Testing Goodness-of-Fit of a Uniform Truncation Model

Micha Mandel* and Rebecca A. Betensky

Department of Biostatistics, Harvard School of Public Health,

Boston, Massachusetts 02115

* email: mmandel@hsph.harvard.edu

August 25, 2006

SUMMARY

Several goodness-of-fit tests of a lifetime distribution have been suggested in the literature; many take into account censoring and/or truncation of the event times. In some contexts, a goodness-of-fit test for the truncation distribution is of interest. In particular, better estimates of the lifetime distribution can be obtained when knowledge of the truncation law is exploited. In cross-sectional sampling, for example, there are theoretical justifications for the assumption of a uniform truncation distribution, and several studies have used it to improve the efficiency of their survival estimates. The duality of lifetime and truncation in the absence of censoring enables methods for testing goodness-of-fit of the lifetime distribution to be used for testing goodness-of-fit of the truncation distribution. However, under random censoring, this duality does not hold and different tests are required. In this paper, we introduce several goodness-of-fit tests for the truncation distribution and investigate their performances in the presence of censored event times using simulation. We demonstrate the use of our tests on two data sets.

KEY WORDS: Paired-test; Random censoring; Residual-lifetime; Size-bias; Survival.

1 Introduction

Left truncation is a well studied model in survival analysis in which lifetimes are observed conditionally on being greater than random truncation times. Studies have shown that estimation of the lifetime distribution can be improved when the truncation law is known (Wang, 1989, 1991, Huang and Wang, 1995, Asgharian, M'Lan, and Wolfson, 2002, de Uña-Álvarez, 2004a). In certain cases, such as in cross-sectional sampling, there are theoretical justifications for the assumption of a uniform truncation distribution (Laslett, 1982, Vardi, 1982, 1989). For example, under the assumption of uniformity, Asgharian M'Lan and Wolfson (2002) estimated survival of dementia patients and de Uña-Álvarez (2004a) estimated the distribution of unemployment spells of women. However, goodness-of-fit testing of the uniform assumption was conducted by informal graphical tests, as suggested by Wang (1991).

In addition to the importance of the uniform assumption for efficient estimation of the lifetime distribution, the truncation distribution may be of independent interest. In cross-sectional sampling, truncation signifies the incidence rate of the onset of a disease (Keiding, 1991, Asgharian et.al., 2006), and it is often of interest to test for constancy of that rate. Another example arises in AIDS studies and is discussed in Section 5.

In this paper, we propose several formal goodness-of-fit tests for the assumption of a uniform truncation distribution and to a lesser extent for other specified distributions. Since the assumption of a uniform truncation distribution is most helpful in small data sets, for which the nonparametric maximum likelihood estimator has large variance and may not even exist, our goal is to develop tests that are valid for small samples and do not require estimation of the failure or truncation distributions. In Section 2, we present our model and describe some probabilistic properties of the uniform truncation model and briefly review nonparametric estimation. We exploit these properties in Section 3 to introduce tests for uniformity. In

Section 4, we present a simulation study in which we compare the performances of these tests, and in Section 5 we apply them to the Channing House data (Hyde, 1977) and to AIDS incubation data (Kalbfleisch and Lawless, 1989, Wang, 1989). We conclude the paper with a discussion in Section 6.

2 The Truncation Model

2.1 The Left Truncation Right Censoring Model

Let $T^* \sim G^*$ and $X^* \sim F^*$ be independent positive random variables. In this paper, we consider sampling from the truncation region $T^* < X^*$, i.e., observations are realizations of $(T^*, X^*)|T^* < X^*$. We distinguish between elements of the truncated and non-truncated probability spaces by superscripting the latter with asterisks. Thus, for example, $X \sim F$ has the law of $X^*|T^* < X^*$ and $T \sim G$ has the law of $T^*|T^* < X^*$. To emphasis their different role, we call T^* the truncation time and X^* the lifetime.

In addition to truncation, data are often subject to censoring. We consider the random censoring model where censoring occurs after truncation, and de facto the censoring is of the residual lifetime $R = X - T$ and not of X. Thus, there is a random variable $C \sim F_C$ that is independent of (T, X) and the data comprise of realizations from (T, Z, Δ), where $Z = T + \min(R, C)$, $\Delta = I\{R \leq C\}$ and I is the indicator function. This type of censoring is considered by Vardi (1989), Wang (1991) and others. In the framework of cross-sectional sampling, C represents the length of the follow-up period.

2.2 Properties

Several important properties of the uniform truncation model are described next and will be used in Section 3 as the basis for goodness-of-fit tests.

Let $t_{\min}$ and $t_{\max}$ be the lower and upper limits of the support of G^*, i.e., $t_{\min} = \inf\{t|G^*(t) > 0\}$ and $t_{\max} = \sup\{t|G^*(t) < 1\}$, and let $x_{\min}$ and $x_{\max}$ be the corresponding values of F^*. We assume: (I) $0 \leq t_{\min} \leq x_{\min}$ and $t_{\max} \leq x_{\max} < \infty$; (II) $x_{\min} < t_{\max}$, i.e., $P(T^* < X^*) < 1$; (III) G^* and F^* have continuous densities g^* and f^*, respectively, which are strictly positive over their support. Assumption (I) is needed for identifiability (e.g., Woodroofe, 1985) and to avoid technical difficulties of dealing with improper uniform distributions. Assumption (II) states that the truncation is "real", and the continuity assumption (III) holds in most real life examples. The density of the pair (T, X) is

$$f_{T,X}(t,x) = \frac{g^*(t)f^*(x)}{\mu} \qquad t < x \tag{1}$$

where $\mu = \int_0^\infty g^*(u)\bar{F}^*(u)du$ and $\bar{F}^* = 1 - F^*$. The marginal densities of T and X are

$$g(t) = \frac{g^*(t)\bar{F}^*(t)}{\mu} \qquad t_{\min} < t < t_{\max}, \tag{2}$$

$$\text{and} \qquad f(x) = \frac{f^*(x)G^*(x)}{\mu} \qquad x_{\min} \leq x \leq x_{\max}. \tag{3}$$

