

Estimating Episode Lengths when Some Observations Are Probably Censored

Allen C. Goodman^{1,*†}

Yingwei Peng²

Janet R. Hankin³

David E. Kalist⁴

Stephen J. Spurr¹

December 2003

Forthcoming, *Statistics in Medicine*

¹ Department of Economics, Wayne State University, 2074 FAB, Detroit, MI 48202, U.S.A.

² Department of Mathematics and Statistics, Memorial University of Newfoundland, HH-3051, Mathematics Building, St. John's, NL A1C 5S7, Canada.

³ Department of Sociology, Wayne State University, 2228 FAB, Detroit, MI 48202, U.S.A.

⁴ Department of Economics, Shippensburg University, 324 Grove Hall, Shippensburg, PA 17257, U.S.A.

KEY WORDS: Accelerated failure time model; Censored samples; Hazard functions; Monte Carlo simulation; Probit models; Simulated likelihood.

The research was partially supported by Grant DA10828 from the National Institute on Drug Abuse (NIDA), and from a grant from the Blue Cross Blue Shield Foundation of Michigan. We thank Gary Chase and Thomas McLellan for their advice and technical expertise, and two reviewers for their comments. The findings are ours and do not represent NIDA, the Blue Cross Blue Shield Foundation, or the individual universities.

***Corresponding Author:** Allen C. Goodman, Department of Economics, Wayne State University, 2145 FAB, 656 W. Kirby, Detroit, MI 48202 USA

† e-mail: allen.goodman@wayne.edu.

Estimating Episode Lengths when Some Observations Are Probably Censored

Abstract

This paper analyzes a case in censored failure time data problems where some observations are potentially censored. The traditional models for failure time data implicitly assume that the censoring status for each observation is deterministic. Therefore they cannot be applied directly to the potentially censored data. We propose an estimator that uses resampling techniques to approximate censoring probabilities for individual observations. A Monte Carlo simulation study shows that the proposed estimator properly corrects biases that would otherwise be present had it been assumed that either all potentially censored observations are censored or that no censoring has occurred. Finally we apply the estimator to a health insurance claims database.

KEY WORDS: Accelerated failure time model; Censored samples; Hazard functions; Monte Carlo simulation; Probit models; Simulated likelihood.

1. Introduction

Failure time models have become increasingly common in characterizing health services phenomena. Researchers typically consider which appropriate hazard function to use, how to handle censored samples, whether the assumption of eventual failure is merited, and which variables affect the hazard. This paper discusses a class of problems in which at least some of the sample observations are “probably censored.” Unlike failure times with observed censoring information, censoring information for some treatment episodes is not observable by the end of the episodes due to the definition of the health care episode length and the termination of studies. These problems also arise in other scientific areas. Because of unavailable or missing censoring information for some treatment episodes, traditional failure time models for data with given censoring information are not appropriate for modeling the distribution of the episode data. Failure to account for censoring will generate biased duration lengths, and for right-censored samples the estimates will be too small. However, treating *all* of the potentially censored observations as censored will overcompensate for the censoring bias.

In this article we propose a new method to analyze the probably censored data. The method properly handles the missing censoring information in an accelerated failure time model. We also apply the method to analyze mental health/substance abuse treatment episodes. Section 2 provides a detailed description of the definition of treatment episodes and reasons why probably censored data may arise. Section 3 presents an accelerated failure time model for treatment episodes and a new estimation method for the model that allows probably censored data. Section 4 reports a simulation study to investigate the performance of the proposed estimation method. In Section 5, the proposed model and

method are applied to treatment episodes constructed from a health insurance database with a detailed analysis. Section 6 provides further discussion and conclusions.

2. Treatment Episodes

Episode definition and analysis have provided an intuitive yet elusive framework for analyzing health services utilization and costs. Utilization and costs generally begin when illness or treatment episodes begin, and end when the episodes end, irrespective of the calendar. It is important to understand episode lengths and patterns to determine personnel and facility needs, and health care costs, and to assure health care quality. Screening procedures or health care treatments that can prevent or shorten episodes affect health care use and health care costs.¹

Our analysis focuses on treatment rather than illness episodes. The distinction is important because illness episodes, especially for chronic illnesses with vague symptoms and imprecise beginning dates, may be difficult to define in terms of start, duration, and conclusion. Treatment episodes, in contrast, can be defined with beginning and ending treatment events.

Episodes are based on a 30-day break point. Starting with treatment initiation, an episode includes all of the events that occur within 30 days of previous ones, so an episode may last for months or even years. We use the 30 day cut-off because several major agencies define proportion discharged from a treatment setting (particularly inpatient) and readmitted within 30 days as a key clinical “performance indicator.” Also, most of the

¹ Stern, Merwin, and Holt [1] show that information on client characteristics available from inpatient stay records is useful in predicting not only the length of inpatient stay but also the length of the subsequent community stay after hospitalization. Their findings demonstrate that client characteristics may be used to target discharge planning for those at greater risk of rapid readmission to inpatient care.

agencies wish to see patients transferred from more intensive care (inpatient, intensive outpatient) to less intensive care settings (outpatient), again within 30 days of discharge. The fact that an individual receives repeated treatments at intervals within 30 days of previous treatments indicates that the episode is not yet completed. Goodman et al. present a detailed review of the literature on episode definition [2].

By this definition, an episode ends if there has been no treatment event for 30 days. Hence a series of 6 outpatient visits, 28 days apart from each other are considered to be one episode lasting 140 days (from the day of the first treatment to the day of the last one). However, events separated by 31 days or more fall into separate episodes.

The study population was selected from a large health insurance claims database of 36 self-insured employers, for all treatment events starting January 1, 1989, and ending December 31, 1991, so episode lengths have a distribution with a minimum of 1 day and a maximum of 1,095 days. The database includes all claims of all beneficiaries less than 65 years of age (to avoid Medicare overlap) who incurred at least one drug abuse or alcoholism treatment event in the 3-year period.

