\boldsymbol{\beta}})^{-1}\{\textstyle\sum \mathbf{U}_i(\hat{\boldsymbol{\beta}})\mathbf{U}_i(\hat{\boldsymbol{\beta}})^{\mathrm{T}}\}\mathbf{I}(\hat{\boldsymbol{\beta}})^{-1},$$

where $\mathbf{U}_i$ is the contribution of the ith pair to the score vector and $\mathbf{I}$ is the matrix of second derivatives of $\ln(L_{\mathrm{IWM}})$. One can then use the IWM estimators with their robust variances to make inferences about the marginal survival time distributions.

This discussion on the IWM applies generally to bivariate dependence models provided the marginal distributions have been correctly specified. For example, the IWM approach can be applied to bivariate models in Crowder (1985), Hougaard (1986), and Johnson and Kotz (1972).

4. Efficiency Considerations

The asymptotic relative efficiency (ARE) can be used as a measure of the amount of precision lost by the use of the IWM instead of the true bivariate model. The ARE for a given parameter is the ratio of the appropriate diagonal element of the inverse of the expected information matrix for the true bivariate model to the corresponding term in the matrix $\boldsymbol{\Lambda}(\boldsymbol{\beta})$ for the IWM given above.

As an example, assume the margins are exponential with a binary covariate but no censoring. The marginal hazard is $\lambda(t) = \alpha \exp(\beta X)$ where X is, say, a treatment indicator. In this case, the formula for the ARE is given in Appendix 2. Figure 1 displays the ARE for estimators of the treatment effect ($\hat{\beta}$) as a function of θ and of the percentage of

