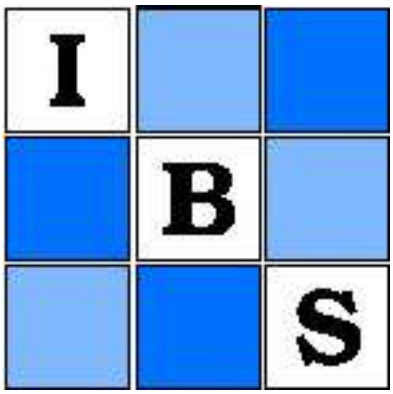
Title
Semi-Parametric Estimation in Failure Time Mixture Models

Author
Jeremy M. G. Taylor

Publication Date
2011-10-24



BIOMETRICS 51, 899–907
September 1995

Semi-parametric Estimation in Failure Time Mixture Models

Jeremy M. G. Taylor

Department of Biostatistics, University of California,
Los Angeles, California 90024-1772, U.S.A.

SUMMARY

A mixture model is an attractive approach for analyzing failure time data in which there are thought to be two groups of subjects, those who could eventually develop the endpoint and those who could not develop the endpoint. The proposed model is a semi-parametric generalization of the mixture model of Farewell (1982). A logistic regression model is proposed for the incidence part of the model, and a Kaplan–Meier type approach is used to estimate the latency part of the model. The estimator arises naturally out of the EM algorithm approach for fitting failure time mixture models as described by Larson and Dinse (1985). The procedure is applied to some experimental data from radiation biology and is evaluated in a Monte Carlo simulation study. The simulation study suggests the semi-parametric procedure is almost as efficient as the correct fully parametric procedure for estimating the regression coefficient in the incidence, but less efficient for estimating the latency distribution.

1. Introduction

A typical assumption in survival analysis states that, if there had been no censoring, the event would eventually occur for every subject. However, it is not infrequent, when considering endpoints other than death, that the event would never occur for some fraction of the subjects. Some examples exist in the field of radiation research. Patients with tumors of the head and neck are frequently treated with radiation only. The endpoint of most interest for this treatment is local recurrence, and it is known that only between 5 and 50% of patients will experience local recurrences depending on the size of the tumor. The remaining patients will not have recurrences, because all the tumor cells have been killed by the radiation. Furthermore, it is extremely unlikely, because of the kinetics of tumor growth, that there will be any recurrences later than 5 years after treatment. Another example from radiation research is when the endpoint is spinal cord paralysis. In an animal experiment, paralysis can occur if a high dose of radiation is given, and if such paralysis does occur it will nearly always be within a well defined time window. Also it is known that if the dose is small enough paralysis will never occur. In both these situations there is a non-zero probability that the endpoint will not occur. Analysis of examples of this type have been considered previously (Farewell, 1977, 1982; Pack and Morgan, 1990; Sposto, Sather, and Baker, 1992; Larson and Dinse, 1985) using a "mixture model" or "cure model."

The objective in many analyses using this type of censored data is to investigate the effect of covariates on the outcome. A natural question is whether standard methods of survival analysis, such as the Cox proportional hazards model or accelerated failure time models, can be used for these data. In these standard methods it is assumed that all censored observations would have eventually developed the endpoint, although quite possibly at times beyond the observed range.

A simple example of a mixture model is a logistic-Weibull. Let $(d_i, t_i, \mathbf{Z}_i)$ be the observations, where $\mathbf{Z}_i$ is a vector of covariates, t_i is the observed or censored time, and $d_i = 2$ (if the event occurred), $d_i = 1$ (if it is known that the event cannot occur), and $d_i = 0$ (for a censored observation). In many applications there would be no observations with $d_i = 1$. Let D_i indicate the two groups; thus $D_i = d_i$ if $d_i \neq 0$, and D_i is unknown if $d_i = 0$. The logistic-Weibull model is

$$P(D_i = 2|\mathbf{Z}_i) = \exp(\mathbf{Z}_i\beta)/[1 + \exp(\mathbf{Z}_i\beta)],$$

Key words: Cure model; EM algorithm; Kaplan–Meier estimator; Latency; Logistic regression; Long term incidence; Radiation therapy.

and

$$P(T > t|D = 2) = \exp[-(\lambda t)^{\gamma}].$$

An attractive feature of the model is that it contains two parts which can be interpreted separately, in particular the long term incidence and the latency distribution.