The model $G^* = \mathrm{U}(0, t_{\max})$ exhibits some interesting and useful features. When $t_{\max} = x_{\max}$, the densities in (2) and (3) above are $\bar{F}^*(t)/\mathbb{E}X^*$ and $xf^*(x)/\mathbb{E}X^*$, the residual lifetime distribution of a stationary renewal process and the size-biased version of F^*, respectively. Furthermore,

Theorem 2.1. *Suppose that $t_{\max} = x_{\max} = b < \infty$, $t_{\min} = x_{\min} = 0$ and let $Q = T/X$ with distribution F_Q. Under the truncation model (1), $F_Q = \mathrm{U}(0,1)$ iff $G^* = \mathrm{U}(0,b)$.*

Theorem 2.1, proved in Web Appendix A, characterizes the uniform truncation model in terms of the distribution of the ratio Q. This will be used later as a basis for a goodness-of-fit

test. Also, when $G^* = \mathrm{U}(0, b)$, Q and X are independent, as can be seen from the proof of the theorem, and an efficient way of generating (T, X) is by generating independently X from (3) and Q from $\mathrm{U}(0, 1)$ and letting $T = XQ$. This will be used to generate simulated data and to construct bootstrap estimates of variance.

Another quantity of interest is the residual lifetime, $R = X - T$. R has density

$$dH(r)/dr = h(r) = \frac{\int_r^{x_{\max}} g^*(z - r) f^*(z) dz}{\mu} \qquad 0 < r < x_{\max} - t_{\min}. \tag{4}$$

Inspection of (2) and (4) shows that when G^* is uniform, $t_{\min} = 0$ and $t_{\max} = x_{\max}$, then $H = G$. We propose a test based on this equivalency of distributions in the next section. The assumptions concerning the limits of the supports are important since when $x_{\max} > t_{\max}$ or $t_{\min} > 0$, G and H are supported on different intervals, and hence cannot be equal. Also, if $x_{\max} < t_{\max}$, then the shape of g^* over $(x_{\max}, t_{\max})$ does not affect G and H and a non-uniform G^* can be constructed such that $H = G$. Interestingly, $H = G$ does not imply that G^* is uniform even when $x_{\max} = t_{\max}$. For example, if F^* assigns mass 1 to $x_{\max}$, then $G = H$ if g^* is symmetric, i.e., $g^*(t) = g^*(x_{\max} - t)$ for all t. However, subject to assumptions (I)-(III) above, Asgharian et.al. (2006) show that

Theorem 2.2. *Suppose that $t_{\max} = x_{\max} = b < \infty$ and $t_{\min} = x_{\min} = 0$. Under the truncation model (1), g^* is constant iff T and R have the same law.*

When g^* signifies the incidence rate of some event, it is often of interest to test for a constant versus an increasing or decreasing rate. The next theorems, proved in Web Appendix A, are concerned with this situation.

Theorem 2.3. *Under the assumptions of Theorem 2.2, if g^* is decreasing (increasing) then G is stochastically smaller (larger) than H.*

Theorem 2.4. *Under the assumptions of Theorem 2.1, if g^* is decreasing (increasing) then Q is stochastically smaller (larger) than U, where $U \sim \mathrm{U}(0, 1)$.*

Finally, we note that for the general problem of testing $H_0 : G^* = G_0$ for a known continuous G_0, (i) G_0 is strictly increasing hence $\{G_0(T^*) < G_0(X^*)\} \Longleftrightarrow \{T^* < X^*\}$, (ii) the random variables $G_0(T^*)$, $G_0(X^*)$ are independent, and (iii) under H_0, $G_0(T^*) \sim \mathrm{U}(0,1)$. This observation enables extension of $H_0 : G^* = \text{Uniform}$ to $H_0 : G^* = G_0$ in the uncensored case by simply transforming the data using G_0. Similarly, in the uncensored case for the problem $H_0 : F^* = F_0$ for a known continuous F_0, the tests presented here can be employed by reversing the time, i.e., by using the transformation $\bar{F}_0$ and the variables $\bar{F}_0(X) < \bar{F}_0(T)$. The hypothesis $H_0 : F^* = F_0$ is of interest for a better estimation of G^* in the case of right truncation where T^* is the lifetime of interest and X^* is the truncation variable.

2.3 Estimation

This section reviews estimation of (F^*, G^*) under the non-parametric truncation model, i.e., without any constraints on G^* and F^*. The estimate will be used in the following sections for goodness-of-fit testing. For estimation under the uniform truncation model, the reader is referred to Vardi (1989).

Instead of estimating $F^*(\cdot)$ and $G^*(\cdot)$, it is necessary to estimate the conditional distributions $F^*(\cdot|x_0) = P(X^* \leq \cdot | X^* \geq x_0)$ and $G^*(\cdot|t_0) = P(T^* \leq \cdot | T^* \leq t_0)$ for some values t_0, x_0 (e.g., Wang, Jewell and Tsai, 1986). Thus, we artificially impose the additional truncation $T^* \leq t_0$ and $X^* \geq x_0$. Consider n independent realizations $(t_1, z_1, \delta_1), \ldots, (t_n, z_n, \delta_n)$ of (T, Z, Δ) and let $m(y) = \sum_i I\{t_i \leq y \leq z_i\}$ and $d(y) = \sum_i I\{z_i = y, \delta_i = 1\}$. Assuming no ties, $F^*(\cdot|x_0)$ is estimated by the product-limit equation

$$F^*_{PL}(y|x_0) = 1 - \prod_{\{i: x_0 \leq z_i \leq y\}} \frac{m(z_i) - d(z_i)}{m(z_i)}, \tag{5}$$

and $G^*(\cdot|t_0)$ is estimated by an inverse weights formula as the discrete distribution

$$dG^*_{IW}(t_i|t_0) = \frac{\{1 - F^*_{PL}(t_i|x_0)\}^{-1}}{\sum_{\{i: t_i \leq t_0\}} \{1 - F^*_{PL}(t_i|x_0)\}^{-1}} I\{t_i \leq t_0\}. \tag{6}$$