Drug abuse treatment episodes were defined when the initial event had a principal International Classification of Disease (ICD-9) diagnosis of 292 (drug psychoses), 304 (drug dependence), or 305.1-305.9 (drug abuse). Alcoholism treatment was defined by an initial principal ICD-9 diagnosis of 303 (alcohol dependence), 305.0 (alcohol abuse), or 291 (alcohol psychoses). Psychiatric comorbidities (ICD-9 codes 290, 293-299, 300-302, and 306-319) were also identified. Remaining diagnoses were defined as surgery (inpatient or outpatient), or “medical” (all other groups defined through ICD-9 diagnoses). Obstetric-gynecological treatment for women was excluded to address gender comparisons.

Inpatient events consist of all services provided between and including the first and last dates of admissions involving at least an overnight stay. All other services constitute outpatient events, as defined by the employer providing the data. Thus five different diagnostic categories of treatment use are defined for both inpatient and outpatient care, for a total of 10 categories, each to be analyzed separately. This analysis examines the length of the first treatment episode within the three-year window.

Any episode with events occurring after December 1, 1991 may be right censored because an unobserved event may occur after the end of the data collection window (December 31, 1991), but within 30 days of a previous event, and thus be part of the episode. The length of any episode starting between January 1 and January 31, 1989 may also be right censored, because the 30-day treatment window may have begun before January 1, before data collection began.

Assuming that no episodes are censored leads to downward biases in estimating median lengths. However, assuming that all potentially censored episodes are censored almost certainly biases episode lengths upward. Although one solution would be to drop all potentially censored observations, we prefer an alternative approach because:

1. Omitting observations sacrifices information. In our three-year database (Table 5) between 12.4% and 28.3% of the episodes in the inpatient episode categories are potentially censored. Among the outpatient categories, 44.6% of the initial psychiatric episodes are potentially censored, and hence potentially discarded.
2. Omitting observations may lead to selection bias by differentially deleting patients with insurance coverage limitations. Moreover, specific diagnoses (often mental health or substance abuse) may be selectively deleted, again due to insurance limitations.

3. Potential censoring in this case occurs at the beginning or at the end of a calendar year. Health insurance may limit utilization or expenditure levels by calendar year, so utilization and/or costs may be concentrated at the beginning (pent up demand from the previous year) or at the end of a year (exhausting the year's benefits). Systematically omitting such episodes will likely bias estimates of episode length, utilization, and costs downward, leading to potentially serious financial consequences for providers and insurers who use these estimates.

In the following sections, we specify a regression model for the episode data and we propose a method to address the probable censoring.

3. Parametric Failure Time Models for Probably Censored Episodes

We seek to characterize the length of a specific episode defined by treatment and location, denoted by T using the following log-linear model:

$$\log T = X'\beta + Y'\gamma + F'\theta + \sigma W. \quad (1)$$

This is an accelerated failure time model. The unknown model parameters are β , γ , θ , σ , and parameters in the distribution of random error W . Variables X' refer to individual level indicators including subject's age, gender, and employment status. Variables Y' refer to coinsurance and insurance deductible levels. With perfect information and with no variation among employees in the firm, these would be the definitive measures for the individual. Vector F' characterizes the employer where the subject works and is the primary insured, or has coverage as the dependent of a family member who works there. We characterize employers with employer-specific year-specific measures such as mean age, mean employment status, or mean percentage male, for example, since employer

health care packages and policies may reflect the types of workers that are covered.

The distributions considered for T include Weibull (W follows the standard extreme value distribution), lognormal (W follows the standard normal distribution) and gamma distributions. More complicated distributions of T such as the extended generalized gamma distribution can also be considered [3, 4]. In this article, we will use the generalized gamma (EGG) distribution for length of an episode. The corresponding density function of W is

$$f(W) = \frac{|q|}{\Gamma(q^{-2})} (q^{-2} \exp(qW))^{q-2} \exp(-\exp(qW)q^{-2}) \quad (2)$$

where $\Gamma(\bullet)$ denotes the complete gamma function, and q is a free shape parameter. The Weibull, lognormal and gamma distributions for T can be obtained as special cases of the EGG distribution when $q = 1$, $q = 0$, and $q/\sigma = 1$ respectively.

Let δ_i be a censoring indicator that equals 0 for right censored observations and 1 for uncensored observations. Given values of δ_i , the likelihood function is written as:

$$L = \prod_{i=1}^n f(t_i)^{\delta_i} S(t_i)^{1-\delta_i} . \quad (3)$$

where $f(t)$ and $S(t)$ are the density and survival functions of T . However, treatment episodes exemplify a class of responses in which censoring is probabilistic. An episode may begin two weeks before the end of the data collection period. Because all treatment within 30 days of the episode initiation is considered as part of the episode, it is possible although not certain that the episode is censored. Hence the values of some of censoring indicators δ_i are unknown, or these δ_i are random variables with Bernoulli distributions.

Probable censoring can be found in other applications as well. Hougaard [5] notes that menopause is defined as the time of the last menstrual bleeding, but it is not possible to say whether a particular occurrence of bleeding is the last until a whole lifespan has passed.

Observers generally wait for one year before determining that menopause occurred at that time, but if a woman dies after a half year without bleeding “it is forever unknown whether she should count as having obtained menopause.” Alternatively the estimated age at which menopause occurs is uncertain.

Van der Laan and McKeague [6] refer to “missing censoring indicators” rather than to probable censoring. They propose an efficient estimator of the survival function of a single sample, and covariate effects are not allowed in the estimation. McKeague and Subramanian [7] obtain another estimator of the survival function of a single sample, and extend it to the Cox regression to model the effects of covariates. However, it appears difficult, if not impossible, to extend their method to model (1).