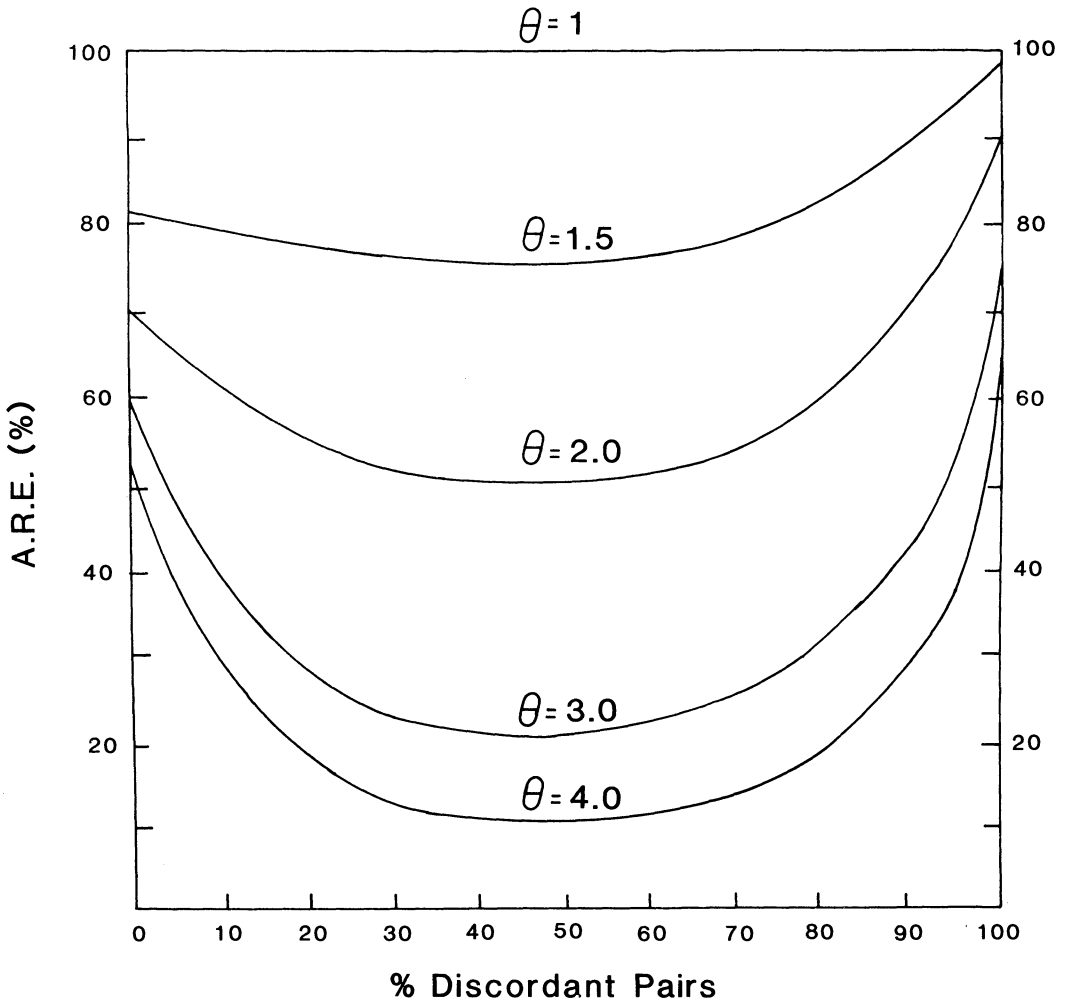

Figure 1. Asymptotic relative efficiency (ARE) of the estimator of the treatment effect from the bivariate model to the estimator from the independence working model (IWM), as a function of θ (the association parameter) and the percentage of discordant pairs.

discordant pairs. A discordant pair is a pair whose members have different (discordant) covariate vectors. A concordant pair is a pair whose members have identical (concordant) covariate vectors.

For a particular value of θ, the ARE is a nonsymmetric U-shaped function of the percentage of discordant pairs. The ARE of the IWM estimator is the greatest when the discordance is 100%. For example, there is only a 10% loss of efficiency when $\theta = 2$, which corresponds to a correlation of .64 between the survival times of the pair members in this setting. For 100% concordance, the IWM estimator loses only 30% efficiency when $\theta = 2$. However, for intermediate values of discordance, the IWM estimator does the worst, with a loss of 50% efficiency when $\theta = 2$.

Some insight into why Figure 1 is U-shaped can be gleaned by considering the problem of estimating the treatment effect from paired normally distributed data. That is, let Y_1 and Y_2 have a bivariate normal distribution:

$$\begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} \sim \text{BVN}\left(\begin{pmatrix} \beta X_1 \\ \beta X_2 \end{pmatrix}, \begin{bmatrix} \sigma^2 & \sigma_{12}^2 \\ \sigma_{12}^2 & \sigma^2 \end{bmatrix}\right),$$

where X_1, X_2 are treatment indicators (1 and -1), and β measures the treatment effect. When there is 100% discordance, the maximum likelihood estimator (MLE) of the treatment effect is the difference between the sample means. When there is 100% concordance, the MLE is the difference between the mean of all individuals with $X = 1$ and the mean of all individuals with $X = -1$. In both cases, 100% concordance and 100% discordance, the MLEs correspond to the IWM estimators. When there is 50% concordance, the IWM estimator equals the MLE only when the correlation is .50. Therefore, the ARE will be less than 1.0. This simple example suggests that the ARE of the IWM estimator relative to the MLE has a U-shaped curve as a function of the percentage of discordant pairs.

5. Example: The Diabetic Retinopathy Study

Diabetic retinopathy is a complication associated with diabetes mellitus consisting of abnormalities in the microvasculature within the retina of the eye. It is the leading cause of new cases of blindness in patients under 60 years of age in the United States and is the major cause of visual loss elsewhere in many industrialized countries (Murphy and Patz, 1978).

The Diabetic Retinopathy Study (DRS) was begun in 1971 to study the effectiveness of laser photocoagulation in delaying the onset of blindness in patients with diabetic retinopathy.

Patients with diabetic retinopathy in both eyes and visual acuity of 20/100 or better in both eyes were eligible for the study. One eye of each patient was randomly selected for treatment and the other eye was observed without treatment.

The endpoint used to assess the treatment effect was the occurrence of visual acuity less than 5/200 at two consecutively completed 4-month follow-ups. The total study size was 1,742 patients followed over several years. A 50% sample of the high-risk patients as defined by DRS criteria is the analysis subset ($N = 197$). On the basis of DRS findings, approximately one-half of the untreated eyes and one-quarter of the treated eyes should achieve the outcome after 5 years of follow-up.

