

Nonparametric methods for employment termination times with competing causes

MARA TABLEMAN* AND WERNER A. STAHEL

A dataset of termination times of a company is analyzed by fitting a competing risk model, taking left truncation and right censoring into account. The analysis intends to provide evidence for a claim of age discrimination. The model imposes no functional form on the parameters of interest, namely the survivor, cause-specific hazard, and cumulative incidence functions. The empirical estimators are modified to account for delayed entry. Inference is primarily based on smoothed empirical hazard functions. The approach provides a frequentist alternative to the Bayesian analysis advocated by Kadane and Woodworth (2004) and is easy to implement using existing software.

AMS 2000 SUBJECT CLASSIFICATIONS: Primary 62N02; secondary 62G05.

KEYWORDS AND PHRASES: Age discrimination, competing risks model, conditional probability function, flow data, kernel estimator of hazard, left truncation, nonparametric, time-varying covariate.

1. INTRODUCTION

Models for failure times describe either the survivor function or the hazard rate and their dependence on explanatory variables. In the example presented here, the question of whether older employees are fired by a certain company more often than their younger colleagues is investigated for providing evidence in a legal claim. Such cases fall under the USA's Age Discrimination in Employment Act of 1967, which federally protects individuals 40 years or older against age discrimination in employment decisions concerning hiring, firing, and promotion.

Employees may leave the company voluntarily for *other reasons*, such as to join a different company, relocation to a new geographical region for family reasons, retiring, or early death. These other reasons compete with *involuntary* termination (firing). Voluntary and involuntary terminations are thus “competing risks.”

In such legal cases flow data are collected to examine a specified observation period. The data observed, consisting of the day of termination, counted from the study begin date, are subject to *right censoring* as subjects may still be

employed with the firm at the end of the study period. Some data will also be subject to *left truncation* as some subjects are hired after the study start date. These subjects are said to have *delayed entry* into the study. Some modifications of nonparametric methods are needed for this situation.

Kadane and Woodworth (2004) advocate a Bayesian analysis of employment decisions. They use a hierarchical proportional hazards modeling approach to investigate a claim of age biased firing practices within a company. One of their examples is referred to as the Case K employment data. They conclude the data do support the claim. Here, we present a frequentist analysis of this dataset, which, in contrast to their Bayesian model, is distribution free, meaning that it imposes no functional form on the survival time distribution, and avoids the choice of a prior for the parameters of interest. The purpose of this case study is to apply nonparametric methods, which have been primarily used in medical clinical trials and epidemiology, in such a legal setting.

The dataset is presented in Section 2. In Section 3 the competing risks model with left-truncated and right-censored (LTRC) data is discussed. The population quantities of interest, namely the survivor, hazard, and cumulative incidence functions for the competing risks, are presented, and the modified empirical estimators are introduced. Kernel estimates of the empirical cause-specific hazard function for each age group form the most direct basis of inference. The log ratio of smoothed hazards and bootstrap pointwise confidence limits summarize the evidence. In Section 4 the methods are implemented and results for the Case K data are discussed. With the practitioner in mind, R code is interwoven. The estimated cumulative incidence functions are discussed in Section 4.1. Inference based on empirical estimates of conditional probability functions is conducted in Section 4.2. The conditional odds ratio of *involuntary* termination is defined. Bootstrap confidence intervals are used to draw inference. The evidence provided by comparing smoothed empirical hazard functions is presented in Section 4.3. In Section 4.4, the semiparametric proportional hazards model is used, and we show that standard diagnostics for this model would also lead to an adequate analysis. Section 5 is a brief discussion.

2. THE DATASET

The “case K” dataset `CaseK` is available from the archive `statLib` located at lib.stat.cmu.edu/datasets under

*Corresponding author.

Table 1. Cross tabulation of termination type by age group

age	Status			
	<i>involuntary</i>	<i>other</i>	censored	
≥ 40	79	47	88	214
< 40	17	53	113	183
	96	100	201	397

employment. For 416 subjects employed by the company involved, birth, hire, and end of employment dates, as well as cause of termination were recorded. Through personal communication, we learned the study period began on June 1, 1990 and ended on Sep 2, 1994. The spreadsheet in Table 2 in the Appendix indicates the nature of the original data.

Upon checking the individual cases, we found four cases had missing birth dates. Fourteen had hire dates after the study end date and one had an exit date before the study start date which excludes them from the analysis as they are outside the study period. Hence, our reference dataset contains 397 employees.

A cross tabulation of age group by failure type is given in Table 1. Of those *involuntarily* terminated, 82.3% are age 40 or older, which clearly exceeds the percentage, 54%, of this age group among the employees. Expressed differently, 37% of the protected age group were fired, compared to 9% of those under 40. This simple analysis suggests clear evidence for age discrimination. Note, however, that young employees leave the company more frequently than their older colleagues for other reasons, and therefore, a more detailed analysis is needed.

