

A formal test for the stationarity of the incidence rate using data from a prevalent cohort study with follow-up

Vittorio Addona · David B. Wolfson

Received: 31 May 2005 / Accepted: 3 May 2006
© Springer Science+Business Media, LLC 2006

Abstract In a prevalent cohort study with follow-up, the incidence process is not directly observed since only the onset times of prevalent cases can be ascertained. Assessing the “stationarity” of the underlying incidence process can be important for at least three reasons, including an improvement in efficiency when estimating the survivor function. We propose, for the first time, a formal test for stationarity using data from a prevalent cohort study with follow-up. The test makes use of a characterization of stationarity, an extension of this characterization developed in this paper, and of a test for matched pairs of right censored data. We report the results from a power study assuming varying degrees of departure from the null hypothesis of stationarity. The test is also applied to data obtained as part of the Canadian Study of Health and Aging (CSHA) to verify whether the incidence rate of dementia amongst the elderly in Canada has remained constant.

Keywords Prevalent cohort · Right censoring · Left truncation · Incidence process · Forward and backward recurrence times

1 Introduction

Prevalent cohort survival data arise when prevalent cases are recruited and followed until failure or censoring. These data are typically left truncated and right censored. A prevalent sampling method alleviates the main practical

V. Addona
Department of Mathematics and Computer Science, Macalester College, St-Paul, Minnesota,
USA

D. B. Wolfson (✉)
Department of Mathematics and Statistics, McGill University, Burnside Hall,
805 Sherbrooke Street West, Montréal, Québec, Canada, H3A 2K6
e-mail: david@math.mcgill.ca

difficulties associated with the sampling of incident cases, but results in observations which are not a random sample from the target “incident” population. In particular, it is well known that cases obtained through a prevalent sampling scheme tend to have longer survival than cases obtained in an incident cohort study. Although the context of this paper is medical applications, prevalent cohort studies are not restricted to the medical field (see, for example, Lancaster 1990).

An important feature of any prevalent cohort study is that the incidence process of onset times is only partially observed, since only the onset times of prevalent cases may be ascertained. We refer to the fully observed incident process as the underlying incidence process to differentiate it from the process of partially observed onset times of the prevalent cases. When the underlying incidence process is a stationary Poisson process, we refer simply to “stationarity”. Whether or not stationarity holds in any particular situation is important for at least three reasons: (1) The well-known epidemiological relationship, “prevalence = incidence \times mean duration”, depends on the assumption that the underlying incidence process is stationary. Keiding (1991) makes explicit the assumptions needed to validate this relationship, although he clearly states that the ascertainment of stationarity is not the subject of his paper. (2) When estimating the survivor function from a prevalent cohort study with follow-up, there is a gain in efficiency if one can assume that the underlying incidence process is stationary (Asgharian et al. 2002; Wang 1991). (3) A goal of a study might be to ascertain whether the underlying disease incidence rate is constant or not. In this paper, we propose a formal test of stationarity using prevalent cohort survival data, where the prevalent cases are recruited over a short time period and followed up for a fixed study period until failure or censoring. Henceforth, a prevalent cohort study with follow-up is given this meaning.

To our knowledge, no formal method exists to verify stationarity of the underlying incidence process using data from a prevalent cohort study with follow-up. Wang (1991) proposed a method for estimating the left truncating distribution, which is Uniform under stationarity. Her method, however, provides no goodness-of-fit test to the Uniform distribution; Kolmogorov-type tests rely on the assumption that the estimator of the distribution function is derived from a set of independent and identically distributed (i.i.d.) observations, while Wang’s estimator is not obtained from such data.