The estimates given by (5) and (6) are jointly the nonparametric maximum likelihood estimate of $(F^*(\cdot|x_0), G^*(\cdot|t_0))$ (Wang, 1991), which exists if $m(z) > 0$ for all $x_0 \leq z \leq \max(z_i)$ (Wang et.al., 1986).

In choosing x_0 and t_0, one should bear in mind that these values impact the parameters that are estimable through this conditional approach: $F^*(\cdot|x_0)$ and $G^*(\cdot|t_0)$. The values x_0 and t_0 should be as close as possible to the values that define the original parameters (i.e., $x_0 = 0$ and $t_0 = \infty$), but should be such that the sizes of the risk sets $m(z_i)$ are sufficiently large so the estimators have small mean square errors. In practice, choice of these values is guided by the data (see Wang, 1991, Li and Doss, 1993).

3 Goodness-of-Fit Tests

This section presents two classes of goodness-of-fit tests for the hypothesis $H_0 : G^*\ uniform$ based on specific features of the uniform truncation model given in Theorems 2.1 and 2.2 and two additional tests derived by general principles. The tests are developed for a sample of n independent triplets (T_i, Z_i, Δ_i) (or pairs (T_i, X_i) for uncensored data), where the usual convention of uppercase letters denoting random variables and lowercase letters denoting the corresponding realizations is used.

3.1 Compare $\hat{F}_Q$ to $\mathrm{U}(0,1)$

Theorem 2.1 reduces the problem of testing uniformity of G^* to testing $Q \sim \mathrm{U}(0,1)$. This can be conducted in various ways and a natural omnibus test is the two-sided one-sample Kolmogorov-Smirnov test, which compares the empirical distribution of $Q_i = T_i/X_i$ to the uniform distribution. For the important case of monotone alternatives, Theorem 2.4 suggests the use of the one-sided one-sample Kolmogorov-Smirnov test. The procedure inherits the

consistency property of the Kolmogorov-Smirnov test and will eventually detect any non-uniform G^* satisfying the conditions of Theorem 2.1. It can be applied to the more general problem of $H_0 : G^* = G_0$, throughout application of the transformation discussed at the end of Section 2.2. However, since the test statistic is a function of the X_i's, the test is appropriate only for uncensored data.

3.2 Compare G to H

Theorem 2.2 replaces the problem of testing uniformity of G^* to that of testing equivalency of G and H. Asgharian et.al. (2006) use this property for graphical examination of the uniform truncation model and in this section we present formal goodness-of-fit tests.

As the truncation and residual lifetime within individuals are not independent, a paired test is needed. For uncensored data, the Wilcoxon sign-rank test can be used and an exact P-value is easily obtained. Theorem 2.3 states that for a monotone g^*, G and H are stochastically ordered, and since the Wilcoxon sign-rank test is consistent for ordered alternatives, it is a consistent test for a uniform against a monotone g^*.

In the general case, recall that T is uncensored and R is randomly censored by C, hence Jung's (1999) class of weighted log-rank tests for paired censored data can be utilized. This class takes the form

$$W = \sqrt{n} \int_0^\infty w_n(u)\{\hat{\Lambda}_R(du) - \hat{\Lambda}_T(du)\}, \tag{7}$$

where $\hat{\Lambda}_R$ and $\hat{\Lambda}_T$ are the Nelson-Aalen estimators of Λ_R and Λ_T, the cumulative hazard functions of R and T, based on the marginal data $(z_1 - t_1, \ldots, z_n - t_n, \delta_1, \ldots, \delta_n)$ and $(t_1, \ldots, t_n)$, and $w_n(\cdot)$ is a nonnegative bounded predictable process satisfying mild conditions of convergence to a function $w(\cdot)$ (see Jung, 1999). Jung proves that W has a limiting normal distribution and gives the formula of the variance under a general depen-

dence structure. He also shows that the test is consistent provided that $\int_0^\infty w(u)\{\Lambda_R(du) - \Lambda_T(du)\} \neq 0$. Let $m_T(u) = \sum_i I\{t_i \geq u\}$ and $m_R(u) = \sum_i I\{z_i - t_i \geq u\}$ and write $w_n(u) = n^{-1}\tilde{w}_n(u)m_T(u)m_R(u)/\{m_T(u) + m_R(u)\}$. The paired log-rank test uses $\tilde{w}_n = 1$ and the Prentice-Wilcoxon statistic uses $\tilde{w}_n(u) = \hat{S}(u)$, where $\hat{S}$ is a left continuous version of the Kaplan-Meier estimate from the pooled sample. These special cases of w_n satisfy the regularity conditions needed for asymptotic normality (Gangnon and Kosorok, 2004) and are used in the simulation study and data analysis reported in Sections 4 and 5.

Other tests for paired censored data exist and can be used to compare G and H, but a comprehensive review is beyond the scope of the current paper. The reader is referred to Woolson and O'Gorman (1992) who review and compare several of the tests suggested in the 1980's, and to Murray (2001) who reviews more recent work and suggests a family of tests based on weighted survival differences (similar to (7) with survival estimators replacing the hazard estimators).