The primary question of the DRS study was to assess the effectiveness of the laser photocoagulation treatment. Secondary questions were whether the survival times for the eyes of a patient were related and whether the treatment and type of diabetes were related.

Since the outcome is defined as occurring on two consecutive visits, there is an inherent lag time built into the model. This lag time was estimated by the smallest time-to-event, which was 6.5 months. Hence, all times have this lag subtracted from them in the analysis.

The exponential and Weibull distributions were considered as possible marginal distributions. Tables 1a and 1b present the maximum likelihood results when these distributions are fit to the data in the independence and nonindependence cases. The best model in the independence as well as nonindependence case was the Weibull having margins with the same shape but differing scale parameters. Exponential and Weibull probability plots as well as generalized residual plots (Cox and Snell, 1968) were used to check the goodness of fit of the marginals.

There is a proportional hazards interpretation of this result. If $h_1(t) = \alpha t^{\alpha-1}\exp(\beta_1)$ and $h_2(t) = \alpha t^{\alpha-1}\exp(\beta_2)$ are the marginal hazard functions for the treated and untreated eyes, respectively, then they are related by

$$h_1(t) = h_2(t)\exp(\beta_1 - \beta_2).$$

That is, the hazard of diabetic retinopathy in the treated eye is proportional to that in the untreated eye with a relative risk given by $\exp(\beta_1 - \beta_2)$.

The first hypothesis of interest is whether there is a treatment effect. Statistically, this question translates into whether the marginal distributions are the same—that is, whether

Table 1a
Maximum likelihood results when fitting the bivariate model for the independence case

Model	Log-likelihood (df)	Independence ($\theta = 1$)[a] Shape ($\hat{\alpha}$)		Scale ($\hat{\beta}$)	
Exponential					
Same scale	−853.23 (196)	1.00 (−)		−4.50 (.08, .09)	
Different scale[b]	−840.96 (195)	1.00 (−)		−4.95 (.14, .14)	−4.14 (.10, .11)
Weibull					
Same shape, scale	−847.96 (195)	.80 (.04, .02)		−3.74 (.16, .12)	
Different shape[b]	−837.24 (194)	.69 (.04, .02)	.88 (.04, .03)	−3.74 (.16, .12)	
Different scale[b]	−836.38 (194)	.81 (.04, .02)		−4.23 (.20, .16)	−3.44 (.17, .17)
Both different[b]	−836.33 (193)	.78 (.05, .02)	.82 (.05, .04)	−4.14 (.25, .16)	−3.49 (.22, .16)

[a] The point estimate is given with standard errors from the IWM and robust approach, respectively.
[b] When different shape or scale parameters are estimated, the table presents the estimate for the treated margin first and the untreated margins second.

Table 1b
Maximum likelihood results when fitting the bivariate model for the nonindependence case

Model	Log-likelihood (df)	Nonindependent ($\theta \neq 1$)[a] θ	Shape ($\hat{\alpha}$)		Scale ($\hat{\beta}$)	
Exponential						
Same scale	−849.65 (195)	1.51 (.23)	1.00 (−)		−4.49 (.09)	
Different scale[b]	−834.06 (194)	1.82 (.28)	1.00 (−)		−4.94 (.13)	−4.14 (.10)
Weibull						
Same shape, scale	−844.86 (194)	1.54 (.24)	.84 (.02)		−3.93 (.09)	
Different shape[b]	−830.67 (193)	1.87 (.31)	.71 (.03)	.90 (.02)	−3.78 (.09)	
Different scale[b]	−829.61 (193)	1.89 (.31)	.81 (.02)		−4.20 (.14)	−3.42 (.10)
Both different[b]	−829.58 (192)	1.89 (.31)	.79 (.03)	.82 (.02)	−4.13 (.14)	−3.48 (.10)

[a] The point estimate is given with the standard error from the Clayton–Oakes approach.
[b] When different shape or scale parameters are estimated, the table presents the estimate for the treated margin first and the untreated margins second.

the scale parameters are the same ($\beta_1 = \beta_2$). Table 2 presents the likelihood ratio test and Wald's test. There is a large treatment effect in either case but it is greater when one (correctly) takes the association into account.