The failure time T is day in study period when termination occurred. Thus, T^* equals the exit date (or end date of study period if still employed then), expressed in days after the study start. The delayed entry time X is days until entry into the study period, which is 0 for those who were working in the company at the study start. Censored individuals include all cases which were still employed at the end of the study period. There are two failure types: *involuntary* termination, coded as 1, and exit for *other* reasons (quit, retired, died, etc.), as 2. For example, in Table 2, observation 2 is censored with $X = 0$. The 3rd observation has failure type 2 with $X = 248$. Observation 15 (the 14th after deletion of the cases mentioned above) corresponds to an employee with $X = 210$ and who was *involuntarily* terminated at 24-JAN-1992, which corresponds to $T = 603$.

A complication arises with employees that turn 40 within the study period. It is resolved by splitting such cases into two, the first one generating an observation censored the day before the 40th birthday, belonging to the young age group, and the second, with delayed entry time equal to the day before the 40th birthday, attributed to the protected age group. The new dataset `d.casekext` then counts 430 records. In the Appendix we explain the R function `julian` used to convert dates into days in the study period. The generated variables are displayed in Table 3.

3. COMPETING RISKS MODEL WITH LTRC DATA

A nonparametric estimator of the survivor function under LTRC. Let us first consider overall survival, overlooking type of termination. The termination time T is assumed to be a continuous random variable with distribution function F . The survivor function $S(t) = 1 - F(t)$ is the probability of termination after day t . Let C denote the right censoring time and X , the delayed entry time. The random times X , T , and C are assumed to be independent. In the LTRC model we observe the triple (T^*, X, δ) , where

$$(1) \quad T^* = \min(T, C) \text{ and } \delta = \begin{cases} 1 & \text{if } T \leq C, \\ 0 & \text{if } C < T, \end{cases}$$

conditional on $X < T^*$. The data we observe are n independent and identical random triples $\{T_i^*, X_i, \delta_i\}$ for which $X_i < T_i^*$. For an introduction to left truncation, see, for example, Tableman and Kim (2004), Chapter 7.3.

An analogue to the Kaplan-Meier (1958) estimator (K-M), obtained by adjusting the risk set at each failure time to account for delayed entries, provides a nonparametric consistent estimator for the survivor function S . It is defined as follows: The adjusted risk set at time t consists of all employees with delayed entry time x_i before t that are still employed just before time t ,

$$(2) \quad \mathcal{R}(t) = \{i \mid x_i < t \leq t_i^*\}.$$

Let $t'_1 < \dots < t'_{n'}$, $n' \leq n$, denote the set of distinct ordered uncensored observed event times (T_i^* with $\delta_i = 1$). Let n_k be the size of the risk set at time $t = t'_k$, and d_k be the number of uncensored observations with $T_i^* = t'_k$; that is, $n_k = \#\{i \mid x_i < t'_k \leq t_i^*\}$ and $d_k = \#\{i \mid t'_k = t_i^*, \delta_i = 1\}$. Then, the modified K-M estimator is

$$(3) \quad \widehat{S}^{adj}(t) = \prod_{t'_k \leq t} \left(\frac{n_k - d_k}{n_k} \right).$$

As $\widehat{S}^{adj}(t)$ estimates the probability of termination after day t , the left-hand limit $\widehat{S}^{adj}(t-)$ estimates the probability of being at risk at day t . This estimator, referred to as the product-limit estimator under LTRC, was proposed by Tsai, Jewell, and Wang (1987). They show it is the nonparametric maximum likelihood estimator of S . Further, they establish asymptotic normality and provide a consistent estimator for the asymptotic variance. When all $x_i = 0$, $\widehat{S}^{adj}(t)$ reduces to the K-M estimator for right-censored data.

Competing risks setting. In the competing risk setting, we study jointly the random variable $T =$ failure time and $j = 1, 2, \dots, J$ distinct, exclusive, and exhaustive failure types (or causes). The J causes are called competing risks as the occurrence of any one precludes the occurrence of the others. Let ε denote the failure type. The observations now consist of random triplets $\{T_i^*, X_i, \varepsilon_i \delta_i\}$.

Population quantities of interest, which provide an overall description of company and employee termination activity, are the cumulative incidence functions (CIF) for type j failure, which are the subdistribution functions F_j defined as

$$(4) \quad F_j(t) = \Pr(T \leq t, \varepsilon = j).$$

It is called a subdistribution function as the total probability is less than one. The $F_j(t)$ are assumed to be continuous with subdensities $f_j(t)$. The term CIF was coined by Kalbfleisch and Prentice (1980, pages 168–169). It is also known as the marginal probability, the crude incidence, and the absolute cause-specific risk. The survivor function for T is now given by

$$(5) \quad \begin{aligned} S(t) &= \Pr(T > t) \\ &= 1 - \sum_{j=1}^J \Pr(T \leq t, \varepsilon = j) = 1 - \sum_{j=1}^J F_j(t). \end{aligned}$$