Van der Laan and Hubbard [8] and van der Laan and Robins [9] consider another example of probable censoring when an event is reported with delay. In this example, monitoring times $U_1 < U_2 < \dots < U_{k-1} < T < U_k$ and reporting times $A_1 < A_2 < \dots < A_k$ satisfy $A_j = U_j$ if $j < k$, and $A_k > U_k$. Suppose at analysis time C that death has not yet been reported. (i.e. $A_k > C$) and $U_{j-1} < C < U_j$. Analysts cannot be sure that death did not occur between U_{j-1} and C since they only know that $T > U_{j-1}$. If they set C at U_{j-1} and let T be censored at U_{j-1} , the censoring variable is now a function of T , implying that censoring is no longer independent of T , and leading to estimation biases. The authors propose Inverse Probability Censoring Weighted (IPCW) estimators to address the problem.

As the authors indicate, the censoring status of an observed time in [8] or [9] is either right censored or interval censored, while in our problem the status is either right censored or uncensored. The difference between the two problems arises from the two distinctive mechanisms that produce the probable censoring. Therefore, it is unclear

whether it is appropriate or not to apply methods for data with delayed report time to treatment episodes in model (1).

In the econometric literature, Pohlmeier and Ulrich [10] and Santos Silva and Windmeijer [11] find incomplete illness spells because the spells may have started before the observation period, and/or may end after the observation period. Santos Silva and Windmeijer observe that the observed number of spells S , and total number of visits

$V = \sum_{j=1}^S R_j$ are likely to be underestimated:

... The first obvious consequence is that S has to be interpreted as the number of illness spells the individual suffers during the observation period. Therefore S is larger than or equal to both the number of complete spells and the spells started in that period, either of which would be more interesting to model ... R_j should be viewed as the number of visits from the j^{th} spell that occurred during the observation period, which may be smaller than the total number of visits in the j^{th} illness spell (P. 77).

Like Hougaard, they offer no suggestions on how to address the problem.

Romeo [12] discusses problems when inconsistencies at the “seams” between surveys make it impossible to determine with certainty when one spell ends and the next begins. He develops a duration model that uses questionnaire answers to quantify duration data inconsistencies and he estimates the model for two different surveys. However, his method is specific to the particular databases, and is not readily applicable to the insurance claims used in our research.

In the method we propose here, let $p_i = P(\delta_i = 1)$, i.e., p_i is the probability that the individual is uncensored. Assuming that the δ_i are observed, the likelihood function of model (1) can be written as:

$$L_d(\beta | \mathbf{d}) = \prod_{i=1}^n f(t_i)^{\delta_i} S(t_i)^{1-\delta_i} p_i^{\delta_i} (1-p_i)^{1-\delta_i} . \quad (3')$$

However some values of δ_i are not available. Let D be the collection of all possible values of $\mathbf{d} = (\delta_1, \dots, \delta_n)$. If n_d is the number of probably censored patients, the size of D is 2^{n_d} .

The final likelihood function is:

$$L(\beta) = \sum_{\mathbf{d} \in D} L_d(\beta | \mathbf{d}).$$

If n_d is too large, calculation of $L(\beta)$ will be infeasible. One solution is to approximate it by

Monte Carlo methods. One approximation is $L(\beta) \approx \frac{2^{n_d}}{r} \sum_{j=1}^r L_d(\beta | \mathbf{d}_j)$ where \mathbf{d}_j is

randomly selected from D with respect to the uniform distribution that assigns equal probability to each of the 2^{n_d} members in D , and r is the number of draws.

However, the Monte Carlo approximation approach to the likelihood function has some limitations. Because we view δ_i as a random variable for the possibly censored observations, we cannot estimate (3') using existing procedures in SAS or other statistical packages. To ease the computational burden, we propose a *probit* estimator that uses comparable episode information to estimate censoring probabilities.

Consider the set of episodes that included December 1991 and were possibly censored. Data available indicate starting date, episode length at the last event, and episode type and location. We propose to use episodes that included December 1989 (and December 1990) to estimate the probability that an episode that displays a December 1991 event actually will extend into the following (unobserved) month. If 1989 (or 1990) were in fact the last year of the database, then episodes that extended into 1990 (1991) would be censored – episodes that did not extend into 1990 (1991) would not. The probit model can be used to make the 1989 (1990) prediction, and then applied to December data for 1991.

Our database contains two relevant types of December 1989 (and 1990) episodes for

estimating censoring for 1991 episodes. They are noted in Figure 1.

(Figure 1 – Probit Censoring Adjustments)

1. *Type O* (ongoing) episodes started before December 1 and were ongoing at the date of the last observed event (after December 1), thus possibly censored. We know episode length on that date. Censoring probabilities for episodes extending into December 1991, that started prior to December 1, 1991, will be predicted by probit regressions for 1989 and 1990 Type O episodes.
2. *Type A* (after) episodes started after December 1. Since this *starting date* was within 30 days of the probable censoring date, these episodes were possibly censored. As with the Type O episodes, we know how long these episodes lasted, but we do not know the sequencing of events. Censoring probabilities for December 1991 episodes that started after December 1, 1991 will be predicted by probit regressions for 1989 and 1990 Type A episodes.

The probit analyses determine whether the 1989 (and 1990) episodes were censored.

For Type O episodes:

$$Pr(\delta = 1) = \beta_0 + \beta_1 * \text{Episode Length (as of 12/1)} + \beta_2 * \text{Episode Location (inpatient or outpatient)} + \sum_{type=1}^k \beta_{3k} \text{Treatment}_k \quad (4a)$$

For Type A episodes:

$$Pr(\delta = 1) = \nu_0 + \nu_1 * \text{Start date} + \nu_2 * \text{Episode Location (inpatient or outpatient)} + \sum_{type=1}^k \nu_{3k} \text{Treatment}_k \quad (4b)$$

For episodes ongoing on December 1, 1991 we apply the Type O predictor. For episodes starting after December 1, 1991 we apply the Type A predictor.

The alternative estimator for episodes starting within the *first* 30 days of the data

window (January 1 through January 30, 1989) is analogous. We use January 1990 and January 1991 episodes to generate censoring probabilities for January 1989.

The estimation process follows this sequence:

1. Assign probability values to the possibly censored observations.
2. Draw from a probability distribution to assign coding as either censored ($\delta = 0$) or uncensored ($\delta = 1$) to possibly censored observations. Over repeated drawings individual observations will approach their asymptotic probability of being censored.
3. Estimate equation (1) and calculate survival rates.
4. Repeat Steps 2 and 3 to build a distribution of medians, vectors of β , γ , θ , and σ terms, and 95% confidence intervals around them.