Taking a different approach, Asgharian et al. (2006) showed that stationarity is, under mild assumptions, equivalent to the equality of the forward and backward recurrence time distributions. They suggested a graphical method, based on plots of Kaplan–Meier estimates of the backward and forward recurrence time distributions, to check the validity of stationarity. A formal test based on the distance between the Kaplan–Meier estimates of the backward and forward recurrence time distributions is not obvious since the backward and forward recurrence times are correlated. For this reason, Asgharian et al. did not present a formal test for stationarity. We exploit and extend the characterization of stationarity given by Asgharian et al. and show

how a test for matched pairs of right censored data (Wei 1980) may be adapted to give a test for stationarity.

The organization of this paper is as follows: In Sect. 2, we carefully describe the problem and present the notation that we use in this paper. In Sect. 3, we present a formal test for stationarity. In Sect. 4, we extend the results of Asgharian et al. by characterizing certain types of departures from stationarity. In Sect. 5, a power study is carried out to assess the performance of the test under various scenarios. In Sect. 6, the test is applied to data that were collected as part of the Canadian Study of Health and Aging (CSHA), to ascertain whether the incidence rate of dementia remained constant in Canada over the period 1971–1991.

2 Description of the problem and notation

Data from prevalent cohort studies in the medical field arise as follows: Let X_1, X_2, \dots, X_m be m i.i.d. positive random variables representing the lifetimes of individuals from onset of a certain condition (initiating event) to an end point of interest (terminating event). Let the X_i 's have survivor function $S(x) := P(X_i > x)$, with probability density function $f(x)$. Let $\tau_1, \tau_2, \dots, \tau_m$ be the m calendar times of onset corresponding to X_1, X_2, \dots, X_m and let τ^* be the calendar time of recruitment into the study. Individual i is observed in the study only if $X_i \geq \tau^* - \tau_i$, and we therefore say that X_i is left truncated with left truncation time $T_i = \tau^* - \tau_i$. Since the onset times are random, the truncation times are random variables, with distribution function denoted by G , and density g .

Let Y_1, Y_2, \dots, Y_n be the *observed* left truncated lifetimes, with $n \leq m$, recalling that not all subjects will survive long enough to be observed in the study. That is, $P(Y_i > x) = P(X_i > x | X_i > T_i)$. We write $Y_i = Y_i^{\text{bwd}} + Y_i^{\text{fwd}}$, where Y_i^{bwd} is the time from onset to recruitment into the study or the “backward recurrence time”, and Y_i^{fwd} is the time from recruitment to failure, or the “forward recurrence time”. Let $Y_i^{\text{bwd}} \sim f_{\text{bwd}}$ and $Y_i^{\text{fwd}} \sim f_{\text{fwd}}$, where f_{bwd} and f_{fwd} are, respectively, the backward and forward recurrence time densities. Since Y_i^{bwd} and Y_i^{fwd} are obviously negatively correlated, this must be taken into account when constructing a test for stationarity.

In practice, a cohort of subjects who agree to participate in a study will not be assembled at a single instance in time but over a short time period. Further, as is commonly assumed in all prevalent cohort studies, the prevalent cases are, in fact, identified over a time interval, $[\tau^*, T]$, say. The instances of case ascertainment, $\tau_1^*, \tau_2^*, \tau_3^*, \dots, \tau_n^*$, therefore, fall in the interval $[\tau^*, T]$. We assume, as do Keiding (1991) and Wang (1991), and most others, that T is close enough to τ^* to ensure that all n subjects identified as prevalent cases were, in fact, prevalent cases at time τ^* , and that no case prevalent at τ^* , is lost before time T . With these constraints one need place no further restriction on the τ_i^* 's. The time instance, τ^* , is termed *prevalence day*. Of course, in practice, these assumptions will be only approximately true, in that a small number of

incident cases may occur over $[\tau^*, T]$, and some of the prevalent cases at τ^* , may be lost before T . All the results are approximately valid under these assumptions. Allowing for incident cases and loss of prevalent cases is a much harder problem which has, as far as we can ascertain, not been addressed in the literature.