The population quantity most relevant to the plaintiff's claim of age discrimination is the cause-specific hazard function $h_1(t)$, where $h_j(t)$ defined to be

$$(6) \quad \begin{aligned} h_j(t) &= \lim_{\Delta t \rightarrow 0} \frac{\Pr(t < T \leq t + \Delta t, \varepsilon = j \mid T > t)}{\Delta t} \\ &= \lim_{\Delta t \rightarrow 0} \frac{\Pr(t < T \leq t + \Delta t, \varepsilon = j)}{\Delta t \Pr(T > t)} \\ &= \frac{f_j(t)}{S(t)}. \end{aligned}$$

This quantity measures the rate – the intensity – of type 1 failure at time t amongst all employed at time t . The main question we seek to answer is: Was the hazard of *involuntary* termination on a particular day in the study period significantly greater for those employees 40 years or older compared to their younger colleagues? This question is addressed in Section 4.3.

The cause specific CIF is now reexpressed as

$$(7) \quad F_j(t) = \int_0^t f_j(u) du = \int_0^t S(u) dH_j(u)$$

where $H_j(u) = \int_0^u h_j(v) dv$ is called the cumulative hazard of type j failure.

Another population quantity of interest, which is also useful for the plaintiff's case, is the conditional probability function CP_j defined as

$$(8) \quad \begin{aligned} CP_j(t) &= \Pr(\text{type } j \text{ failure by } t \mid \\ &\quad \text{not failed from other causes by } t). \end{aligned}$$

This quantity is appealing in that it has a straightforward interpretation. It is the proportion of subjects who have experienced type j failure by time t among those surviving all other causes through time t . It is also useful when interested

in quantifying the odds of a type j failure by time t since its total probability is one, in contrast to the CIF (4). To see this, suppose there are two failure types 1 and 2. By (5), $1 - S(t) = F_1(t) + F_2(t)$. Then

$$(9) \quad CP_1(t) = \frac{F_1(t)}{1 - F_2(t)} = \frac{F_1(t)}{S(t) + F_1(t)}$$

which increases from zero to one as $0 \leq t \rightarrow \infty$, since $S(t)$ goes to zero.

Empirical estimators adjusted to account for LTRC. Define the counting processes

$$N_j(t) = \sum_{i=1}^n I(X_i < T_i^* \leq t, \varepsilon_i \delta_i = j)$$

for $j \geq 1$, and

$$Y(t) = \sum_{i=1}^n I(X_i < t \leq T_i^*).$$

The $N_j(t)$ counts the number of type j failures by t and $Y(t)$ counts the number of subjects with delayed entry before t and still at risk just prior to t . The empirical estimator of the CIF for type j failure adjusted for delayed entry is given by

$$(10) \quad \widehat{F}_j^{adj}(t) = \int_0^t \widehat{S}^{adj}(u-) \frac{dN_j(u)}{Y(u)},$$

where $\widehat{S}^{adj}(t-)$ is the left-hand limit of \widehat{S}^{adj} (3) computed over all observed failure times t_i^* irrespective of failure type. $\widehat{S}^{adj}(t-)$ is defined to be 0 when $Y(t) = 0$, and $0/0 = 0$. For the case when left truncation is not present, that is, all $x_i = 0$, the estimator $\widehat{F}_j^{adj}(t)$ reduces to that studied by Aalen (1978a). He proves strong consistency and weak convergence when the data are right-censored. For those who may not be familiar with computing the Stieltjes integral in (10), the formula yields the following sum: Let $t'_1 < \dots < t'_{n'}$ denote the set of distinct ordered observed times as before. Then $\widehat{F}_j^{adj}(t)$ can be expressed as follows:

$$(11) \quad \widehat{F}_j^{adj}(t) = \sum_{t'_k \leq t} \widehat{S}^{adj}(t'_k-) \frac{d_{jk}}{n_k},$$

where n_k = number of subjects with delayed entry before t'_k that are still at risk just prior to t'_k , and d_{jk} = number of type j failures at t'_k . (The censoring times may be included in the sum since the $d_{jk} = 0$ for these.) To understand the marginal probability estimator $\widehat{F}_j^{adj}(t)$, observe that $\widehat{S}^{adj}(t'_k-)$ estimates the probability of being at risk at t'_k and d_{jk}/n_k estimates the probability of a type j failure at t'_k conditional on being at risk at t'_k ; hence their product estimates the probability of a type j failure at t'_k . Therefore, $\widehat{F}_j^{adj}(t)$ estimates the probability of type j failure by time t .