There are a number of variations on this simple logistic-Weibull mixture model. Farewell (1982) uses a logistic-Weibull model and assumes that the latency distribution does depend on covariates. Larson and Dinse (1985) use a proportional hazards model for the latency with a step-function for the baseline hazard. Lo et al. (1993) uses a similar model, but with the baseline hazard determined by piecewise linear splines. Yamaguchi (1992) uses a general class of accelerated failure time models for the latency distribution. Bentzen et al. (1989) allow the covariates to influence both the long term incidence and the latency, but using the same linear combination. This is achieved by assuming λ is a function of $P(D = 2|\mathbf{Z})$. Another simple variation is to use other link functions, instead of the logistic, to model the long term incidence. Kuk and Chen (1992) develop a semi-parametric model. In their model the long term incidence depends on the covariates through a logistic link, and the latency period depends on covariates in a proportional hazards structure with unspecified baseline hazard function.

Despite its appeal there are a number of problems associated with the mixture model. One is the identifiability problem (Farewell, 1986; Laska and Meisner, 1992). This can manifest itself as a very high correlation between the intercept term of the logistic model and the shape parameter of the Weibull or as a flat likelihood surface which is far from quadratic near its maximum. There is sometimes very little information in the data about the tail of the latency distribution and yet this tail plays an important role in determining the long term incidence probability. For example, if the latency distribution has a heavy tail, with a substantial amount of probability beyond the range of the observations, then this will force the long term incidence of the event to be higher than if the latency distribution has a very light tail. This problem is less likely to arise in situations where there are a substantial number of censored observations ($d_i = 0$) at times greater than the vast majority of the observed event times ($d_i = 2$). It is more likely to arise if, for example, a Kaplan–Meier survival plot of all the data does not show a clear level plateau. However, even if a Kaplan–Meier plot does show a clear plateau, this is no guarantee of identifiability.

A second problem is that maximum likelihood estimates of β may be infinite for certain configurations of the observations. This is not unique to failure time mixture models; a similar problem occurs with standard logistic regression.

A third problem with the logistic-Weibull mixture model is that it may be too restrictively parametric in nature. The purpose of this paper is to describe a semi-parametric version of the failure time mixture model. For the case in which the latency distribution does not depend on covariates, an estimator is given in which a Kaplan–Meier type estimator replaces the Weibull distribution. This method will be referred to as the logistic/Kaplan–Meier approach. Thus our approach is similar in spirit to that of Kuk and Chen, but the estimation method is different. Kuk and Chen first estimate the regression coefficients by maximizing a stochastic approximation to the marginal rank likelihood, and then obtain an estimate of the baseline hazard function. An attractive feature of their approach is that by defining a marginal rank likelihood they eliminate the large number of parameters in the baseline latency distribution. In our approach the regression coefficients and hazard function are jointly estimated.

Section 2 contains a description of the estimator. The estimator arises naturally out of the EM algorithm approach for finding the maximum likelihood estimate (MLE) as described by Larson and Dinse (1985). Section 3 contains numerical results for a specific data set and from a simulation study. Some important issues and possible extensions are discussed in Section 4.

2. Model and Analysis

Following the notation of Larson and Dinse (1985) there are two possible categories for each individual denoted by $D = 1$ and $D = 2$. If the observation is censored before either type of event $D = 1, 2$ is observed then $d = 0$. The incidence model is given by

$$P_2(\mathbf{Z}) = P(D = 2|\mathbf{Z}) = \frac{e^{\beta \mathbf{Z}}}{(1 + e^{\beta \mathbf{Z}})},$$

where $\mathbf{Z}$ are the covariates, and $P_1(\mathbf{Z}) = 1 - P_2(\mathbf{Z})$. The survival function for the failure time in Group 2 is denoted by $S(t) = P(T > t|D = 2)$ with associated hazard $h(t)$. There is no survival function associated with $D = 1$, because it represents long term survival.

The observation for the ith individual is the triple $(d_i, t_i, \mathbf{Z}_i)$. If $d_i = 1$ the contribution to the observed data likelihood is $P_1(\mathbf{Z}_i)$, if $d_i = 2$ the contribution to the likelihood is $P_2(\mathbf{Z}_i)h(t_i)S(t_i)$, and if $d_i = 0$ the contribution to the likelihood is $P_1(\mathbf{Z}_i) + P_2(\mathbf{Z}_i)S(t_i)$. In many applications there would be no observations with $d_i = 1$.