2.1 The censoring mechanism

In a prevalent cohort study, suppose that individual i has censoring time $C_i^* = Y_i^{\text{bwd}} + C_i$, where C_i , which we call the “residual censoring time”, is the time from recruitment until the individual is censored. We assume that $P(C_i^* > T_i) = 1$ (see, for example, Wang 1991) and we thus observe only $\min(C_i^*, Y_i)$. Often, however, the backward recurrence times are fully observed, and we shall suppose here that this is the case, and that the observed data are:

$$(Y_i^{\text{bwd}}, Y_i^{\text{obs}}, \epsilon_i) \quad i = 1, 2, \dots, n, \quad (1)$$

where $Y_i^{\text{obs}} = \min(Y_i^{\text{fwd}}, C_i)$ and $\epsilon_i = \mathbf{1}[Y_i^{\text{fwd}} \leq C_i]$ indicates whether subject i has been followed until failure. Since C_i^* and Y_i have Y_i^{bwd} in common, we thus have *informative censoring* (Vardi 1989). It is still reasonable to assume, however, that the residual censoring time, C_i , is independent of both Y_i^{fwd} and Y_i^{bwd} , noting that the assumed independence between C_i and Y_i^{fwd} corresponds to the usual random censoring assumption.

3 A formal test for stationarity

When the underlying incidence process is a Poisson process with constant intensity $\lambda(t) \equiv \lambda$, we shall, for brevity, simply refer to “stationarity”. The truncation time distribution, under stationarity, is Uniform, conditional on the number of incident times in $(0, \tau^*)$.

In general, if the initiation times occur over an interval $(0, \tau^*)$ according to an arbitrary Poisson process with intensity function $\lambda(t)$, then an onset at calendar time $(\tau^* - x)$ for some $x \in (0, \tau^*)$ implies a truncation time of x , and we have that,

$$g(x) = \frac{\lambda(\tau^* - x)\mathbf{1}[0 < x < \tau^*]}{\Lambda(\tau^*)} \quad (2)$$

where $\Lambda(u) = \int_0^u \lambda(v)dv$, and g is the truncation time density (see, for example, Ross 2003).

The problem under consideration here is: knowing only the backward recurrence times, and the (possibly right censored) forward recurrence times, test

$$\begin{aligned} H_0 : & \quad \lambda(t) = \lambda \quad \text{for all } t \leq \tau^* \\ \text{vs. } H_a : & \quad \lambda(t) \neq \lambda \quad \text{for some } t \leq \tau^*. \end{aligned} \quad (3)$$

We shall, ultimately, narrow the class of alternative hypotheses. The key to our simple test of (3) is Theorem 1 of Asgharian et al. stated below:

Theorem 1 (Asgharian et al. 2006) *Let X and T be, respectively, the true lifetime and left truncation time with corresponding density functions $f(x)$ and $g(t)$ (possibly improper), having the same support. Suppose X and T are independent. Let $\int_0^\infty g(t)S(t) dt < \infty$ and one of the following two conditions hold:*

- (A) *The truncation density, g , has an absolute maximum with unbounded support.*
- (B) *The truncation density, g , is continuous with bounded support. Then the truncating density, g , is constant if and only if $f_{\text{bwd}} \equiv f_{\text{fwd}}$.*

Theorem 1 establishes that the hypothesis (3) of stationarity holds if and only if

$$H_0 : F_{\text{bwd}}(x) = F_{\text{fwd}}(x) \quad \text{for all } x \geq 0. \quad (4)$$

Although an inspection of the Kaplan–Meier estimates of $F_{\text{bwd}}(x)$ and $F_{\text{fwd}}(x)$, plotted together, provides a quick indication of whether or not there may be non-stationarity, Asgharian et al. provide no formal test of (4).

Now, (4) may be described as the hypothesis that the population ensemble of backward recurrence times has the same distribution as the distribution of the ensemble of forward recurrence times. However, a two-sample test that allows for censoring, such as a log rank test (Mantel 1966), is precluded since although the between pair backward and forward recurrence times are independent, within pairs they are not; in a sense we have matched pairs, with one member of the pair (the forward recurrence time) possibly randomly right censored.