For the Case K study there are two competing risks. Pepe and Mori (1993) discuss the empirical estimator of the CP function for a type 1 failure when data are subject to right censoring, which consists of plugging \widehat{F}_j^{adj} into equation 9,

$$(12) \quad \widehat{CP}_1^{adj}(t) = \frac{\widehat{F}_1^{adj}(t)}{1 - \widehat{F}_2^{adj}(t)}.$$

With uncensored and non-left-truncated data, $\widehat{CP}_1^{adj}(t)$ is simply the number of subjects who have been *involuntarily* terminated by t divided by the number who did not leave employment for *other* reasons by time t . In the same article, Pepe and Mori show that the large sample distribution of $\widehat{CP}_1^{adj}(t)$ is normal when left truncation is not present, and they provide a consistent variance estimator.

Lastly, to estimate the cause-specific hazard function (6) at time t , we smooth the empirical values

$$(13) \quad \hat{h}_j(t'_k) = \frac{d_{jk}}{n_k},$$

where t'_k , d_{jk} , and n_k were previously defined for expression (11). The sum of these values over all $t'_k \leq t$, $\widehat{H}^{adj}(t) = \sum_{t'_k \leq t} d_{jk}/n_k$, is a modified version of the Nelson-Aalen (1972, 1978b) estimator of cumulative hazard of type j failure. It is constructed with risk sets adjusted to account for delayed entries. This was used to estimate $H_j(u)$ in (7), as seen in (11). The estimates in (13) are the same when the risks competing with j are treated as censored.

The kernel estimator of $h_j(t)$ is given by

$$(14) \quad \hat{h}_j^{kernel}(t) = \frac{1}{b} \sum_{k=1}^{n'} \mathcal{K}\left(\frac{t - t'_k}{b}\right) \frac{d_{jk}}{n_k},$$

where the *kernel* \mathcal{K} is a bounded function which has integral 1 and vanishes outside $[-1, 1]$, and b is the *bandwidth*. The kernel estimator smoothes the occurrence/exposure rates – the increments d_{jk}/n_k in $\widehat{H}^{adj}(t)$. This estimator is a modified version of the kernel estimator of intensity proposed by Ramlau-Hansen (1983). He establishes consistency and asymptotic normality. Andersen et al. (1993, Chapter IV.2.2) generalize these results to include left truncation.

4. IMPLEMENTATION AND RESULTS FOR THE CASE K DATA

We use the statistical computing package R. It's free and downloadable from www.r-project.org. The R function `survfit` together with `Surv` yields the values essential to compute the estimates in the foregoing section for each age group $g = 0, 1$. Details are shown in the Appendix.

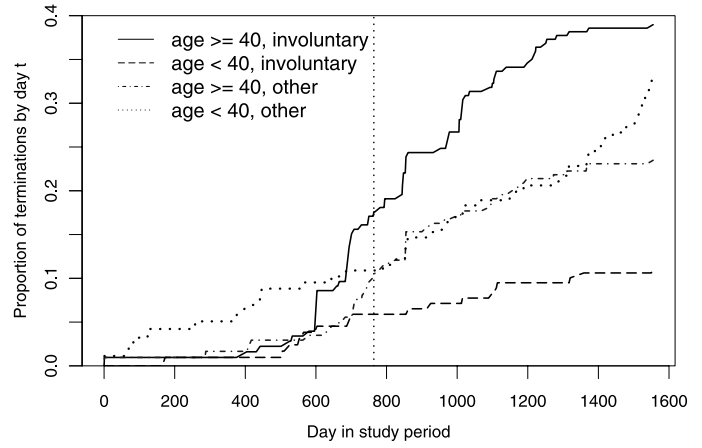


Figure 1. Estimated cumulative incidence curves for involuntary and other within age group. The vertical dotted line indicates day of plaintiff's dismissal.

4.1 Descriptive summary provided by empirical estimator of CIF

The estimated CIF's, $\widehat{F}_{gj}^{adj}(t)$ (11) for each combination of the age groups $g = 0, 1$ and termination types $j = 1, 2$ are displayed in Figure 1. At about day 600 the firm's decisions regarding *involuntary* termination become apparent. The rate of this activity in the protected age group dramatically increases whereas the rate within the younger group, remains rather low. This results in the dramatic divergence of CIF's. By the end of the study period, 39% of the protected group were affected in comparison to only 11% of the younger group.

However, the firm can also use Figure 1 to defend its actions. The CIF's for *other* reasons reveal that by day 600 about 10% of the younger group left voluntarily, whereas such activity amongst the older group was much lower (3.6%). The firm may say that they had no other choice but to assign layoffs more frequently to the older group. The Figure shows, however, that the two curves meet around day 760. Nevertheless, we seek more convincing evidence.