The EM algorithm can be used to maximise the likelihood. The E-step of the algorithm consists of assigning a fraction of each censored observation to the two categories. The assigned fraction to category j is the conditional probability that the individual will eventually be in category j given that no event has occurred by time t. These weights are

$$W_2(t|\mathbf{Z}) = P(D = 2|\mathbf{Z}, T > t) = \frac{P_2(\mathbf{Z})S(t)}{P_2(\mathbf{Z})S(t) + P_1(\mathbf{Z})} \tag{1}$$

and $W_1(t|\mathbf{Z}) = 1 - W_2(t|\mathbf{Z})$. From this weighting scheme the complete (or pseudo) datum $\{(d_i, t_i, \mathbf{Z}_i, \mathbf{g}_i), i = 1, \ldots, n\}$ is constructed, where $\mathbf{g}_i$ is vector of length 2 with

$$g_{ij} = I(d_i = j) + I(d_i = 0)W_j(t_i|\mathbf{Z}_i)$$

The M-step of the algorithm consists of maximizing, regarding g_{ij} as fixed, the log-likelihood of this complete data, which can be written as $L(P) + L(S)$, where

$$L(P) = \sum_{i=1}^{n} \sum_{j=1}^{2} g_{ij} \log P_j(\mathbf{Z}_i)$$

and

$$L(S) = \sum_{i=1}^{n} \{I(d_i = 2)\log h(t_i) + g_{i2}\log S(t_i)\}$$

The attractive feature of the EM algorithm for this problem is that the two components of the complete data log-likelihood can be maximized separately.

Larson and Dinse assumed a piecewise exponential function for $h(t)$; in this paper we assume a non-parametric form for $S(t)$ in which jumps in S only occur at the times of events $D = 2$. Let $\tau_1, \tau_2, \ldots, \tau_k$ be the set of distinct observed failure times, with c_j events at τ_j and m_j censored in the interval $[\tau_j, \tau_{j+1})$. Denote this set of m_j observations by H_j.

Let $f(\tau_j) = P(T = \tau_j)$, $S(t) = 1 - \Sigma_{j:\tau_j<t} f(\tau_j)$ and $\lambda_j = P(T = \tau_j|T \geq \tau_j) = f(\tau_j)/S(\tau_j)$, then

$$f(\tau_j) = \lambda_j \prod_{\ell=1}^{j-1} (1 - \lambda_\ell)$$

$$S(t) = \prod_{j:\tau_j<t} (1 - \lambda_j),$$

$$S(\tau_j) = \prod_{\ell=1}^{j-1} (1 - \lambda_\ell), \quad \text{and}$$

$$S(\tau_j + 0) = \prod_{\ell=1}^{j} (1 - \lambda_\ell).$$

With this notation the $L(S)$ part of the complete data likelihood is

$$\prod_{j=1}^{k} \left\{ [f(\tau_j)]^{c_j} \prod_{h \in H_j} [S(\tau_j + 0)]^{w_h} \right\}$$

where w_h is the fractional weight for this censored observation determined by equation (1), i.e., for $h \varepsilon H_j$

$$w_h = \frac{P_2(\mathbf{Z}_h)S(\tau_j + 0)}{P_2(\mathbf{Z}_h)S(\tau_j + 0) + P_1(\mathbf{Z}_h)}. \tag{2}$$

Thus

$$L(S) = \sum_{j=1}^{k} \left\{ c_j \log\lambda_j + \sum_{\ell=1}^{j-1} c_j \log(1 - \lambda_\ell) + \sum_{h \in H_j} w_h \left[\sum_{\ell=1}^{j} \log(1 - \lambda_\ell) \right] \right\}.$$

It can be shown that this is maximized by

$$\lambda_j = \frac{c_j}{\sum_{r=j}^{k} (c_r + \sum_{h \in H_r} w_h)}. \tag{3}$$

Kuk and Chen (1992) obtain an analogous expression. In the denominator of (3), which represents the number at risk, only a certain fraction of each censored observation is considered at risk. If each $w_h = 1$, then (3) reduces to the Kaplan–Meier estimator. The estimators obtained from this approach using the EM algorithm will be referred to as the logistic/Kaplan–Meier estimators.

Hypothesis tests and profile likelihood confidence intervals for the parameter β can be constructed by using the observed data likelihood. In particular differences in twice the log-likelihood are compared to critical values of the chi-squared distribution. A potential problem with this standard asymptotic inference is the large number of nuisance parameters in the hazard. This will be addressed in a simulation study.