3.3 Other Tests

3.3.1 Embed the model into a parametric family

This approach embeds G_0 into a family of distributions $\{G_\theta; \theta \in \Theta\}$ and tests for $\theta = 0$. For uncensored data, a test based on the maximum conditional likelihood of $T_i|X_i = x_i$,

$$\text{argmax}_{\theta\in\Theta} \prod_{i=1}^{n} \frac{g_\theta(t_i)}{G_\theta(x_i)}$$

can be used. Asymptotic properties of such estimators are given in Andersen (1970) and are valid under the usual regularity conditions applied to $g_\theta(t)/G_\theta(x)$ under the sampling law $g^*(t)f^*(x)/P(T^* < X^*)$. For asymptotic results concerning the random left truncation model see Wang (1989) and Li, Qin and Tiwari (1997). Specifically, the uniform distribution can be tested by embedding it within the Beta(α, β) family of distributions and using the

likelihood ratio test.

The test is tailored to alternatives $\{G_\theta; \theta \neq 0\}$ against which it is consistent (see Andersen, 1970). It is also applicable to G_0 other than the uniform and for the more general problem of testing a parametric model. However, extension of the test to censored data is not straightforward since the conditional likelihood of $T|Z$ or $T|Z, \Delta$ includes terms other than G^* that are unknown, i.e., depend on F^*. Inference on θ based on the full likelihood under the semi-parametric model of Wang (1989) $\{G^* = G_\theta\ ,\ F^*$ unspecified$\}$ has not been studied yet for censored data and is a topic for future research.

3.3.2 Compare the distribution under H_0 to a nonparametric estimate

The tests suggested above use properties of the model that eliminate estimation of F^* and G^*. This is a significant advantage for small samples where truncation introduces difficulties in estimation (see Section 2.3). For large samples, Wang (1991) studies the properties of the nonparametric maximum likelihood estimator (NPMLE) of G^* and showes that its asymptotic distribution is Gaussian. This can be used in formal tests. Let $I_1, \ldots, I_k$ be a partition of $[0, t_0]$, then

$$\zeta = \sqrt{n}\ \{G^*_{IW}(I_1|t_0) - G_0(I_1)/G_0(t_0), \ldots, G^*_{IW}(I_k|t_0) - G_0(I_k)/G_0(t_0)\}$$

has an asymptotic mean zero normal distribution with a covariance matrix Σ. Thus, under the null, the statistic $\zeta'\hat{\Sigma}^{-1}\zeta$ has an asymptotic chi-square distribution with $k-1$ degrees of freedom, where $\hat{\Sigma}$ is a consistent estimator for Σ. This test is a special case of generalized Pearson chi-square tests developed by Li and Doss (1993) for a more general situation where G^* assumes a parametric form and the partition is data dependent.

Consistent formulae for $\hat{\Sigma}$ can be found in Wang et.al. (1986) for the uncensored case and in Wang (1991) for censored data. The simulation study described in the next section

reveals that use of the asymptotic formula for $\hat{\Sigma}$ results in an overly optimistic test. This, and the untractable form of the asymptotic variance in the censored case, suggest the use of a bootstrap estimate for Σ, similar to that used by of Qin and Wang (2001) (see also Wang, 1991, Bilker and Wang, 1997 and Asgharian et.al., 2002). This consists of the following steps

1. Choose x_0 and t_0 as described in Section 2.3 and calculate $F^*_{PL}(\cdot|x_0)$. Use $(z_1 - t_1, \ldots, z_n - t_n, 1-\delta_1, \ldots, 1-\delta_n)$ to calculate $\hat{F}_C$, the estimate of the censoring distribution F_C, by the Kaplan-Meier formula.

2. Generate B samples from $\hat{F}^{SB}$, where $\hat{F}^{SB}(dx) \propto \min(x, t_0) F^*_{PL}(dx|x_0)$. Denote the observations by x_{bi} $(i = 1, \ldots, n\,,\, b = 1, \ldots, B)$. Independently, generate c_{bi} from $\hat{F}_C$ and u_{bi} from $\mathrm{U}(0,1)$.

3. Estimate $G^*(\cdot|t_0)$ from the data $(t_{bi}, z_{bi}, \delta_{bi}) = (u_{bi} \min(x_{bi}, t_0), \min(x_{bi}, t_{bi}+c_{bi}), I\{x_{bi} < t_{bi} + c_{bi}\})$ as described in Section 2.3 and denote the estimate by $G^*_{IWb}(\cdot|t_0)$. Calculate $\zeta_b = \sqrt{n}\, \{G^*_{IWb}(I_1|t_0) - G_0(I_1)/G_0(t_0), \ldots, G^*_{IWb}(I_k|t_0) - G_0(I_k)/G_0(t_0)\}$.

4. Calculate $\hat{\Sigma} = B^{-1} \sum_b \zeta_b \zeta_b'$.

Steps 2 and 3 utilize the fact that for G^* uniform and the truncation model $\{\max(T^*, x_0) \leq \min(X^*, t_0)\}$, $F_X(dx) \propto \min(x, t_0) F^*(dx) I\{x > x_0\}$ and $T|X = x \sim \mathrm{U}(0, \min(x, t_0))$. Alternatively, one can generate x^* from $F^*_{PL}(\cdot|x_0)$ and t^* from $\mathrm{U}(0, t_0)$, retain the value if $\{\max(t^*, x_0) \leq \min(x^*, t_0)\}$ and proceed until the data contain n observations.

This procedure is very similar to Wang's "obvious bootstrap" method for truncated data (Wang, 1991). We note that a rigorous proof for the consistency of the method for the variance of G^*_{IW} has not yet been given, although a simulation study suggests its validity (Bilker and Wang, 1997).