4.2 Inference based on an empirical estimator of CP function

The comparison of type 1 conditional probabilities does, however, make it more difficult for the firm to explain its actions. In context, $CP_{g1}(t)$ is the probability of involuntary termination by day t , given that the subject has not left for other reasons by time t . Estimates are computed for each age group using expression (12). Figure 2 confirms what was detected in Figure 1. As of day 600, $\widehat{CP}_{11}^{adj}(t)$ jumps up sharply and continues a steady increase while $\widehat{CP}_{01}^{adj}(t)$, after a gentle jump between day 500 and 600, flattens. By day 764 (the day the plaintiff was dismissed), among those

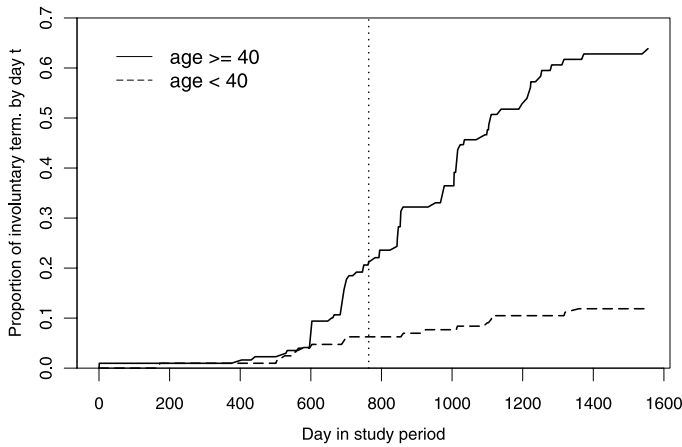


Figure 2. Estimated conditional probability curves for involuntary termination in each age group. The vertical dotted line indicates day (764) of plaintiff's dismissal.

in the protected group who did not leave for other reasons, about 21% were terminated by the firm in comparison to only 6% among the younger group. By the end of the study period, the gap widens to 64% versus 12%.

The conditional odds ratio (OR) is defined to be

$$\text{OR}(t) = \frac{\text{CP}_{11}(t)}{1 - \text{CP}_{11}(t)} \div \frac{\text{CP}_{01}(t)}{1 - \text{CP}_{01}(t)}$$

and its estimator $\widehat{\text{OR}}$ is given by plugging in the adjusted estimators $\widehat{\text{CP}}_{gj}^{adj}$. Pointwise 95% bootstrap confidence limits for the true conditional log-OR of *involuntary* termination are displayed in Figure 3.

For obtaining these limits, 1000 bootstrap samples were drawn, stratified within the groups of employees which had their 40th birthday by their day $X_i + 1$, those with 40th birthday after their exit time T^* , and those 33 who moved from the unprotected to the protected group within their observation period $(X_i + 1, T_i^*]$. For each employee in the third group, there are two records in the extended dataset, and both are selected for the bootstrap sample. This avoids problems with independence that would arise in principle (though with tiny effect on the result) if the records of the extended dataset were sampled directly. Based on the 1000 bootstrap values for the log odds ratio, pointwise percentile confidence intervals were calculated.

The $\log\widehat{\text{OR}}$ at day 764 is 1.36. The lower and upper confidence limits are 0.49 and 2.41, respectively. That is, among those who did not leave for other reasons by day 764, the odds of *involuntary* termination by day 764 are significantly greater for the protected group with estimated odds 3.9 times the odds for the younger group. If a one-sided test were asked for, its p-value would therefore be less than 2.5%. The $\widehat{\text{OR}}$ continues to increase to about 12 by day 1300 and is fairly constant thereafter.

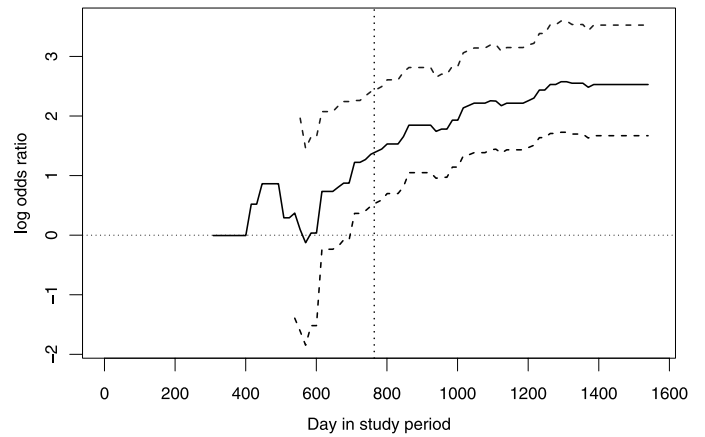


Figure 3. Conditional log odds ratios of involuntary termination, with pointwise 95% bootstrap confidence limits (---). The vertical dotted line indicates the day of plaintiff's dismissal.

A naive estimate of the odds ratio may be obtained from Table 1. The conditional probabilities of $79/(214 - 47) = 0.47$ and $17/(183 - 53) = 0.13$ lead to an odds ratio of 6.0. The discrepancy with the appropriate estimate stems from the Kaplan-Meier type adjustments for estimating the CIF and from splitting the data of those subjects which turn 40 into two records.

In view of the foregoing findings, just how would the firm now explain such grossly disproportionate adverse actions against the protected age group?