The estimator defined by λ_j in (3) does not necessarily lead to $S(t) = 0$ for t greater than the last event time. We found that $S(t) > 0$ for $t > \tau_k$ occurred more often for small sample sizes, when there was a lot of censored observations and when $d_i = 1$ was not possible. In a small simulation study we found that $S(t) > 0$ for $t > \tau_k$ for more than half the data sets in some situations. This occurrence is a manifestation of the identifiability problem in which the tail of the latency distribution is hard to estimate. A slight modification of the estimator defined by equation (3) in the M-step is to force $S(t)$ to be zero beyond the last event time. This forces the weight in equation (2) for these censored observations to be zero, thus essentially classifying those observations as $D = 1$. This procedure effectively eliminates the problem with lack of identifiability. Forcing $S(t)$ to be zero beyond the last event can be justified because one would only contemplate a mixture model in situations where it is clear that there are two groups and in which there is good follow up beyond the time when most of the events occur. This might occur, for example, if a Kaplan–Meier plot of the observations showed a clear plateau at long times. In this situation it might seem reasonable to assign all censored observations beyond the last event to the long term survivor category.

3. Numerical Results

3.1 *Simulation Study*

Two simulation studies were performed. The first small study was to compare the properties of the two logistic/Kaplan–Meier estimator in which $S(t)$ is and is not forced to zero beyond the last event. The second larger study is to compare the properties of the logistic/Kaplan–Meier estimator and the logistic/Weibull estimator.

For both studies data sets of size n ($n = 50$ or 100) are generated from either a true logistic Weibull model or a logistic uniform model. Specifically, $\log[P/(1 - P)] = a + b\mathbf{Z}$, where $P = P(D = 1)$, $a = -.847$, $b = 0$ or -3.0, and $\mathbf{Z}$ is uniformly distributed between $-.5$ and $.5$. The distribution of T is either Weibull $F(T|D = 2) = 1 - \exp[-(\lambda T)^\gamma]$, where $\lambda = .0864$, $\gamma = 2.5$, or T is uniform [5, 20]. The median, 5th, and 95th percentiles of the Weibull are 10, 3.5, and 18.0, respectively, and of the uniform are 12.5, 5.75, and 19.25, respectively. The observed time is $T^* = \min(T, 30.0, V)$, where $V \sim \exp(\mu)$. These configuration choices are designed to address a number of issues, in particular the importance of sample size, the importance of the amount of censoring, and the robustness of the parametric estimator to misspecification of the latency distribution. Six hundred data sets were generated for each configuration.

For some of the simulations we assumed that it is possible for observations to be known to be in the second group, i.e., $d = 1$, when $D = 1$ and $V > 30$. In other cases we assumed d was never equal to 1, so $d = 0$, if $D = 2$ and $T^* < T$ or if $D = 1$.

The observed proportion of observations in each data set with $d_i = 2$ was either approximately .44 if $\mu = 25$ or .65 if $\mu = 250$. The observed proportion with $d_i = 0$ was either all of the rest or, in the case that $d_i = 1$ is allowed, was .45 if $\mu = 25$ and .06 if $\mu = 250$.

The quantities considered in the Monte Carlo study were the median and interquartile range of a and b; the power of a likelihood ratio test of $b = 0$; the bias, variance, and mean squared error (MSE) of estimates of the conditional latency distribution $F(t)$, at $t = 3, 9, 15, 21$; and the bias, variance, and MSE of the predicted value of $P(D = 2|\mathbf{Z})$ at $\mathbf{Z} = -.5, 0$, and $.5$. In addition we considered the

bias, variance, and MSE of $P(D = 2|\mathbf{Z})(1 - S(t))$, the predicted probability of the event occurring before $t = 15$ and $t = 21$ for $\mathbf{Z} = -.5$, 0, and .5. The median and interquartile range of a and b were considered, because it is possible for the maximum likelihood estimates to be infinitely large.

In the first simulation study we found that the estimator, in which $S(t)$ was forced to zero beyond the last event, had a 10–30% smaller interquartile range for a and b and gave infinite estimates for a and b less often. When $S(t)$ was not forced to zero the estimated latency distribution was biased, particularly at $t = 21$, and had larger variance. Similarly the predicted probabilities were biased and had larger variance when $S(t)$ was not forced to zero. The size of a nominal 5% level test of $b = 0$ was in the range .04–.065 when $S(t)$ was forced to zero and was in the range .05–.09 when $S(t)$ was not forced to zero. In summary, it appeared that requiring $S(t)$ to be zero beyond the last event was a better estimator with respect to every quantity considered.