The second hypothesis is whether the treated and untreated times-to-event are associated. Using the best model from Table 1, we can test the independence hypothesis by the likelihood ratio test ($\chi^2 = 13.5$), the square of Wald's test ($\chi^2 = 7.9$), and the square of the score test ($\chi^2 = 11.6$). Because the null value of θ is on the boundary of the parameter space, these test statistics are compared to an equal mixture of a chi-square distribution

Table 2
Tests and estimates of the treatment effect when $\theta = 1, \theta \neq 1$

Test	$\theta = 1.00$	$\theta = 1.89$
Likelihood ratio test[a]	23.16	30.50
[Wald's test][2 a]	22.28	28.73
Estimates	[b]	[c]
$\beta_1 = \beta_2$	−3.74 (.16, .12)	−3.93 (.09)
$\beta_1 \neq \beta_2$	−4.23 −3.44 (.20, .16) (.17, .17)	−4.20 −3.42 (.14) (.10)

[a] Compared to a χ^2 distribution with 1 degree of freedom.
[b] The point estimate is given with the standard error from the IWM and robust approach, respectively.
[c] The point estimate is given with the standard error from the Clayton–Oakes approach.

Table 3
Maximum likelihood results from the Clayton–Oakes model; Weibull model with type of diabetes as a binary covariate (0: *Juvenile*, 1: *Adult onset*)

Model	Log-likelihood (df)	$\hat{\theta}$	Shape ($\hat{\alpha}$)	Scale ($\hat{\beta}_1$)		Type of diabetes ($\hat{\beta}_2$)	
Weibull (same shape, different scale)[a]	−829.61 (193)	1.89 (.31)	.81 (.02)	−4.20 (.14)	−3.42 (.10)	—	
Weibull with type of diabetes as:							
Main effect[a]	−829.37 (192)	1.92 (.32)	.81 (.02)	−4.26 (.16)	−3.47 (.12)	.125 (.18)	
Interaction with treatment	−825.28 (191)	2.01 (.34)	.81 (.02)	−4.02 (.17)	−3.59 (.14)	−.47 (.29)	.37 (.20)

[a] When different scale or covariate parameters are estimated, the table presents the estimate for the treated margin first and the untreated margin second.

with 1 degree of freedom and the degenerate distribution at zero (Self and Liang, 1987). We conclude that there is a strong positive association between failure times for the two eyes.

Diabetes can be classified into two general groups by the age at onset: juvenile and adult diabetes. Since the two diseases have very different courses, it is desirable to examine how this may affect the time to loss of visual acuity.

Weibull probability plots support the assumption of proportional hazards for the type of diabetes within each margin. Table 3 presents maximum likelihood results modelling type of diabetes (1) as a main effect and (2) as an interaction with treatment (margins).

There is a significant decline in the log-likelihood [$\chi^2(1) = 8.66$, $P < .005$] for the interaction model. That is, within each level of the covariate, there is a significant treatment effect [for juvenile onset, the square of Wald's test statistic is 5.6 ($P = .01$), whereas for adult onset the square of the statistic is 7.9 ($P = .003$)]. Figure 2 presents the Kaplan–Meier survival curves for the treated and untreated eye, broken down by the type of diabetes. We see that the interaction arises because the treatment is much more effective for adult onset diabetics than for juvenile onset diabetics.