4.3 Evidence provided by comparing smoothed empirical hazard functions

Comparison of smoothed hazard estimates provides the most compelling evidence because it specifically describes the firm's actions at any given time in the study period. Because there are very few firings near the start of the observation period and heavy right censoring at the end, smoothing estimates are unreliable before day 300 and after about day 1400. The kernel estimates (14) are computed for each age group using the R function `density` and displayed in Figure 4. Clearly, between day 400 and 1300, the protected group is at a far greater risk of *involuntary* termination than their younger colleagues. The curve for the older group dramatically increases beginning about day 400 and peaks about day 900.

The log-relative risk, estimated by the log ratio of the smoothed hazards, and pointwise bootstrap 95% confidence limits are displayed in Figure 5. The log-relative risk is clearly not constant nor linear. The interval captured by the vertical gray lines, day 440 to 1476, is where the lower confidence limits lie above zero. The log-risk of being fired on day 764 is 1.79, with a confidence interval from 1.1 to 2.69. Thus, the risk is estimated to be about 6 times higher

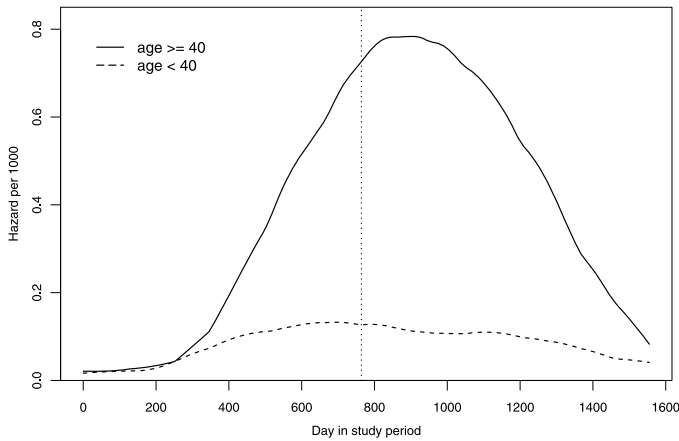


Figure 4. Smoothed estimates of cause-specific hazards for involuntary termination. The Epanechnikov kernel $\mathcal{K}(t) = 0.75(1 - t^2)$, $|t| \leq 1$ was used with bandwidth (standard deviation) of $155.5\sqrt{5} = 348$. The vertical dotted line indicates day 764.

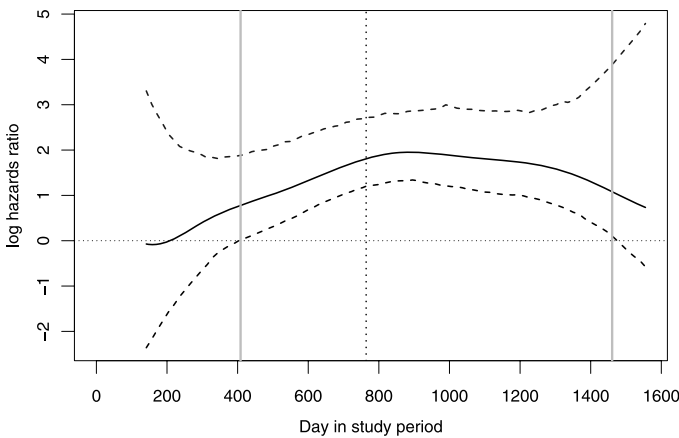


Figure 5. Log of the ratio of smoothed hazards of involuntary termination, with pointwise 95% bootstrap confidence limits (---). The vertical dotted line indicates the day of plaintiff's dismissal, and the gray lines, the days when the lower confidence limit crosses zero.

for the protected group to be involuntarily terminated on day 764.

4.4 Cox proportional hazards model with diagnostics

Let's fit the data to a Cox proportional hazards (*PH*) model, which assumes

$$h_{cox}(t) = h_0(t) \exp(\beta g),$$

where $h_0(t)$ is some unspecified baseline hazard function. The model explicitly implies that log hazards ratio β is con-

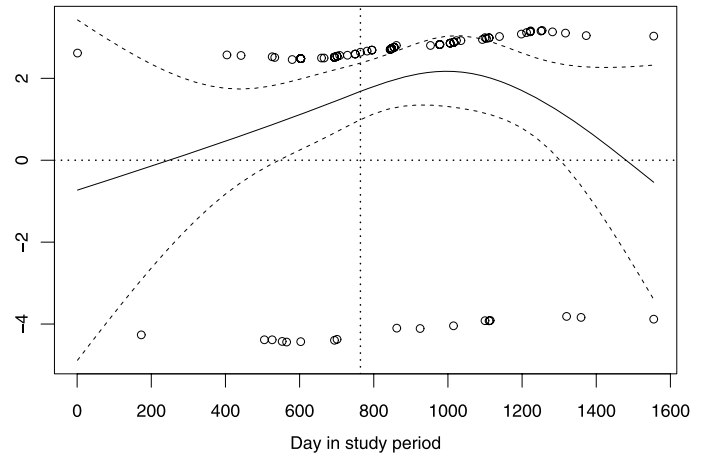


Figure 6. Scaled Schoenfeld residuals, incremented by $\hat{\beta}_{cox}$, resulting from fitting the Cox *PH* model $h(t) = h_0(t) \exp(\beta g)$ accounting for delayed entry. The lines show a spline smooth estimating $\beta(t)$ along with ± 2 standard error bands.

stant over the study period. We know this assumption is violated. Anyhow, if we had started with the Cox *PH* model – a reasonable place to start – good data diagnostics on the residuals would have led us to reject the model and try a refinement of it.