4 Simulation

We conducted an extensive simulation study to evaluate the performance of the tests. We considered all combinations of truncation and lifetime distributions belonging to {U(0,1), Beta(1,1.5), Beta(1.5,1), Beta(2,4), Beta(4,2), Beta(3,3)}. These exemplify densities that are constant, decreasing, increasing, right tailed, left tailed and unimodal-symmetric. For different sample sizes ($n = 50, 100$ and 200), we simulated 400 data sets from each of the 36 possible models and calculated the P-values of the tests. In addition, we subjected the same data sets to two types of censoring. The first used a fixed censoring time, C_α^f, which was motivated by cross-sectional studies in which censoring represents the end of follow-up. The second used random censoring times, C_α^r, generated from $F_C(x) = x^\gamma$. We selected C_α^f and C_α^r such that the probability of censoring is $1 - \alpha$ for $\alpha = 0.25, 0.50, 0.75$. Tables 1 and 2 present the results of several of the models that show the most interesting features. We include detailed tables of the simulation results and graphs presenting the properties of the different models in Web Appendix B.

In Table 1, for uncensored data, we compare the power of the Kolmogorov-Smirnov test (KS), the Wilcoxon sign-rank test (WSR), the Paired Log-Rank test (PLR), the Prentice-Wilcoxon test (PW), the conditional likelihood ratio test (CLR) and the Pearson chi-square test with asymptotic (CH2a) and bootstrap (CH2b) variance estimates. The tests have level 0.05. To calculate the chi-square tests, we partitioned $(0, t_0)$ into k equal length intervals, where t_0 was chosen to be the 90th empirical percentile of $z_1, \ldots, z_n$ and $k = 3$ or 4 for $n = 50, 100$ or 200. This choice of t_0 ensured that the risk sets would not be too small, and the choice of k reflects an increase in the number of intervals and in the number of observations in each interval when sample size increases. The bootstrap estimate of variance was based on 200 bootstrap samples. The lifetimes are all generated from F^*=Beta(3,3).

All tests except CH2a are of the desired level. The CH2a test is anti-conservative, and although its level reduces when sample size increases, it is still as high as 0.08 for $n = 200$. The CLR performs very well and in most models has the highest power (this is more pronounced in other models for F^* presented in Web Appendix B). It is hard to rank the other tests. The WSR, PLR and PW tests perform well in most cases but failed to reject the model G^* =Beta(2,4). This poor performance is seen also for the model {G^*=Beta(2,4), F^*=Beta(1,1.5)}, but not in the other models we considered (see Tables 1-3 in Web Appendix B). The KS and CH2b are consistent for general alternatives and perform reasonably well where WSR, PLR and PW fail, but in other models their performance is moderate. Use of one-sided tests for monotone alternatives improves the power of the tests (see Table 4 in Web Appendix B). For example, for the model {G^*=Beta(1.5,1), F^*=Beta(3,3)} with $n = 50$, the one-sided KS and WSR tests have power of .698 and .813, respectively, which is an improvement over .620 and .690, the power of the corresponding two-sided tests given in Table 1.

Table 2 compares the CH2b, PLR and PW tests for censored data (the estimated levels of the tests are given in Table 6 in Web Appendix B). In general, the power of the tests increases with sample size and decreases with the probability of censoring. However, for the PLR and PW tests, the fixed censoring time imposes testing based only on the left tail of the truncation and residual lifetime distributions, and may change the performance of the test for models in which the hazards of these distributions cross. This is illustrated by the model {G^*=Beta(2,4),F^*=Beta(1,1.5)}, for which the power of the PLR test is smallest for $P(\Delta = 0) = 0.50$ and the power of the PW test is highest for $P(\Delta = 0) = 0.75$ (see Figure 4 in Web Appendix B for comparison of the hazards).

To summarize, if the data are not censored and the Beta family is a reasonable model, we recommend use of the CLR test. When simplicity and ease of programming are important

factors, then the KS or WSR tests can be employed and give exact P-values. For censored data and monotone alternatives, the PW or PLR tests with one-sided P-values perform very well and should be used. If no knowledge about G^* is available, the CH2b is probably the test of choice, at least for moderate and large data sets.

[Tables 1 and 2 about here]

5 Data analysis

We further illustrate the tests using two well known truncated data sets. Our first example concerns the Channing House data (Hyde, 1977). The data contain lifetimes of residents in a retirement community in Palo Alto, California. The lifetime X^* is left truncated by the age at entry to the community T^* and it is right censored due to end of study or loss to follow-up. Wang (1991) analyzed the data and estimated the lifetime distribution of male residents conditionally on potential age being greater than $x_0 = 866$ months (she did not truncate G^*, i.e., $t_0 = \infty$). She also estimated the distribution of age at entry and, using her graphical test concluded that the uniform assumption is reasonable. However, the PLR, PW and CH2b P-values are 0.057, 0.033 and 0.021, respectively, which provide some evidence against the uniform truncation assumption. A careful look at the NPMLE of G^* suggests uniformity of the age at entry distribution only after 800 months (Wang, 1991 Figure 3 (b)). Repeating the analysis for age at entry greater than 800 (i.e., changing the truncation distribution to $G(\cdot)/\{1 - G(800)\}$) yielded P-values of 0.457, 0.407 and 0.156. Only one individual from Wang's original cohort of those observed to survive beyond age 866, had entered the community before age 800 months and was thus excluded from the second analysis. Considering this individual as an outlier, the benefit of using the uniform assumption is demonstrated in Figure 1. On the left side, the unconditional estimate of

$1 - F^*(\cdot|866)$ that exploits the uniform assumption and the conditional estimate that does not are shown. The first was calculated using the EM algorithm of Vardi (1989), and the second by the product-limit method described in Section 2.3. The right side compares the widths of 95% pointwise confidence intervals. We calculated them using 5000 simple bootstrap samples (Gross and Lai, 1996). Although the survival estimates are very close, and both lie inside the intersection of the confidence regions, the variance of the unconditional estimator is substantially smaller than that of the conditional estimator at the left tail. This same feature is illustrated by Asgharian et.al. (2002) using data on survival with dementia.