Based on likelihood function (3'), one could theoretically estimate equations (1), (4a), and (4b) jointly. Such joint estimation is not now tractable, but our proposed iterative probit method is equivalent to using (3') except that p_i is estimated separately by the probit estimator.

4. A Simulation Study

Since the proposed estimator is new, we present a Monte Carlo simulation to demonstrate its properties. We generate a sample of 5,000 episode lengths using the random number generator with:

$$T = e^{2.4055} e^{\sigma W}, \text{ or: } \log T = 2.4055 + \sigma W, \quad (5a)$$

where W is an error term. This generates a median of approximately 11 days and a mean of about 30 days (both figures are consistent with our data on health care treatment episodes).

We use a lognormal distribution in the simulation because it generally characterizes our episode distribution, and it permits a nonmonotonic hazard function.

A second experiment adds a (0, 1) binary variable with a mean of 0.5 and a

coefficient of 0.1, representing either a gender or a treatment effect of approximately 10%.

We again specify the parameters to provide a median of 11 days and a mean of 30 days:

$$T = e^{2.3537 + 0.1 * X} e^{\sigma W}, \text{ or: } \log T = 2.3537 + 0.1X + \sigma W, \quad (5b)$$

Again, we use the standard normal error term.

We randomly assign each of the 5,000 lengths to a date between 0 and 1095 (approximating a three year data window), and we randomly assign (probability = 0.5) each date as the episode's starting or the ending date. We then assign "possible censoring status" if the episode started:

1. before date 0 or extended past date 1095. This would occur, for example, if we drew a 60-day episode that ended on February 15, 1989 (thus extending back before January 1, 1989). We assign an "observation date" by drawing randomly from a uniform distribution over the first thirty days of January (1989).
2. between dates 0 and 30, or between dates 1065 and 1095. We assign "observed start or end dates" by drawing from a uniform distribution between the starting date and either Day 1 or Day 1065, as applicable.

This process provides those lengths that we would see if there were no censoring, or *true* episode lengths, and those lengths that are recorded due to episode censoring, or *observed* episode lengths.

We then estimate four probit regressions using data from dates 335 – 395, and dates 700 – 760, as specified in equation (4a) or (4b). The probit regressions calculate a censoring probability for each "possibly censored" data point. We use that probability to assign "censored" status and estimate a duration model, conditional on the vector of variable δ generated by the probit probability; we repeat the process 25 times for each

probit. We replicate the entire process (generate data, estimate probit regressions, assign censored status, estimate parameters) 500 times.

We evaluate the estimator properties using the following parameters and medians:

1. True sample medians.
2. “Almost true” (AT) – estimated if the true (uncensored) lengths were known.
3. Uncensored (NC) – estimated if the observed episode lengths were analyzed with no censoring adjustment.
4. All-censored (AC) – estimated if the observed episode lengths were analyzed assuming that all potentially censored lengths were, in fact, censored.
5. Probably censored (PC) – estimated using the probable censoring method.

(Tables 1a and 1b – Monte Carlo Analyses)

Table 1a evaluates equation (5a). We summarize the estimates with the mean of the mean estimates, and the median of the median estimates. We evaluate them with MSE (mean squared error – for the mean relative to the true value), and bias (for both the mean and the median estimates, relative to the true value).

The AT model is the “gold standard” because it replicates the true values. Of the four models, the AT model provides the minimum MSE and the smallest biases. However, it contains information on the true lengths that is not available to the analyst.

By MSE standards, PC is as good as AT in estimating the intercept and only slightly worse in estimating the median. It is much better than either NC or AC. With respect to bias, PC is slightly worse than AT by either the mean or the median measure. PC bias measures are between one-eighth and one-sixteenth the size of the NC and AC measures.

Table 1b compares the various estimators for the case in which the covariate X is introduced. All three of the models (NC, AC, and PC) provide good estimates of the intercept and the β parameter, based on bias calculation. Both the NC and the AC estimates

of the intercept and β have smaller MSEs than the PC estimate, presumably because the PC estimate does more random sampling. However, the PC estimator provides the best estimates of the median by MSE criteria, only slightly larger than AT, and one-fifth to one-twelfth as large as NC and AC.

Regarding bias, the mean of means and the median of medians for PC compare most favorably with the AC and the NC estimators. For medians the mean bias is -0.0698 days (-0.6 percent of the true median), compared with -0.4596 days (-4.2 percent) for the NC estimator, and +0.7761 days (+7.1 percent) for the AC estimator.

In sum, the PC parameter estimates compare favorably with other methods, and the estimated medians are closer to the true medians than models which assume that all possibly censored observations are censored (AC), or that none are (NC).

5. Application to Treatment Episodes

This section presents our empirical results when applying the PC estimator to the database. Table 2 shows the probit censoring adjustments. Tables 3 and 4 present outpatient and inpatient episode duration estimates. Table 5 compares the survival rates across a wide range of inpatient and outpatient episodes and across censoring specifications.

Probit Censoring Adjustments

Table 2 displays the probit censoring adjustments. All were estimated with two years of data. For *Type O - December* episodes estimated with 1989 and 1990 data, the positive coefficient on EPS_LENGTH (0.0027) indicates that episodes of greater length on December 1 are more likely to extend past December 31. This implies censoring when the probit equation is applied to possibly censored 1991 observations. Inpatient episodes, with

a coefficient on EPS_LOC of -0.2661, are less likely to imply censoring. ALC, DRUG, PSYCH, or MED all imply more likely censoring than the omitted surgery category.

(Table 2 – Probit Estimate of Censoring Probabilities)

For the *Type A – December* adjustment, also estimated for 1989 and 1990, the coefficient of 0.0303 for episode starting date (START) implies that the later in December the episode starts, the more likely it will extend past December 31. Inpatient episodes are more likely to imply censoring, and again ALC, DRUG, PSYCH, or MED are all more likely to imply censoring than the (omitted) surgery category.