More specifically, as an alternative to the above model with a simple coefficient β , Grambsch and Therneau (1994) consider a time-varying coefficient $\beta(t) = \beta + \theta g(t)$, where $g(t)$ is a specified function of time. They show the scaled Schoenfeld residual has approximate mean $\theta g(t)$. From theory and by Monte Carlo simulation studies they show a smoothed scatter plot of the *scaled Schoenfeld residuals* plus $\hat{\beta}_{cox}$ (estimates resulting from fitting the Cox *PH* model) versus time gives a direct estimate of $\beta(t)$, and show how to test if the time-varying term $\theta g(t)$ improves the fit significantly. When the *PH* model is adequate, the spline smooth would appear horizontal, indicating $\beta(t)$, the log hazards ratio, is constant. The R function `cox.zph` produces Figure 6 and tests the default hypothesis $H_0 : \theta = 0$ for $g(t) = t$.

The p-value obtained for a linear dependence of β on time, that is, for $g(t) = t$, is 0.47. Hence, the test fails to reject $\theta = 0$, indicating no significant time dependence. But clearly, Figure 6 shows the log hazards ratio is neither constant nor linear over time and, in fact, is a smoother version of the curve in Figure 5. When we specify $g(t) = (t - \bar{t})^2$, where \bar{t} is the average of the uncensored observed times, the p-value decreases to 0.053. Our point here is to alert the reader that the results from testing the default hypothesis can be misleading. The test depends on an appropriate specification of $g(t)$ and has low power. Figure 6 is indeed welcome validation of the goodness of the kernel estimator (14) of the empirical hazard estimates (13) and the power of the bootstrap confidence limits for the log hazards ratio.

5. DISCUSSION

The data have been analyzed using nonparametric frequentist methods. The results reveal the hazards ratio – the relative risk of the protected group being involuntarily terminated on a given day in the observation period – to be neither constant nor linear, but to have a maximum near day 900. The bootstrap confidence limits show significant discrimination against employees aged 40 and above was present in involuntary terminations between days 440 and 1476, and, in particular, on the day of the plaintiff’s dismissal. The conditional log-odds ratio of involuntary termination by a certain day provides strong supporting evidence of disproportionate adverse actions against the protected age group. The results provide ample descriptions of the data which would certainly help to establish a *prima facie* case against the defendant – the company.

Kadane and Woodworth (2004) conducted a Bayesian analysis and claim that a frequentist approach would not lead to useful conclusions. They “presume that a frequentist would opt for a constant odds model” (page 188). Meanwhile, as we have demonstrated, established diagnostics for the Cox *PH* model would lead to an extension by a time-varying coefficient, and would arrive at descriptions similar to ours and to the Bayesian findings. It remains unclear why Kadane and Woodworth accuse frequentists of “unexamined subjectivity” when selecting a functional form of a model (page 192). All good statisticians share the tradition of examining model adequacy with diagnostics. The present nonparametric analysis avoids having to choose a functional form for the survival time distribution, let alone prior distributions. The question of which analysis – Bayesian versus frequentist – is easier to explain to a juror shall be left open.

APPENDIX

The original data for this study follows the pattern given in Table 2.

Table 2. Flow Data on employment termination for the Period June 1, 1990 to September 2, 1994. Dates are of the form MM DD YYYY. DOB : Date of Birth; DOH : Date of Hire; DOX : Date of Exit (termination), 99 99 1999 : still employed at study end; TI : Termination Indicator, 1 = Involuntary, 0 = Other. Observation 233 stems from the plaintiff in the legal case

obs	DOB	DOH	DOX	TI
1	11 24 1972	02 11 1991	99 99 1999	0
2	03 22 1955	03 04 1985	99 99 1999	0
3	11 13 1941	02 04 1991	10 02 1992	0
...
15	04 16 1930	12 28 1990	01 24 1992	1
...
233	04 01 1928	01 02 1990	07 03 1992	1
...

In this Appendix, we document the essential R code needed to obtain the results of this paper.

The R function `julian` converts dates into days, like

```
dox <- ifelse(t.dd$DOX=="99 99 1999", 1556,
             julian(as.Date(t.dd$DOX, format="%m %d %Y"),
                   origin=as.Date("1990-05-31")))
```

(since ‘‘99 99 9999’’ is used for most of those who were still employed at the study’s end) and similarly for DOH. The generated variables are shown in Table 3.