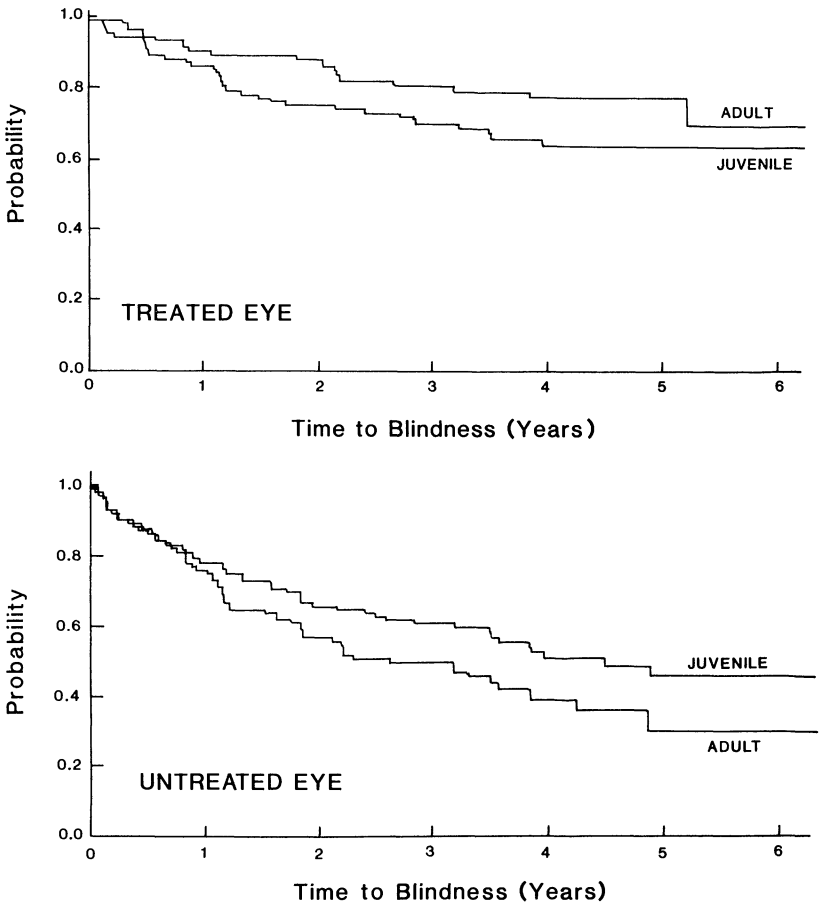

Figure 2. Kaplan–Meier curves for the time to blindness in the treated and untreated eyes of individuals with juvenile or adult onset diabetes.

6. Discussion

Many clinical trials involve the comparison of two independent groups. In clinical trials of eyes or kidneys, say, the natural pairing violates the independence assumption. We have compared two ways of analyzing survival data from such trials.

The computationally simpler IWM approach ignores the association within the pair. This can result in severe loss of information for estimation of the treatment effect, depending on the amount of association and the percentage of discordant pairs.

The Clayton–Oakes model can measure and account for the association between pair members, and thus permits efficient estimation of treatment effects. Although this model as extended here to allow censoring and covariates is quite flexible, it does have some limitations. First, the association modelled must be positive. This is unlikely to be an important restriction in most situations of interest. The model has been extended by Genest and Mackay (1986) to allow negative association. However, the support of the resulting distribution depends on θ. Second, the methods for inference proposed here are fully parametric as compared to the proportional hazards technique (Holt and Prentice, 1974). Care must be taken to assure that the appropriate form of baseline hazard is used. Finally, these methods are very computer-intensive because of the relatively complicated form of

the likelihood function and its derivatives and the fact that the maximization of the likelihood is subject to an inequality constraint on θ.

Not considered in this paper is the Holt and Prentice (1974) technique. Like the IWM, it treats the association as a nuisance. However, the Holt–Prentice technique does allow estimation of the relative risk using discordant pairs. This can lead to serious efficiency loss if there are many concordant pairs. The main difference between the Holt–Prentice model and the Clayton–Oakes model (besides its semiparametric nature) is that the proportional hazards assumption applies to the conditional (rather than the marginal) hazards.

An additional area of inquiry is optimal design of paired-data experiments. As mentioned above, efficient estimation in the Holt–Prentice model uses only discordant pairs. A similar result is obtained for the Clayton–Oakes and IWM models. This can be seen by inspecting the variance of the treatment effect for the Clayton–Oakes and IWM models as shown in Appendix 2. When there are equal numbers of treated and untreated eyes in the study, both of these variances are at a minimum when all pairs are discordant.