The January adjustments estimated for 1990 and 1991, to be applied to episodes starting in January 1989, are qualitatively similar. The *Type O – January* adjustment implies that the longer the current episode lasts after January 31, the more likely that it will have started before January 1, implying censoring. The *Type A – January* adjustment, with a negative start date coefficient, indicates that the later in January the episode started, the less likely that it will have started before January 1. Note that the coefficients of the start dates for both Type A adjustments are similar in absolute impacts (0.0303 for December; 0.0320 for January). This implies that the closer the start date to the last (for December) or first (for January) day of the window, the more likely it will either extend past December 31, or start before January 1. When applied to the possibly censored episodes, this increased likelihood implies more probable censoring.

Outpatient and Inpatient Duration Estimates

Having estimated the censoring probability equations, we draw from the probability distribution to assign coding as either censored ($\delta = 0$) or uncensored ($\delta = 1$). We estimate

equation (1) using the EGG distribution.² The process is repeated 1,000 times for each diagnosis-location combination to generate distributions of parameter estimates, medians, and confidence intervals. Tables 3 and 4 provide detailed estimates of drug treatment episodes. Table 5 compares estimated median lengths by diagnosis-location categories.

Table 3 (outpatient drug care) and Table 4 (inpatient drug care) display the episode length determinants. Column 4 in each table presents the AC model with variables separated into employee and employer categories. In Table 3, for example, a 0.1 increase in employer male percentage increases outpatient episode length by 16% ($e^{0.1 \times 1.4989}$); holding employer constant, men's (PT_MALE = 1) episodes are 2.8% shorter ($e^{-0.0283}$) than women's (PT_MALE = 0).

(Table 3 – Outpatient Drug Episode Length Estimates – EGG distribution)

Continuing with Table 3, the distinction between employee and employer variables has a particular interpretation for both coinsurance rates and deductibles. An employer that raises its coinsurance rate by 0.1 for all employees will see increases in variables EPS_CPR (Employee effect), with coefficient +1.1891 and CPR_AVG (Employer effect), with coefficient -2.2913. The net impact will be a 9.9% decrease ($e^{0.1 \times (1.1891 - 2.2913)}$) in episode length. A \$100 increase in insurance deductible has similar joint effects, with a net 18.9% decrease ($e^{100 \times (0.0044 - 0.0065)}$) in episode length. Assuming that all potentially censored observations are censored yields an estimated AC median of 15.97 days.

The PC adjustment lowers the estimated median to 11.30 days, or about 4.7 days shorter than the AC estimate. The nonparametric 95% confidence interval is plus or minus

² All parameters were estimated with the SAS LIFEREG procedure. A set of nested hypotheses (available from the senior author) shows the EGG distribution to be superior to the simpler ones.

0.35 days, and is approximately symmetric.

We also compared PC median estimates of β , γ , and θ , and their estimated 95% confidence intervals to the AC model. Finding that these medians differ from the AC model estimate does not imply that they (or the AC estimates) differ significantly from 0, but it does indicate that they differ, and that the medians provide consistent estimates of the parameter values. Consider the PC median coefficient value of 0.3216 if a patient is the primary beneficiary (PT_SELF). The resampling method suggests that the PC estimate of 0.3216 is preferable to the AC coefficient of 0.3943, which lies outside the 95% confidence interval (0.2747 to 0.3696).³

(Table 4 – Inpatient Drug Episode Length Estimates – EGG distribution)

Table 4 presents similar findings for inpatient drug estimates. The inpatient estimates also use the EGG distribution. There are separable employee and employer effects. Like the outpatient episodes, the joint effects of coinsurance rate or deductible increases reduce episode lengths. The PC median of 26.06 days is 1.9 days (or about 6.7%) shorter than the AC estimated median. As with the outpatient episodes, the median estimates of the β , γ , and θ parameters in this adjusted model differ slightly from the AC estimates.

Survival Rates

Table 5 compares censoring estimators for nine diagnosis-location combinations.⁴ Looking first at inpatient drug episodes, up to 13.54% may be censored. Ignoring the censoring provides the NC median estimate of 24.76 days; assuming that all of the

³ To compare the resampled median estimates of coefficient β to 0, one could estimate a resampled standard error as the median of the 1,000 standard errors.

⁴ Outpatient medical care had 16,540 observations, almost 4 times the next largest category (outpatient alcohol). Large sample computing constraints prevented our estimating Table 5 survival rates.

potentially censored observations are censored provides the AC median of 27.93 days, or about 12.8% higher. The problem is more severe with outpatient drug care since larger percentages of the outpatient episodes are potentially censored. The estimated AC median is 15.97 days, or 63.8% higher than the NC median.

(Table 5 – Comparison of Survival Rates Across Specifications – EGG distribution)

The PC adjustment finds 5.53% of the inpatient drug episodes to be censored with a corresponding median length of 26.06 days. Again, for the outpatient drug care, the differences are more substantial. The probit adjustment assigns 251, or 41.7% of the 602 potentially censored observations to censored status; the corresponding PC median of 11.30 days is 15.9% higher than the NC median (9.75 days) rather than 63.8% higher (15.97 days), as calculated by AC.

The table summarizes estimated survival rates across the nine episode categories describing treatment initiation with NC, PC, and AC. In all cases, estimates that adjust for censoring probability fall between the “no censoring” and the “all-censoring” estimates.

6. *Discussion and Conclusions*

This article addresses probable censoring in the estimation of hazard functions. It arises because the very nature of health care treatment episodes allows for time to elapse between observed events. If the observation “window” closes before the appropriate period of time lapses, then the duration may be censored. However, assuming that the duration is censored with a probability of one will almost certainly bias duration estimates upward.

We use resampling methods to calculate vectors of β , γ , θ , and σ terms, and 95% confidence intervals. Potentially censored observations are assigned censoring probabilities

based on a probit selection model that accounts for treatment diagnosis and location.