To compute $\hat{S}_g^{adj}(t)$ (eq. 3), the probability of overall survival beyond day t for $\text{age40} = g$, run

```
survfit(Surv(entry,time,status!=0),
       subset=(age40==g), data=d.casekext)
```

Note $\hat{S}_g^{adj}(t-)$ is the estimated probability of being at risk on day t for $\text{age40} = g$. To obtain the unique observed times t'_k , the d_{jk} = number of type j terminations at t'_k , and n_k = number of subjects with delayed entry before t'_k that are still employed just prior to t'_k (see (11)), use

```
fitgj <-
summary(survfit( Surv(entry,time,status==j),
                 subset=age40==g, data=d.casekext), censor=TRUE)
```

Then, `fitgj$time` contains the t'_k , `fitgj$n.event`, the d_{jk} , and `fitgj$n.risk`, the n_k values.

The kernel estimates (eq. 14) are computed by the R function `density` with `weights = dgj/ngj`. (Weights can be specified only in versions 2.2.1 or newer of R.)

The following R code fits the Cox model with risk sets adjusted for delayed entries, tests the default hypothesis $H_0 : \theta = 0$ for $g(t) = t$, and produces Figure 6:

```
(r.cox <-
coxph(Surv(entry,time,status==1)~age40,
      data=d.casekext))
...
      coef exp(coef) se(coef)      z      p
age40  1.52      4.57    0.268  5.66 1.5e-08
...
(r.phtest <- cox.zph(r.cox))
      rho chisq      p
age40  0.073 0.517 0.472
plot(r.phtest)
```

Table 3. The transformed variables which fit into the framework of competing risks with LTRC data

<code>time</code>	$T^* = \text{pmin}(\text{dox}, 1555)$
<code>entry</code>	$X = \text{pmax}(\text{doh}-1, 0)$
<code>status</code>	$\varepsilon = 0$ if censored, 1 if <i>involuntary</i> , 2 if <i>other</i> = $(2-\text{TI}) * (\text{dox} < 1556)$
<code>age40</code>	$g = 0$ if age less than 40, 1 if age 40 or older

The test for a quadratic function $g(t) = (t - \bar{t})^2$ is obtained from `cox.zph(r.cox, transform=function(t) (t-t.bar)^2)`.

Received 19 August 2008

REFERENCES

- [1] AALEN, O. (1978a). Nonparametric estimation of partial transition probabilities in multiple decrement models. *Ann. Statist.* **6** 534–545. [MR0478510](#)
- [2] AALEN, O. (1978b). Nonparametric inference for a family of counting processes. *Ann. Statist.* **6** 701–726. [MR0491547](#)
- [3] ANDERSEN, P. K., BORGAN, O., GILL, R. D., and KEIDING, N. (1993). *Statistical Models Based on Counting Processes*. Springer-Verlag, New York. [MR1198884](#)
- [4] GRAMBSCH, P. and THERNEAU, T. M. (1994). Proportional hazards tests and diagnostics based on weighted residuals. *Biometrika* **81** 515–526. [MR1311094](#)
- [5] KADANE, J. and WOODWORTH, G. (2004). Hierarchical models for employment decisions. *J. Business and Economic Statist.* **22** 182–193. [MR2049920](#)
- [6] KALBFLEISCH, J. D. and PRENTICE, R. L. (1980). *The Statistical Analysis of Failure Time Data*. Wiley, New York. [MR0570114](#)
- [7] KAPLAN, E. L. and MEIER, P. (1958). Nonparametric estimation from incomplete observations. *J. Amer. Statist. Assoc.* **53** 457–481. [MR0093867](#)
- [8] NELSON, W. (1972). Theory and applications of hazard plotting for censored data. *Technometrics* **14** 945–965.
- [9] PEPE, M. S. and MORI, M. (1993). Kaplan-Meier, marginal or conditional probability curves in summarizing competing risks data? *Statist. in Medicine* **12** 737–751.
- [10] RAMLAU-HANSEN, H. (1983). Smoothing counting process intensities by means of kernel functions. *Ann. Statist.* **11** 453–466. [MR0696058](#)
- [11] R Development Core Team (2005). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org>.
- [12] TABLEMAN, M. and KIM, J. S. (2004). *Survival Analysis Using S: Analysis of Time-to-Event Data*. Chapman & Hall/CRC, Boca Raton.
- [13] TSAI, W. Y., JEWELL, N. P., and WANG, M. C. (1987). A note on the product-limit estimator under right censoring and left truncation. *Biometrika* **74** 883–886.

Mara Tableman
Dept. of Mathematics & Statistics
Portland State University
Portland, OR 97207
E-mail address: tableman@pdx.edu

Werner A. Stahel
Seminar für Statistik
Swiss Federal Institute of Technology Zürich
8092 Zürich, Switzerland
E-mail address: stahel@stat.math.ethz.ch