Acknowledgements

The authors would like to thank the two referees for their helpful advice on revision of the manuscript, and Kathleen A. Glassmeyer for typing this manuscript.

Résumé

Le but de ce papier est d'envisager l'analyse paramètrique de données de survie par paires et censurées quand on dispose d'une information additionnelle sous forme de covariables. L'étude de la rétinopathie diabétique en est un exemple sur lequel on évalue l'efficacité de la photocoagulation laser sur le retard de perte d'acuité visuelle. Dans un premier temps on étend le modèle paramètrique de Clayton (1978, *Biometrika* **65,** 141–151) pour introduire l'information apportée par les covariables. Notre seconde approche consiste à obtenir des estimations des paramètres à partir d'un modèle de travail indépendant ainsi que des estimations robustes pour la variance. On compare les approches en termes d'efficacité et sur des considérations de calcul. Un point fondamental dans le choix d'une stratégie d'analyse de données de survie par paires est de savoir si la corrélation intra-paire est un paramètre nuisible on bien un paramètre d'intérêt scientifique intrinsèque. Les approches sont illustrées avec l'étude de rétinopathie diabétique.

References

Abramowitz, M. and Stegun, J. A. (1965). *Handbook of Mathematical Functions.* National Bureau of Standards, Applied Math Series 55. New York: Dover Publications.

Aitkin, M. and Clayton, D. (1980). The fitting of exponential, Weibull, and extreme value distributions to complex censored survival data using GLIM. *Applied Statistics* **29,** 159–163.

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.

Cox, D. R. and Oakes, D. (eds). (1984). *Analysis of Survival Data.* New York: Chapman and Hall.

Cox, D. R. and Snell, E. J. (1968). A general definition of residuals. *Journal of the Royal Statistical Society, Series B* **30,** 248–275.

Crowder, M. (1985). A distributional model for repeated failure time measurements. *Journal of the Royal Statistical Society, Series B* **47,** 447–452.

Genest, C. and Mackay, R. J. (1986). The joy of copulas: Bivariate distribution with uniform marginals. *American Statistician* **40,** 280–283.

Holt, J. D. and Prentice, R. (1974). Survival analysis in twin studies and matched-pair experiments. *Biometrika* **61,** 17–30.

Hougaard, P. (1986). A class of multivariate failure time distributions. *Biometrika* **73,** 671–678.

Johnson, N. L. and Kotz, S. (1972). *Distributions in Statistics: Continuous Multivariate Distributions.* New York: Wiley.

Murphy, R. P. and Patz, A. (1978). New concepts in management of retinal vascular disorder. In *Ophthalmology Update,* International Congress Series, No. 508, K. Mizuno and Y. Mitsui (eds), 111–125. Amsterdam: Excerpta Medica.

Oakes, D. R. (1982). A model for association in bivariate survival data. *Journal of the Royal Statistical Society, Series B* **44,** 414–428.

Royall, R. M. (1986). Model robust confidence intervals using maximum likelihood estimators. *International Statistical Review* **54,** 221–226.

Self, S. G. and Liang, K. Y. (1987). Asymptotic properties of maximum likelihood estimates and likelihood ratio tests under nonstandard conditions. *Journal of the American Statistical Association* **82,** 605–610.

Received March 1987; *revised May* 1988.

Appendix 1

Proof of the Consistency of IWM *Estimators for Any Bivariate Distribution*

To show that an estimator is consistent with respect to a specified distribution, it must be shown that the expected value of the score for that estimator is 0. The expected value is taken with respect to the specified distribution.

For pair i, the score for a particular parameter of $\boldsymbol{\beta}$, β_k, can be derived from $\ln[L_{\text{IWM}}(\boldsymbol{\beta})]$ in Section 3 and is

$$U_i(\beta_k) = \sum_{j=i}^{2} \left[\delta_{ij} \frac{\partial \ln f_{ij}}{\partial \beta_k} + (1 - \delta_{ij}) \frac{\partial \ln S_{ij}}{\partial \beta_k} \right],$$

where $f_{ij} = f_{ij}(t_{ij}, \boldsymbol{\beta})$ is the marginal density for member j of pair i, $S_{ij} = S_{ij}(t_{ij}, \boldsymbol{\beta})$ is the marginal survival function for member j of pair i, and t_{ij}, δ_{ij} are the observed survival times and censoring indicators. The score for all pairs is just the sum over all pairs $i = 1, \ldots, n$.