[Figure 1 about here]

The second example concerns incubation time of HIV, which has been the target of numerous studies. The data were gathered by retrospective ascertainment of AIDS patients who were thought to be infected by blood transfusion, and were reported and analyzed by Kalbfleisch and Lawless (1989) and Wang (1989). Let X^* denote time from infection (transfusion) to last day of enrollment (July 1, 1986) and T^* be the incubation time (from HIV infection to AIDS). Since the enrollment criterion is diagnosis of AIDS, the sample space is truncated to $T^* < X^*$. In this example, the data are uncensored. For illustration purposes, we test whether the parametric results of Kalbfleisch and Lawless (1989) for children and for adults apply to the elderly sub-sample. Specifically, Kalbfleisch and Lawless fit Weibull models with hazards $\alpha t^{\beta-1}$ at t (t is measured in years) to the incubation time. The parameter estimates for children are $\alpha_c = .341$, $\beta_c = 1.845$ and for adults are $\alpha_a = .015$, $\beta_a = 2.353$. The P-values of various tests for the hypothesis $T^* \sim Weibull(\alpha_i, \beta_i)$ $(i = a, c)$ are listed in Table 3. All tests reject the hypothesis that $T^* \sim Weibull(\alpha_c, \beta_c)$ and support the hypothesis that $T^* \sim Weibull(\alpha_a, \beta_a)$. The NPMLEs of the incubation distributions of adults and elderly patients are in good agreement as seen in Figure 3 of Kalbfleisch and Lawless (1989).

The conditional likelihood of elderly people is maximized for $\beta = 2.3$ and $\alpha \to 0$, illustrating that the maximum conditional likelihood does not exist in some applications (hence the CLR cannot be used).

[Table 3 about here]

6 Discussion

The tests described here can be easily extended to the model in which censoring occurs before truncation (e.g., Sun and Zhu, 2000, de Uña-Álvarez, 2004b). In this model, the censoring variable C_b^*, the lifetime X^* and the truncation T^* are independent and the sample space is truncated to $T^* < Z^* \equiv \min(X^*, C_b^*)$. The data are comprised of realizations from (T, Z, Δ), where as before $\Delta = I\{Z = X\}$ and variables without asterisks belong to the truncated space. This model is equivalent to a truncated competing risks model under the assumption of independence. Denote by F_{Z^*} the distribution function of Z^* and assume that F_{Z^*} satisfies the conditions required of F^*, then one can consider (T, Z) as uncensored data for goodness-of-fit purposes, exactly as done before with the uncensored pairs (T, X). That is, for testing hypotheses about G^*, the data can be regarded as uncensored. Huang and Wang (1995) consider competing risk models subject to truncation and censoring and de Uña-Álvarez and Rodríguez-Casal (2006) study such models under uniform truncation. In these models, the truncation is by $T^* < Z^* \equiv \min(X^*, C_b^*)$, but in addition there is an independent variable C_r^* that censors the residual lifetime. The data are comprised of realizations from $(T, \min(Z, T + C_r), \Delta_{br})$ where Δ_{br} indicates whether lifetimes are uncensored, censored by C_b or censored by C_r. For goodness-of-fit purposes, the data can be considered as censored by C_r, after replacing F^* with F_{Z^*}.

Tests based on comparison of G and H are valid only when the support limits of F^* and

G^* satisfy certain requirements (Section 2.2). In several cases, they can still be performed when these requirements do not hold after certain adjustments are made. When $t_{\min} > 0$, one can subtract $t_{\min}$ before performing the tests; this was done for the Channing House data. The more common situation is when the right limit of the support of G^* is larger than that of F^* (as in the AIDS example). In that case, G^* is not identifiable and inference can be made only about $G^*(t|x_{\max}) = P(T^* \leq t|T^* \leq x_{\max})$. This distribution satisfies the conditions regarding the support. The case $t_{\max} < x_{\max}$ cannot be easily overcome. Although one can artificially truncate X at $t_{\max}$ and use the paired tests on the new truncated data, this is problematic under censoring because X is not always observed. In such cases, other tests, such as the CH2b, should be used.

The CLR performed very well in the simulation study, but the search for the maximum conditional likelihood can be a difficult task and in several situations, such as the AIDS example, the maximum is obtained in the boundary of the parameter space. One way to generate good initial values is via the method of moments. By (2), $\mathbb{E}\{T^k/\bar{F}^*(T)\} = \mathbb{E}(T^{*k})/P(T^* < X^*)$. This forms a set of equations that express the parameters as functions of empirical moments and is usually easy to solve. These equations depend on the unknown F^*, which can be replaced with its product-limit estimate to achieve a reasonable starting value. More specifically, let $m_k = n^{-1}\sum_{i=1}^n t_i^k/\bar{F}^*_{PL}(t_i)$, where the product-limit estimate $\bar{F}^*_{PL}$ is given in (5), and solve for θ the set of equations $m_k/m_0 = M_k(\theta)/M_0(\theta)$, where $M_k(\theta) = \mathbb{E}\{T^{*k}|T^* \leq t_0\}$.

Whether or not uniformity is assumed for the truncation distribution, inference for truncated data is commonly done under the assumption of quasi-independence, i.e., $f_{T,X}(t,x) \propto g^*(t)f^*(x)$ on $t \leq x$, where $f_{T,X}$ is the density of (T,X). This assumption is critical for estimation and should be examined before testing uniformity (e.g., Tsai, 1990, and Martin and Betensky, 2005).