In the second simulation study each data set was analyzed using both the logistic/Weibull estimator and the logistic/Kaplan–Meier estimator with $S(t)$ set to zero beyond the last event. For the parameters a and b and the predicted probabilities $P(D = 2|\mathbf{Z})$ the performance of the logistic/Weibull and logistic/Kaplan–Meier were almost identical, with relative efficiencies between .96 and 1.04 for all configurations. Neither method showed consistently higher bias for a, b and $P(D = 2|\mathbf{Z})$. The largest absolute bias for $P(D = 2|\mathbf{Z})$ is .038 for the logistic/Kaplan–Meier and .038 for the logistic/Weibull, both these occurred for $N = 50$ with a large amount of censoring.

The power of the test $b = 0$ was always between .043 and .077 for both estimates when the true b was zero. Thus it appears that the likelihood ratio test has approximately the correct level despite the large number of nuisance parameters. When the true b was -3.0 the difference in power between the two estimators was at most .005.

When the true model was Weibull neither method showed any real bias in the estimate of $F(t)$. When the true distribution was uniform the bias in the estimate of $F(t)$ was always larger for the logistic/Weibull estimator. For the logistic/Kaplan–Meier estimator the largest absolute bias was .017, and 50% of the biases are greater than .002. For the logistic/Weibull estimator the largest absolute bias is .057, and 50% of the biases are greater than .03.

The means and standard deviations of the relative mean squared errors of logistic/Weibull to logistic/Kaplan–Meier are given in Table 1. It can be seen that the Weibull approach is appreciably more efficient than the Kaplan–Meier approach for estimating $F(t)$ when the true distribution is Weibull. The zero entries for the uniform distribution indicate that the non-parametric method always correctly estimates that no events are possible before $t = 3$ or after $t = 21$.

Table 2 shows the relative MSEs for the predicted probability of the event by $t = 15$ and $t = 21$. For $t = 21$ the relative MSEs are very close to 1, because the estimator is almost identical to the long term predicted probability for which the relative efficiency appears to be 1. For $t = 15$, the properties

Table 1
Relative efficiencies of logistic/Weibull to logistic/Kaplan–Meier for estimating $F(t)$ at $t = 3, 9, 15, 21$

True distribution	t			
	3	9	15	21
(a) Weibull Relative MSE Mean (SD)	2.78 (.36)	1.43 (.05)	1.51 (.06)	2.32 (.36)
(b) uniform Relative MSE Mean (SD)	0 (0)	.92 (.25)	1.03 (.12)	0 (0)

of the estimator of the latency distribution begin to play a role. This makes parametric method slightly superior when the true model is Weibull and slightly inferior when the true model is uniform. Other factors in the design of the simulation such as the true value of b, the value of z at which the prediction is made, the amount of censoring, and the sample size had only a minimal effect of the relative MSE.

3.2 *Guinea Pig Data*

Spinal cord paralysis is a potential side effect in radiation therapy if too high a dose is administered to that region. To investigate the relationship between the absorbed dose and paralysis an experiment was performed in which Guinea pigs were irradiated with a range of doses. The total dose is administered in N separate 1.5-Gy doses followed by ten 4.5-Gy doses. There were nine different

Table 2
Relative efficiency of logistic/Weibull to logistic/Kaplan–Meier (ratio of MSEs) for estimating predicted probabilities of the event occurring by fixed times (t)

True distribution	$t = 15$	$t = 21$
Weibull	1.10 (.08)	1.02 (.02)
Uniform	.94 (.08)	1.00 (.07)

values for N ranging from 18 to 44. The doses are given daily except for twice daily on Saturdays and Sundays. From previous experience it is known that if paralysis is to occur it will be between 10 and 18 weeks after the final dose of radiation.

There were 115 animals in the experiment, 58 developed paralysis between weeks 10 and 18 ($d = 2$), 42 did not develop paralysis at the termination of the experiment at 30 weeks ($d = 1$), and 15 were censored at various times between 7 and 17 weeks ($d = 0$). The data were fit by the logistic/Kaplan–Meier method. The estimates for a and b are 7.739 and $-.0862$, respectively, and a 95% profile likelihood confidence interval for b is $(-.138, -.041)$ indicating a significant dose response relationship. Figure 1 shows the estimated conditional latency distribution, and also the estimate of $F(t)$ from the logistic/Weibull fit for comparison. It appears for this data set that the assumption of a Weibull for the latency distribution is appropriate. For comparison the Cox model estimated conditional distribution given paralysis before 20 weeks at a total dose of 93 Gy is shown.