The expected value of the score for all pairs is taken with respect to the bivariate distribution density $f(t_1, t_2) = f(t_1, t_2; \boldsymbol{\beta}, \theta)$:

$$\begin{aligned} \mathrm{E}\left[\sum_{i=1}^{n} U_i(\beta_k); f(t_1, t_2) \right] &= \int_{t_2} \int_{t_1} \sum_{i=1}^{n} U_i(\beta_k) f(t_1, t_2)\, dt_1\, dt_2 \\ &= \sum_{i=1}^{n} \int_{t_2} \int_{t_1} \sum_{j=1}^{2} \left[\delta_{ij} \frac{\partial \ln f_{ij}}{\partial \beta_k} + (1 - \delta_{ij}) \frac{\partial \ln S_{ij}}{\partial \beta_k} \right] f(t_1, t_2)\, dt_1\, dt_2. \end{aligned} \tag{1}$$

Now, if the marginals of the IWM have been correctly specified, that is, if the marginals of the IWM are the same as the marginals of the bivariate distribution, then

$$\int_{t_1} f(t_1, t_2)\, dt_1 = f(t_2, \boldsymbol{\beta}) \quad \text{and} \quad \int_{t_2} f(t_1, t_2)\, dt = f(t_1, \boldsymbol{\beta}). \tag{2}$$

Also, since the IWM estimators are consistent for the parameters of the marginal distributions under the assumption of independence (T_1 is independent of T_2), the expected value of the score for each member of a pair with respect to the marginal is 0. That is, for $j = 1$ and 2,

$$\int_{t_j} \left[\delta_{ij} \frac{\partial \ln f_{ij}}{\partial \beta_k} + (1 - \delta_{ij}) \frac{\partial \ln S_{ij}}{\partial \beta_k} \right] f_{ij}\, dt_j = 0. \tag{3}$$

It can be seen that equation (1) is 0 by applying equations (2) and (3) to equation (1).

Appendix 2

Derivation of ARE *Comparing the Clayton–Oakes Model Estimator of* β *to the* IWM *Model Estimator When There Are Exponential Margins, a Binary Covariate, and No Censoring*

If we assume exponential margins with a binary covariate, X, then the marginal density, f_j, and survivor function, S_j, for each member j of a pair are

$$f_j = f(t_j; \alpha, \beta) = \alpha e^{\beta X_j} \exp(-\alpha e^{\beta X_j} t_j)$$

and

$$S_j = S(t_j;\ \alpha,\ \beta) = \exp(-t_j^{\alpha} e^{\beta X_j}),$$

where α is the scale parameter and β is the parameter of the covariate.

For the Clayton–Oakes model, the bivariate exponential density is

$$f(t_1,\ t_2) = f(t_1,\ t_2;\ \theta,\ \alpha,\ \beta) = \frac{\theta f_1 f_2}{(S_1 S_2)^{\theta} D^{2+[1/(\theta-1)]}}$$

where $D = [(1/S_1)^{\theta-1} + (1/S_2)^{\theta-1} - 1]$ and θ is the parameter that measures association between survival times.

Since there is no censoring, the likelihood is just the product of the joint densities of all pairs:

$$L_{\text{co}}(\theta,\ \alpha,\ \beta) = \prod_{i=1}^{n} f(t_{i1},\ t_{i2}).$$

The asymptotic variance of β is the diagonal element of the inverse of the information matrix that corresponds to β. If we assume $\sum_i \sum_j X_{ij} = 0$ for $i = 1, \ldots, n$ pairs and $j = 1, 2$ members of a pair, then it can be shown that $\hat{\theta}$, $\hat{\alpha}$, and $\hat{\beta}$ are uncorrelated. Therefore, it suffices to compute expression (1):

$$\mathrm{E}\left[\frac{-\partial^2 \ln L_{\text{co}}(\theta,\ \alpha,\ \beta)}{\partial \beta^2}\right]. \tag{1}$$