An important problem that is not addressed here is that of testing $G^* \in \mathcal{G}_\theta$, i.e., the truncation belongs to a parametric family. In spite of its potential utility, only a few papers have studied this model. Wang (1989) showed how to estimate G^* and F^* under the semi-parametric model $\{G^* \in \mathcal{G}_\theta,\ F^*$ unrestricted$\}$, but dealt only with the uncensored case. Li et.al. (1997) derived confidence bands for the estimator of F^* in Wang's model. For the uncensored case, the likelihood ratio test based on the conditional likelihood is applicable. For the censored case, the tests of Li and Doss (1993), which are essentially chi-square tests that replace the observed proportions with consistent estimators, could be used. A referee's suggestion of extending the CLR to the censored case using profile likelihood arguments seems promising and further investigation in this direction is planned.

7 Supplementary Materials

Web Appendices, referenced in Sections 2 and 4 are available under the Paper Information link at the Biometrics website *http://www.tibs.org/biometrics*.

8 Acknowledgments

We are grateful to M. Zelen for a suggestion regarding the test of Section 3.1, and to the associate editor and two referees for helpful comments and suggestions that improve the paper considerably. This research was supported in part by CA075971 and CA105956.

References

[1] Andersen, E. B. (1970). Asymptotic properties of conditional maximum-likelihood estimators (Corr: 71V33 p167). *Journal of the Royal Statistical Society*, Series B: Method-

ological **32**, 283-301.

[2] Asgharian, M., M'Lan, C. E. and Wolfson, D. B. (2002). Length-biased sampling with right censoring: An unconditional approach. *Journal of the American Statistical Association* **97**, 201-209.

[3] Asgharian, M., Wolfson, D. b. and Zhang, X. (2006). Checking stationarity of the incidence rate using prevalent cohort survival data. *Statistics in Medicine* **25**, 1751-1767.

[4] Bilker, W. b. snd Wang M.-C. (1997). Bootstrapping left truncated and right censored data. *Communications in Statistics - Simulation and Computation* **26**, 141-171.

[5] Gangnon, R. E. and Kosorok, M. R. (2004). Sample-size formula for clustered survival data using weighted log-rank statistics. *Biometrika* **91**, 263-275.

[6] Gross, S. T. and Lai, T. L. (1996). Bootstrap methods for truncated and censored data. *Statistica Sinica* **6**, 509-530.

[7] Huang, Y. and Wang, M.-C (1995). Estimating the occurrence rate for prevalent survival data in competing risks models. *Journal of the American Statistical Association* **90**, 1406-1415.

[8] Hyde, J. (1977). Testing survival under right censoring and left truncation. *Biometrika* **64**, 225-230.

[9] Jung, S.-H. (1999). Rank tests for matched survival data. *Lifetime Data Analysis* **5**, 67-79.

[10] Kalbfleisch, J. D. and Lawless, J. F. (1989). Inference based on retrospective ascertainment: An analysis of the data on transfusion-related AIDS. *Journal of the American Statistical Association* **84**, 360-372.

[11] Keiding, N. (1991). Age-specific incidence and prevalence: A statistical perspective. *Journal of the Royal Statistical Society*, Series A: Statistics in Society **154**, 371-396.

[12] Laslett, G. M. (1982). The survival curve under monotone density constraints with applications to two-dimensional line segment processes, *Biometrika* **69**, 153-160.

[13] Li, G. and Doss, H. (1993). Generalized Pearson-Fisher chi-square goodness-of-fit tests, with applications to models with life history data. *The Annals of Statistics* **21**, 772-797.

[14] Li, G., Qin, J. and Tiwari, R. C. (1997). Semiparametric Likelihood Ratio-Based Inferences for Truncated Data. *Journal of the American Statistical Association* **92**, 236-245.

[15] Martin, E. C. and Betensky, R. A. (2005). Testing quasi-independence of failure and truncation times via conditional Kendall's tau. *Journal of the American Statistical Association* **100**, 484-492.

[16] Murray, S. (2001). Using weighted Kaplan-Meier statistics in nonparametric comparisons of paired censored survival outcomes. *Biometrics* **57**, 361-368.

[17] Qin, J. and Wang, M.-C. (2001). Semiparametric analysis of truncated data. *Lifetime Data Analysis* **7**, 225-242.

[18] Sun, L. and Zhu, L. (2000). A semiparametric model for truncated and censored data *Statistics & Probability Letters* **48**, 217-227.

[19] Tsai, W.-Y. (1990). Testing the assumption of independence of truncation time and failure time. *Biometrika* **77**, 169-177.

[20] de Uña-Álvarez, J. (2004a). Nonparametric estimation under length-biased sampling and type I censoring: A moment based approach. *Annals of the Institute of Statistical Mathematics* **56**, 667-681.

[21] de Uña-Álvarez, J. (2004b). Nelson-Aalen and product-limit estimation in selection bias models for censored populations. *Journal of Nonparametric Statistics* **16**, 761-777.

[22] de Uña-Álvarez, J. and Rodríguez-Casal, A. (2006). Nonparametric estimation from length-biased data under competing risks. *Computational Statstics & Data Analysis*, in press.

[23] Vardi, Y. (1982). Nonparametric estimation in renewal processes. *The Annals of Statistics* **10**, 772-785.

[24] Vardi, Y. (1989). Multiplicative censoring, renewal processes, deconvolution and decreasing density: Nonparametric estimation. *Biometrika* **76**, 751-761.

[25] Wang, M.-C. (1989). A semiparametric model for randomly truncated data. *Journal of the American Statistical Association* **84**, 742-748.

[26] Wang, M.-C. (1991). Nonparametric estimation from cross-sectional survival data. *Journal of the American Statistical Association* **86**, 130-143.