Figure 2 shows the predicted proportion free of paralysis at 20 weeks as a function of dose for the logistic/Kaplan–Meier method. The results from the logistic/Weibull method are almost identical,

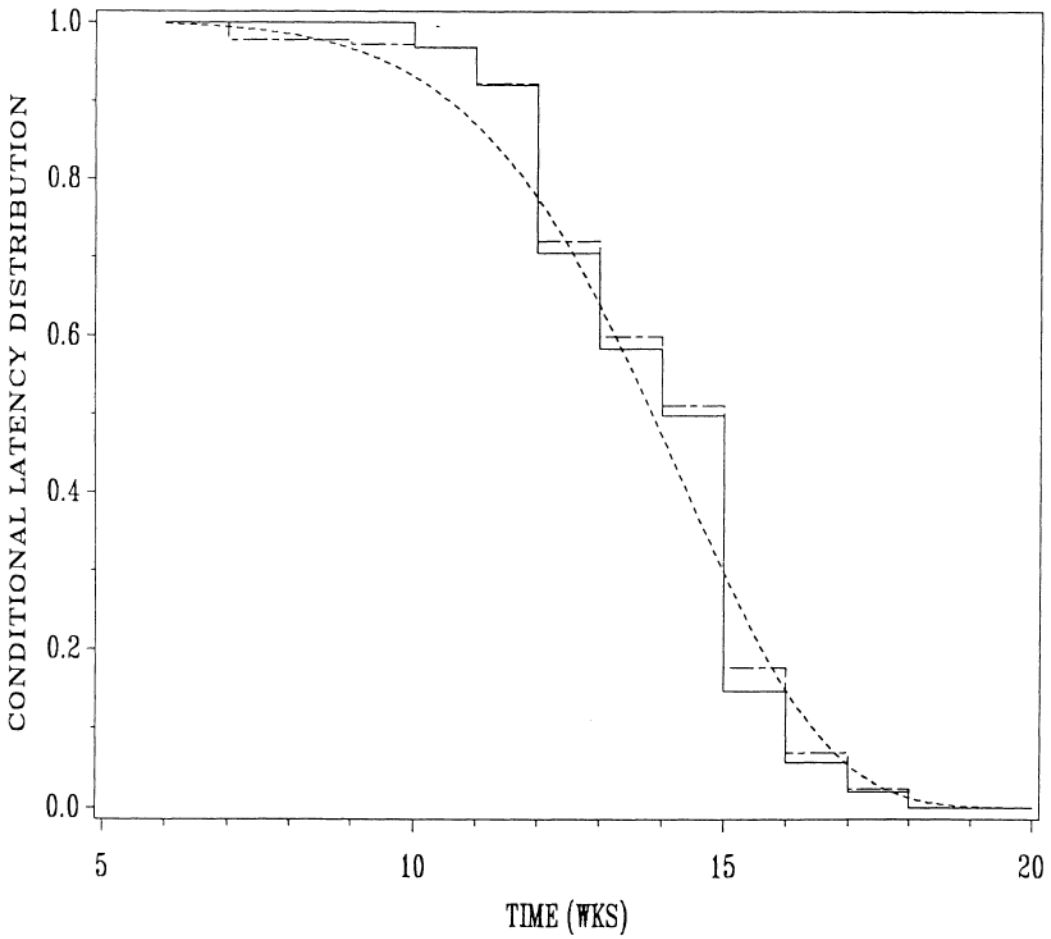


Figure 1. Estimated conditional latency distribution for Guinea pig paralysis experiment. Logistic/Kaplan–Meier estimator (solid step function), Logistic/Weibull estimator (dashed smooth curve), standard Cox model (dashed step function).

and are not shown. The data points and 95% confidence intervals are shown, calculated from separate Kaplan–Meier plots for each dose group. For comparison the results of fitting a standard Cox proportional hazards model are shown. The predictions are similar, although the Cox model does suggest a less steep dose-response relationship, as has been previously observed (Taylor and Kim, 1993).