Expression (1) becomes expression (2) after differentiation, taking expectations, and application of results from Oakes (1982):

$$\left(\sum_i \sum_j X_{ij}^2\right)\left[1 + \frac{2\theta(\theta-1)^2}{3\theta-2}\right] - 2(\theta-1)(2\theta-1)\rho(\theta)\left(\sum_i X_{i1}X_{i2}\right), \tag{2}$$

where

$$\rho(\theta) = \frac{1}{(3\theta-2)(2\theta-1)} + \frac{\theta}{2(3\theta-2)(2\theta-1)(\theta-1)}\left\{\Psi'\left[\frac{1}{2(\theta-1)}\right] - \Psi'\left[\frac{\theta}{2(\theta-1)}\right]\right\}$$
$$+ \frac{1}{2(3\theta-2)(2\theta-1)(\theta-1)}\left\{\Psi'\left[\frac{\theta}{2(\theta-1)}\right] - \Psi'\left[\frac{2\theta-1}{2(\theta-1)}\right]\right\},$$

where $\Psi'(\cdot)$ is the trigamma function (Abramowitz and Stegun, 1965). The variance of β for the Clayton–Oakes model is just the inverse of (2).

For the IWM, the likelihood is the product over all pairs of each of the marginal densities:

$$L_{\text{IWM}}(\alpha,\ \beta) = \prod_{i=1}^{n} f(t_{i1};\ \alpha,\ \beta) f(t_{i2};\ \alpha,\ \beta).$$

The asymptotic variance of β is the diagonal element of $\mathbf{\Lambda}(\alpha, \beta)$ (shown in Section 3) that corresponds to β. Under the assumption that $\sum_i \sum_j X_{ij} = 0$, $\hat{\alpha}$ and $\hat{\beta}$ are uncorrelated and the asymptotic variance of β is the ratio of expression (3) to expression (4):

$$\mathrm{E}\left[\left(\frac{\partial \ln L_{\text{IWM}}(\alpha,\ \beta)}{\partial \beta}\right)^2\right] = \mathrm{E}\left[\left(\sum_i \sum_j X_{ij} - \sum_i \sum_j \alpha X_{ij} e^{\beta X_{ij}} t_{ij}\right)^2\right] = \sum_i \{X_{i1}^2 + X_{i2}^2 + [h(\theta) - 2]X_{i1}X_{i2}\} \tag{3}$$

where $h(\theta) = \Psi'\{1/[2(\theta-1)]\} - \Psi'\{\theta/[2(\theta-1)]\}$, and

$$\mathrm{E}^2\left[-\frac{\partial^2 \ln L_{\text{IWM}}(\alpha,\ \beta)}{\partial \beta^2}\right] = \mathrm{E}^2\left(\sum_i \sum_j \alpha e^{\beta X_{ij}} X_{ij}^2 t_{ij}\right) = \left(\sum_i \sum_j X_{ij}^2\right)^2. \tag{4}$$

The asymptotic variance of β for the IWM is

$$\frac{\sum_i \{X_{i1}^2 + X_{i2}^2 + [h(\theta) - 2]X_{i1}X_{i2}\}}{(\sum \sum X_{ij}^2)^2}. \tag{5}$$

The ARE of the IWM estimator to the Clayton–Oakes estimator is the product of (2) and (5):

$$\left\{\left[1 + \frac{2\theta(\theta - 1)^2}{3\theta - 2}\right] - 2(\theta - 1)(2\theta - 1)\rho(\theta)\left(\frac{1}{2} - \frac{d}{N}\right)\right\}\left\{1 + [h(\theta) - 2]\left(\frac{1}{2} - \frac{d}{N}\right)\right\} \Big/ \sum \sum X_{ij}^2,$$

where

$$\frac{\sum X_{i1}X_{i2}}{\sum \sum X_{ij}^2} = \frac{1}{2} - \frac{d}{N},$$

and d is the number of discordant pairs among the N pairs. Figure 1 graphically displays the ARE as a function of θ and of the percentage of discordant pairs ($100d/N$).