The resulting estimates indicate that just as ignoring the censoring problem provides estimated medians that are too low, assuming all of the potentially censored observations are censored leads to medians that are too high. The method proposed to address the probable censoring can effectively correct the biases and provide appropriate estimates of medians. From a health care system perspective, where episode length is related to utilization and costs, this finding is most important for planning and resource allocation as it helps predict individuals' exposures to the system.

This method may apply to a wide range of applications in which censoring is probabilistic rather than certain. Spells of unemployment or employment, or episodes of illness or health, may be ongoing at times of visits to employment agencies, or physicians, prior to the end of data collection periods. Assuming that all are censored may lead to biases that are potentially as serious as assuming that none are censored. This article provides a method that addresses the issue and suggests an estimator that is easy to use.

Table 1a – Monte Carlo Analysis without a Covariate

True Regression: $\log T = 2.4055 + \sigma W$: standard normal error

		Estimates							
<u>TRUE</u>		<u>Almost True (AT)</u>		<u>No Censoring (NC)</u>		<u>All Censoring (AC)</u>		<u>Probable Censoring (PC)</u>	
<u>Parameter</u>		<u>mean^a</u>	<u>median^b</u>	<u>mean</u>	<u>median</u>	<u>mean</u>	<u>median</u>	<u>mean</u>	<u>median</u>
Intercept	2.4055	2.4062	2.4059	2.3628	2.3625	2.4754	2.4756	2.4003	2.4004
50% Survival	11.0836	11.0938	11.0880	10.6227	10.6176	11.8888	11.8886	11.0281	11.0278
MSE									
Intercept		0.0004		0.0022		0.0053		0.0004	
50% Survival		0.0435		0.2525		0.7039		0.0497	
Bias									
Intercept		0.0007	0.0004	-0.0427	-0.0430	0.0699	0.0701	-0.0052	-0.0051
50% Survival		0.0102	0.0044	-0.4609	-0.4660	0.8052	0.8050	-0.0554	-0.0558

a. Mean values

b. Median values

Table 1b – Monte Carlo Analysis with a Covariate

True Regression: $\log T = 2.3537 + 0.1X + \sigma W$: standard normal error

<u>Parameter</u>	Estimates								
	<u>TRUE</u>	<u>Almost True (AT)</u>		<u>No censoring (NC)</u>		<u>All Censoring (AC)</u>		<u>Probable Censoring (PC)</u>	
		<u>mean^a</u>	<u>median^b</u>	<u>mean</u>	<u>median</u>	<u>mean</u>	<u>median</u>	<u>mean</u>	<u>median</u>
Intercept	2.3537	2.3373	2.3362	2.2962	2.2951	2.4070	2.4053	2.3281	2.3309
β	0.1000	0.0994	0.1022	0.0963	0.0991	0.0985	0.1005	0.1016	0.0999
50% survival	10.8887	10.8831	10.8729	10.4290	10.4171	11.6647	11.6464	10.8188	10.7972
MSE									
Intercept		0.0011		0.0041		0.0037		0.0111	
β		0.0016		0.0016		0.0017		0.0113	
50% survival		0.0424		0.2502		0.6551		0.0488	
Bias									
Intercept		-0.0165	-0.0175	-0.0576	-0.0586	0.0533	0.0516	-0.0257	-0.0228
β		-0.0006	0.0022	-0.0037	-0.0009	-0.0015	0.0005	0.0016	-0.0001
50% survival		-0.0055	-0.0157	-0.4596	-0.4716	0.7761	0.7577	-0.0698	-0.0915

a. Mean values

b. Median values

Table 2 – Probit Estimate of Censoring Probabilities

<u>Variable</u>	Type O - December		Type O - January	
	<u>Coefficient</u>	<u>Std. Error</u>	<u>Coefficient</u>	<u>Std. Error</u>
INTERCEPT	-0.2771	0.0508 *	-0.2561	0.0520 *
EPS_LENGTH	0.0027	0.0001 *	0.0028	0.0001 *
EPS_TYPE	-0.2661	0.0474 *	-0.0424	0.0456
ALC	0.2698	0.0615 *	-0.1302	0.0609 *
DRUG	0.2401	0.0710 *	-0.1708	0.0686 *
PSYCH	0.3852	0.0594 *	-0.0040	0.0585
MED	0.1417	0.0535 *	0.0078	0.0544

<u>Variable</u>	Type A - December		Type A - January	
	<u>Coefficient</u>	<u>Std. Error</u>	<u>Coefficient</u>	<u>Std. Error</u>
INTERCEPT	-10.9882	0.6391 *	0.0263	0.0553
START	0.0303	0.0018 *	-0.0320	0.0016 *
EPS_TYPE	0.5532	0.0617 *	0.7002	0.0560 *
ALC	0.4457	0.0667 *	0.1676	0.0630 *
DRUG	0.4893	0.0777 *	0.0993	0.0713
PSYCH	0.5564	0.0674 *	0.1706	0.0638 *
MED	0.0316	0.0533	-0.0151	0.0530

* indicates 5% significance

Table 3 – Outpatient Drug Episode Length Estimates (EGG)

	(1)	(2)	(3)	(4)
<u>Variable</u>	<u>PC median</u>	<u>PC 95% confidence interval</u>		<u>AC model estimates</u>
<i>Employee</i>				
INTERCEPT	3.0090	2.5255	3.5052	2.9295
PT_MALE	0.0115	-0.0348	0.0581	-0.0283
PT_AGE	-0.0008	-0.0025	0.0010	-0.0005
PT_HRLY	0.0698	0.0183	0.1274	0.0411
PT_ACTIVE	0.3309	0.2713	0.3850	0.4262 *
PT_SELF	0.3216	0.2747	0.3696	0.3943 *
EPS_STR0	-0.0010	-0.0010	-0.0009	-0.0020 *
EPS_CPR	0.8704	0.7168	1.0252	1.1891 *
EPS_DCT	0.0051	0.0047	0.0055	0.0044 *
<i>Employer</i>				
MALE_AVG	1.4497	0.9914	1.9316	1.4989
AGE_AVG	-0.0518	-0.0628	-0.0416	-0.0361 *
HRLY_AVG	-0.0331	-0.1509	0.0797	-0.1580 *
ACTV_AVG	-0.5387	-0.7427	-0.3159	-0.3410
SELF_AVG	1.0271	0.8342	1.2117	1.0132
CPR_AVG	-1.4843	-1.7872	-1.2227	-2.2913 *
DCT_AVG	-0.0073	-0.0081	-0.0064	-0.0065
SCALE	1.9263	1.9062	1.9469	2.0567 *
SHAPE	0.1066	0.0443	0.1554	0.0292 *
survival rate (days)				
median	11.30	10.95	11.62	15.97 *

* All-censored (AC) point estimate lies outside 95% confidence interval of PC point estimate.