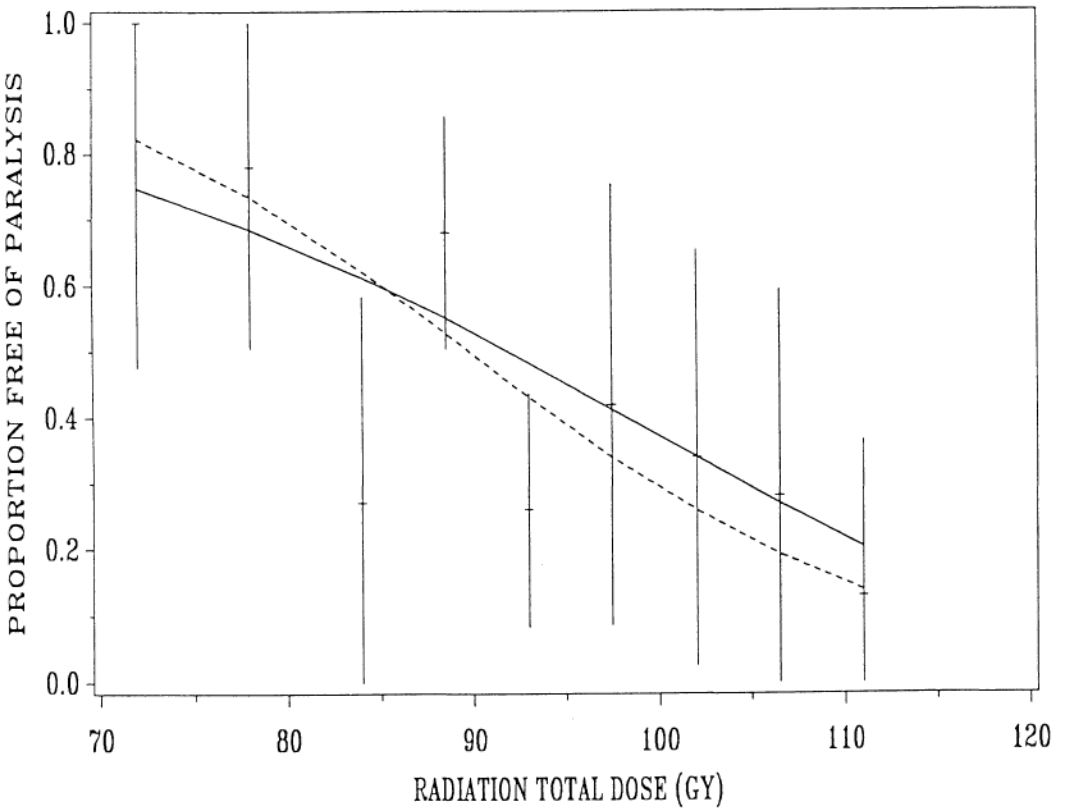


Figure 2. Estimated proportion of Guinea pigs free of paralysis at 20 weeks following radiation. Data points and approximate 95% confidence intervals are separate Kaplan–Meier estimates (±2 SE) from each dose group. Logistic/Kaplan–Meier mixture model estimate (- - - - - - - -). Cox proportional hazards model (————).

4. Discussion

As has been noted by others (Farewell, 1986) there are potential problems in using the mixture model in cases where it may not be adequately justified. The estimator recommended in this paper, because of its good properties, forces $S(t)$ to zero beyond the last event. From a practical point of view this apparently strong assumption may not be unreasonable if there is adequate justification for a mixture model, and if it is not reasonable the suitability of the model might be questioned.

Our work is related to a recent article by Kuk and Chen (1992). They consider a more general semi-parametric-mixture model in which the covariates can influence both the long term incidence and the conditional latency, whereas we only allow the covariates to affect the incidence. The Kaplan–Meier type estimator we use in the M-step is also derived by Kuk and Chen. The difference between the two approaches is that we continually iterate between the incidence regression coefficients and the Kaplan–Meier type estimator in an EM algorithm. In contrast Kuk and Chen first obtain the regression coefficients by maximizing an approximation resulting from a simulation to the marginal rank likelihood and then obtain the Kaplan–Meier type estimator without any iterations. An additional difference is that we allow the possibility that for some observations it is known that the event cannot occur. We are currently working on extending our EM approach to the more general model assumed by Kuk and Chen. A difficulty here is that the latency regression coefficients cannot be estimated separately from the baseline hazard.

In standard logistic regression there are situations where the regression coefficients which maximize the likelihood are infinite (Santner and Duffy, 1989). The logistic/Kaplan–Meier model can

also give infinite regression coefficients, especially for small sample sizes. Specifically if there is a possible fractional allocation of the censored observations (equation 2), such that the design space can be separated by a hyperplane into two groups, then the regression coefficients might be infinite. We observed that including the restriction $S(t) = 0$ beyond the last event had the beneficial effect of reducing, but not eliminating, the occurrence of infinite estimates.

For the situation where $d = 1$ is not possible, if the largest value t occurs for an uncensored observation ($d = 2$), then we observed that the estimate of b is infinite because the allocation of all the censored observations to the group $D = 2$ is possible and maximizes the likelihood. This would appear to be a limitation of the logistic/Kaplan–Meier approach, although in this situation one might question whether it is really valid to think that a non-zero proportion of the subjects will never experience the event and thus whether a mixture model is appropriate.