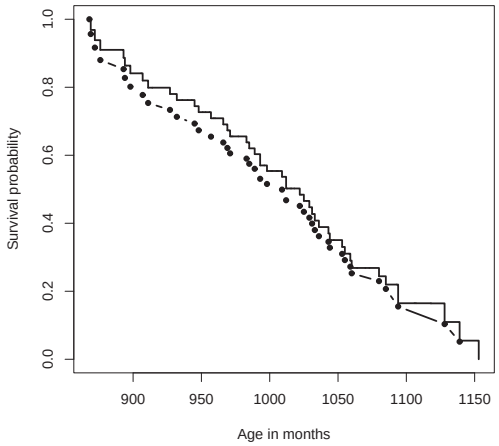
[27] Wang, M.-C., Jewell, N. P. and Tsai, W.-Y. (1986). Asymptotic properties of the product limit estimate under random truncation. *The Annals of Statistics* **14**, 1597-1605.

[28] Woodroofe, M. (1985). Estimating a distribution function with truncated data. *The Annals of Statistics* **13**, 163-177 (Corr: V15 p883).

[29] Woolson, R. F. and O'Gorman, T. W. (1992). A comparison of several tests for censored paired data. *Statistics in Medicine* **11**, 193-208.

Caption for Figure 1:

Figure 1: Comparison of survival estimates (left) and width of 95% pointwise confidence intervals (right) between the unconditional (solid) and conditional (circles) approaches applied to the Channing House data. The pointwise confidence intervals were calculated by the simple method of bootstrap.


Survival probability
0.0
0.2
0.4
0.6
0.8
1.0
900
950
1000
1050
1100
1150
Age in months

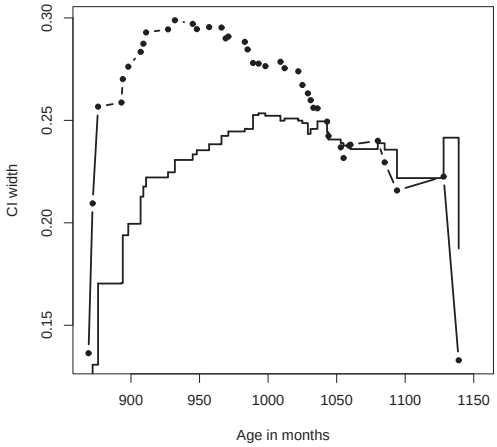

CI width
0.15
0.20
0.25
0.30
900
950
1000
1050
1100
1150
Age in months

Table 1: Estimated power of tests for uncensored data with level 0.05. Results based on 400 replications of different models for G^* and F^*=Beta(3,3).

		distribution of left truncation time (G^*)					
n	test	U(0,1)	Beta(1,1.5)	Beta(1.5,1)	Beta(2,4)	Beta(3,3)	Beta(4,2)
50	KS	.035	.110	.620	.108	.998	1
	WSR	.048	.165	.690	.073	.993	1
	PLR	.068	.200	.688	.135	.960	1
	PW	.058	.165	.728	.075	.995	1
	CLR	.058	.125	.690	.843	1	1
	CH2a	.115	.203	.642	.451	.977	.973
	CH2b	.044	.126	.237	.214	.877	.989
100	KS	.048	.198	.930	.228	1	1
	WSR	.043	.280	.958	.060	1	1
	PLR	.050	.338	.933	.170	1	1
	PW	.045	.275	.963	.058	1	1
	CLR	.048	.228	.958	.995	1	1
	CH2a	.093	.268	.855	.670	1	1
	CH2b	.043	.218	.676	.510	.997	1
200	KS	.048	.358	1	.635	1	1
	WSR	.035	.475	1	.055	1	1
	PLR	.050	.510	.998	.273	1	1
	PW	.035	.490	1	.055	1	1
	CLR	.038	.423	1	1	1	1
	CH2a	.078	.345	.995	.953	1	1
	CH2b	.041	.352	.966	.955	1	1

Table 2: Estimated power of tests for censored data with level 0.05. Based on 400 replications.

			probability and type of censoring					
			$P(\Delta = 0) = 0.75$		$P(\Delta = 0) = 0.50$		$P(\Delta = 0) = 0.25$	
model	n	test	Fixed	Random	Fixed	Random	Fixed	Random
	50	PW	.150	.033	.048	.040	.040	.045
		PLR	.135	.055	.030	.108	.120	.175
		CH2b	.178	.204	.230	.258	.305	.333
G^*=Beta(2,4)	100	PW	.230	.053	.045	.053	.070	.058
F^*=Beta(1,1.5)		PLR	.205	.075	.043	.160	.180	.270
		CH2b	.350	.485	.453	.533	.590	.630
	200	PW	.430	.065	.063	.055	.080	.078
		PLR	.388	.108	.058	.238	.338	.460
		CH2b	.853	.953	.920	.935	.958	.960
	50	PW	.558	.488	.713	.645	.723	.723
		PLR	.550	.498	.683	.595	.663	.645
		CH2b	.086	.146	.103	.138	.125	.118
G^*=Beta(1.5,1)	100	PW	.870	.793	.940	.905	.955	.958
F^*=Beta(2,4)		PLR	.865	.745	.908	.850	.928	.903
		CH2b	.315	.459	.420	.523	.515	.468
	200	PW	.988	.960	1	.998	1	1
		PLR	.990	.935	1	.983	1	.995
		CH2b	.800	.950	.905	.958	.940	.930

Table 3: P-values for testing the distribution of incubation time of elderly patients

Weibull model	WSR	KS	PLR	PW	CH2b
children: $\alpha = .341$, $\beta = 1.845$	$< .001$	$< .001$	$< .001$	$< .001$	$< .001$
adults: $\alpha = .015$, $\beta = 2.353$	.998	.909	.178	.787	.179