Table 4 – Inpatient Drug Episode Length Estimates (EGG)

	(1)	(2)	(3)	(4)
<u>Variable</u>	<u>PC median</u>	<u>PC 95% confidence interval</u>		<u>AC model estimates</u>
<i>Employee</i>				
INTERCEPT	3.4202	3.2633	3.5788	3.4573
PT_MALE	-0.0442	-0.0624	-0.0257	-0.0214*
PT_AGE	-0.0077	-0.0085	-0.0068	-0.0079
PT_HRLY	-0.1388	-0.1591	-0.1193	-0.1347
PT_ACTVE	0.1628	0.1385	0.1846	0.1685
PT_SELF	0.0988	0.0831	0.1140	0.0992
EPS_STR0	-0.0002	-0.0002	-0.0002	-0.0004*
EPS_CPR	-1.1357	-1.2387	-1.0369	-1.0933
EPS_DCT	0.0003	0.0003	0.0004	0.0003
<i>Employer</i>				
MALE_AVG	0.6569	0.5272	0.7932	0.7547
AGE_AVG	0.0051	0.0011	0.0088	0.0052
HRLY_AVG	0.1530	0.1198	0.1877	0.1648
ACTV_AVG	-0.3271	-0.4062	-0.2522	-0.3485
SELF_AVG	-0.3704	-0.4407	-0.2945	-0.3581
CPR_AVG	0.6762	0.5220	0.8382	0.6251
DCT_AVG	-0.0007	-0.0009	-0.0006	-0.0008
SCALE	0.8712	0.8651	0.8770	0.8864*
SHAPE	0.5194	0.5051	0.5348	0.5033*
survival rate (days)				
median	26.06	25.85	26.28	27.93*

* All-censored (AC) point estimate lies outside 95% confidence interval of PC point estimate.

Table 5 – Comparison of Survival Rates in Days Across Specifications (EGG)

Inpatient Drug				Outpatient Drug			
	<u>NC</u>	<u>PC</u>	<u>AC</u>		<u>NC</u>	<u>PC</u>	<u>AC</u>
25%	44.12	46.37	50.17	25%	32.85	40.89	63.94
50% - median	24.76	26.06	27.93	50% - median	9.75	11.30	15.97
75%	12.86	13.62	14.48	75%	2.71	3.02	3.99
95%	4.23	4.58	4.83	95%	0.38	0.43	0.54
Uncensored	2031	1919	1756	Uncensored	2292	2041	1690
Censored	0	112	275	Censored	0	251	602
% Censored	0.00	5.53	13.54	% Censored	0.00	10.96	26.27
Inpatient Alcohol				Outpatient Alcohol			
	<u>NC</u>	<u>PC</u>	<u>AC</u>		<u>NC</u>	<u>PC</u>	<u>AC</u>
25%	44.53	46.74	50.66	25%	30.61	38.46	57.75
50% - median	24.03	25.14	26.94	50% - median	8.92	10.56	14.70
75%	11.91	12.43	13.12	75%	2.55	2.90	3.74
95%	3.60	3.76	3.84	95%	0.41	0.45	0.52
Uncensored	3044	2887	2667	Uncensored	4180	3735	3147
Censored	0	157	377	Censored	0	445	1033
% Censored	0.00	5.16	12.39	% Censored	0.00	10.65	24.71
Inpatient Psych				Outpatient Psych			
	<u>NC</u>	<u>PC</u>	<u>AC</u>		<u>NC</u>	<u>PC</u>	<u>AC</u>
25%	76.37	79.57	89.89	25%	75.87	115.37	269.15
50% - median	38.97	39.68	43.90	50% - median	25.06	30.68	64.42
75%	18.10	18.16	19.63	75%	6.33	7.44	13.99
95%	4.90	4.89	5.12	95%	0.46	0.81	1.28
Uncensored	432	395	357	Uncensored	2484	1930	1376
Censored	0	37	75	Censored	0	554	1108
% Censored	0.00	8.55	17.36	% Censored	0.00	22.31	44.61

Inpatient Surgery				Outpatient Surgery			
	<u>NC</u>	<u>PC</u>	<u>AC</u>		<u>NC</u>	<u>PC</u>	<u>AC</u>
25%	42.26	52.32	72.88	25%	16.92	20.34	29.70
50% - median	17.62	20.31	24.75	50% - median	5.37	6.12	8.26
75%	7.03	7.93	8.9	75%	1.70	1.84	2.30
95%	1.72	2.06	2.24	95%	0.33	0.33	0.36
Uncensored	152	130	109	Uncensored	2536	2301	1942
Censored	0	22	43	Censored	0	235	594
% Censored	0.00	14.25	28.29	% Censored	0.00	9.26	23.42
Inpatient Medical							
	<u>NC</u>	<u>PC</u>	<u>AC</u>				
25%	33.91	39.27	46.66				
50% - median	12.73	13.99	15.61				
75%	4.78	5.15	5.52				
95%	1.17	1.29	1.35				
Uncensored	307	280	254				
Censored	0	28	53				
% Censored	0.00	8.96	17.26				