Because of the potential problems caused by the lack of identifiability when using the mixture model, it is interesting to know whether a standard analysis, such as the Cox proportional hazards model, is satisfactory for this type of data. Certain aspects of this question were investigated in a simulation study and by applying both techniques to two real examples (Taylor and Kim, 1993). We found that generally the methods gave similar results, and that when the true model was a mixture model the logistic/Weibull approach gave smaller MSE than the Cox model for predicting the probability of an event by a specific finite follow-up time for a given covariate vector. In contrast, the Cox model was more efficient at estimating the ratio of two regression parameters. The ratio of regression parameters has an interpretation as the substitutability (Taylor, 1989; Li and Duan, 1989) of one covariate for another and is frequently considered in radiation research.

Acknowledgements

This work was partially supported by National Institute of Health grant CA-45216.

Résumé

Les modèles de mélange constituent une approche intéressante pour l'analyse des données de survie dans lesquelles deux groupes de sujets peuvent être envisagés, ceux qui pourraient éventuellement expérimenter l'événement d'intérêt et ceux qui ne le pourraient pas. Le modèle proposé est une généralisation semi-paramétrique du modèle de mélange de Farewell (1982). Un modèle de régression logistique est proposé pour la partie du modèle concernant l'incidence et une approche de type Kaplan–Meïer est utilisée pour estimer la partie du modèle concernant le temps de latence. L'estimateur se déduit naturellement de l'approche utilisant l'algorithme EM pour ajuster des modèles de mélange en survie comme décrit par Larson et Dinse (1985). La procédure est appliquée à des données expérimentales d'irradiation en biologie et est évaluée par une étude de simulation de Monte Carlo. L'étude de simulation suggère que la procédure semi-paramétrique est presque aussi efficace que la procédure paramétrique correcte pour estimer le coefficient de régression dans l'incidence, mais moins efficace pour estimer la distribution de latence.

References

Bentzen, S. M., Thames, H. D., Travis, E. L., Ang, K. K., van der Schueren, E., Dewit, L., and Dixon, D. O. (1989). Direct estimation of latent time for radiation injury in late-responding normal tissues: Gut, lung, and spinal cord. *International Journal of Radiation Biology* **55,** 27–43.

Farewell, V. T. (1977). A model for a binary variable with time-censored observations. *Biometrika* **64,** 43–46.

Farewell, V. T. (1982). The use of mixture models for the analysis of survival data with long-term survivors. *Biometrics* **38,** 1041–1046.

Farewell, V. T. (1986). Mixture models in survival analysis: are they worth the risk? *Canadian Journal of Statistics* **14,** 256–262.

Kuk, A. Y. C. and Chen, C.-H. (1992). A mixture model combining logistic regression with proportional hazards regression. *Biometrika* **79,** 531–541.

Larson, M. G. and Dinse, G. E. (1985). A mixture model for the regression analysis of competing risk data. *Applied Statistics* **34,** 201–211.

Laska, E. M. and Meisner, J. J. (1992). Nonparametric estimation and testing in a cure model. *Biometrics* **48,** 1223–1234.

Li, K. C. and Duan, N. (1989). Regression analysis under link violation. *Annals of Statistics* **17,** 1009–1052.

Lo, Y. C., Taylor, J. M. G., McBride, W. H., and Withers, H. R. (1993). The effect of fractionated doses of radiation on mouse spinal cord. *International Journal of Radiation Oncology Biology Physics* **27,** 309–317.

Pack, S. F. and Morgan, B. J. T. (1990). A mixture model for interval-censored time-to-response quantal assay data. *Biometrics* **46,** 749–757.

Santner, T. J. and Duffy, D. E. (1989). *The Statistical Analysis of Discrete Data*. New York: Springer-Verlag.

Sposto, R., Sather, H. N., and Baker, S. A. (1992). A comparison of tests of the difference in the proportion of patients who are cured. *Biometrics* **48,** 87–99.

Taylor, J. M. G. (1989). A note on the cost of estimating the ratio of regression parameters after fitting a power transformation. *Journal of Statistical Planning and Inference* **21,** 223–230.

Taylor, J. M. G. and Kim, D. K. (1993). Statistical models for analysing time-to-occurrence data in radiobiology and radiation oncology. *International Journal of Radiation Biology* **64,** 627–640.

Yamaguchi, K. (1992). Accelerated failure-time regression models with a regression model of surviving fraction: An application to the analysis of "permanent employment" in Japan. *Journal of the American Statistical Association* **87,** 284–292.

Received July 1993; *revised January and July* 1994; *accepted August* 1994.