References

1. Stern S, Merwin E, Holt F, Survival models of community tenure and length of hospital stay for the seriously mentally ill: A 10-year perspective, *Health Services and Outcomes Research Methodology* 2001; **2**:117-135.
2. Goodman AC, Hankin JR, Kalist DE, Peng Y, Spurr SJ, Estimating multiple episodes for substance abusers, *Journal of Mental Health Policy and Economics* 2001; **4**:65-77.
3. Yamaguchi K, 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* 1992; **87**:284-292.
4. Peng Y, Dear KBG, Denham JW, A generalized F mixture model for cure rate estimation, *Statistics in Medicine* 1998; **17**:813-830.
5. Hougaard P, Fundamentals of survival data, *Biometrics* 1999; **55**:13-22.
6. van der Laan MJ, McKeague IW, Efficient estimation from right-censored data when failure indicators are missing at random, *The Annals of Statistics* 1998; **26**:164-182.
7. McKeague IW, Subramanian S, Product-limit estimator and Cox regression with missing censoring information, *Scandinavian Journal of Statistics* 1998; **25**:589-601.
8. van der Laan MJ, Hubbard A, Locally efficient estimation of the survival distribution with right censored data and covariates when collection of data is delayed, *Biometrika* 1998; **85**:771-783.
9. van der Laan MJ, Robins JM, *Unified Methods for Censored Longitudinal Data and Causality*, Springer Verlag: New York, 2003.
10. Pohlmeier W, Ulrich V, An econometric model of the two-part decision process in the demand for health, *Journal of Human Resources* 1995; **30**:339-361.

11. Santos Silva JMC, Windmeijer F, Two-part multiple spell models for health care demand, *Journal of Econometrics* 2001; **104**:67-89.
12. Romeo CJ, Controlling for seam problems in duration model estimates, *Journal of Human Resources* 2001; **36**:467-499.

Variable Definitions

Table 2

START – episode start date (January 1 = 1; December 31 = 365)
 EPS_LENGTH – length of Type O episode in days on December 1 (after January 31)
 EPS_TYPE – 1 if inpatient; 0 otherwise
 ALC – 1 if initiating event was alcohol treatment; 0 otherwise
 DRUG – 1 if initiating event was drug treatment; 0 otherwise
 PSYCH – 1 if initiating event was psychiatric treatment; 0 otherwise
 MED – 1 if initiating event was medical treatment; 0 otherwise

Tables 3 and 4

PT_MALE – 1 if patient is male; 0 otherwise
 PT_AGE – patient age in years
 PT_HRLY – 1 if the patient is an hourly employee; 0 otherwise
 PT_ACTIVE – 1 if the patient is currently employed; 0 otherwise
 PT_SELF – 1 if the patient is the primary beneficiary; 0 otherwise
 EPS_STR0 – starting date of episode (January 1, 1989 = 1; December 31, 1991 = 1065)
 EPS_CPR – episode coinsurance rate (0 = no out-of-pocket; 1 = full)
 EPS_DCT – episode deductible in dollars
 MALE_AVG – employer (year-specific) mean of covered employees who are male
 AGE_AVG – employer (year-specific) mean age of covered employees
 HRLY_AVG – employer (year-specific) mean of hourly employees
 ACTV_AVG – employer (year-specific) mean of currently employed subjects
 SELF_AVG – employer (year-specific) mean of employees who are primary beneficiaries
 CPR_AVG – employer (year-specific) mean of coinsurance rate
 DCT_AVG – employer (year-specific) mean deductible payment
 TYP_AVG – employer (year-specific) mean of inpatient v. outpatient episode
 SCALE – statistical parameter σ
 SHAPE – statistical parameter q

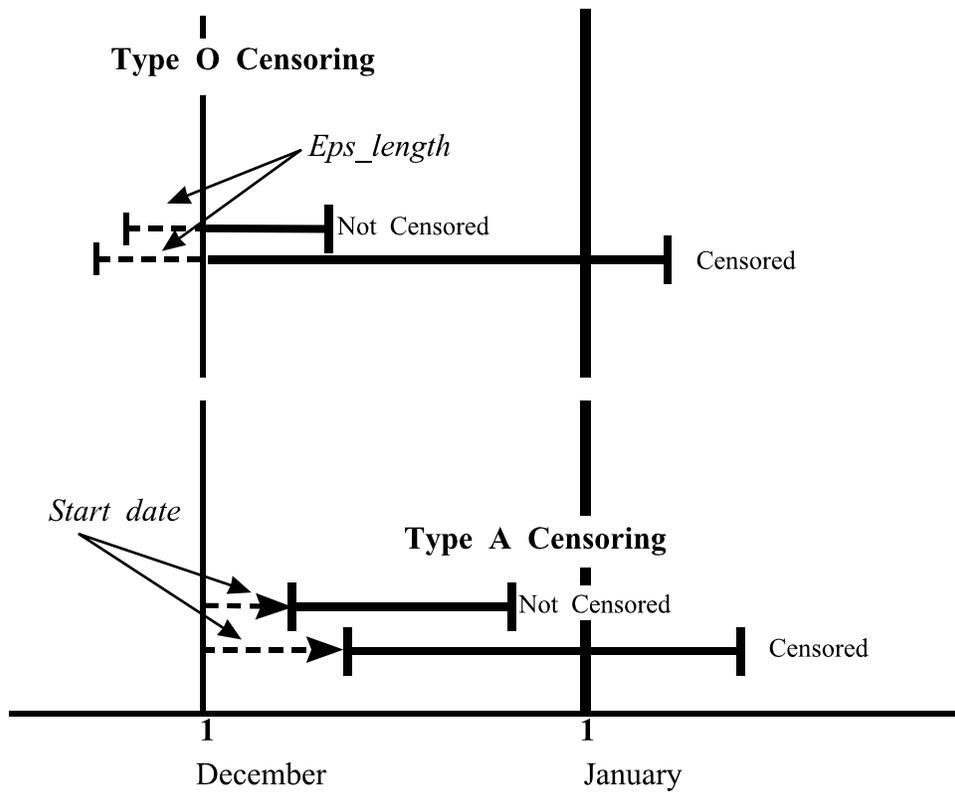


Figure 1 Probit Censoring Adjustments for End-of-Year Episodes