Wei (1980) proposed a matched pairs test using the same scoring function as Gehan (1965), who constructed a modified two-sample Wilcoxon rank sum test that allows for censoring. Wei's test makes use of both within pair and between pair comparisons, and we adapt it to test for stationarity.

Suppose that $(X_1^0, Y_1^0), \dots, (X_n^0, Y_n^0)$ are i.i.d. random vectors with common distribution function $H^0(s, t)$. Let $F^0(s)$ and $G^0(t)$ be the marginal distribution functions of the X_i^0 's and Y_i^0 's, respectively. Assume that H^0 , and hence F^0 , and G^0 , are continuous distribution functions. Suppose that the null hypothesis is,

$$H_0 : H^0(s, t) = H^0(t, s) \quad \text{for all } (s, t) \in \mathbb{R}^2. \quad (5)$$

Alternative hypotheses which may be of interest are those indicating that the X_i^0 's tend to be larger than the Y_j^0 's. It is important to note that while bivariate symmetry of a joint distribution function implies equality of the marginal distributions, in general, the converse is not true. Thus, the

hypothesis in (5) is not, in general, equivalent to the hypothesis in (4). We show later, however, that in the particular case where the random variables in each pair correspond to forward and backward recurrence times, the two hypotheses are, indeed, equivalent, opening the way to the use of Wei's test of (4).

In Wei's paper (1980), the observed data are $(X_1^*, Y_1^*), \dots, (X_n^*, Y_n^*)$, where $X_i^* = \min(X_i^0, U_i)$ and $Y_i^* = \min(Y_i^0, V_i)$, along with $\delta_i = \mathbf{1}[X_i^* = X_i^0]$ and $\epsilon_i = \mathbf{1}[Y_i^* = Y_i^0]$, for $i = 1, \dots, n$. The random variables U_i and V_i are the censoring variables of the X_i^0 's and Y_i^0 's, respectively, with $U_i \sim J(s)$ and $V_i \sim K(t)$. It is assumed that the pairs (X_i^0, Y_i^0) and (U_i, V_i) are independent for all $i = 1, \dots, n$, and that U_i is independent of V_j for $i \neq j$.

Define $\tilde{F}(s) = P(X_i^* \leq s, \delta_i = 1)$ and $\tilde{G}(t) = P(Y_i^* \leq t, \epsilon_i = 1)$. Let the scoring function $\Psi(x_i^*, y_j^*, \delta_i, \epsilon_j) = \mathbf{1}[x_i^* > y_j^*, \epsilon_j = 1] - \mathbf{1}[x_i^* < y_j^*, \delta_i = 1]$. Also, let $W_n = \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n [\Psi(X_i^*, Y_j^*, \delta_i, \epsilon_j) - p]$, where $p = E(G(X)) - E(\tilde{F}(Y))$.

Theorem 2 (Wei 1980) *As $n \rightarrow \infty$, $\sqrt{n}W_n$ converges in distribution to a Normal random variable with mean 0 and variance σ^2 .*

Wei suggests two possible consistent estimators of σ_0^2 , the null value of σ^2 . We denote the preferred estimator by $\hat{\sigma}_0^2$. Wei's asymptotically nonparametric test compares observed values of $\sqrt{n}W_n/\hat{\sigma}_0$ to a standard Normal distribution.

We may use Wei's procedure to test for stationarity, identifying (X_i^0, Y_i^0) with $(Y_i^{\text{bwd}}, Y_i^{\text{fwd}})$ and (U_i, V_i) with $(+\infty, C_i^*)$, for $i = 1, 2, \dots, n$, provided we can establish the equivalence of (4) and (5). The null hypothesis in (5) always implies the null hypothesis in (4), and when paired data arise as backward and forward recurrence times from a prevalent cohort study, Lemma 1 states that the converse is also true.

Lemma 1 $Y^{\text{bwd}} = Y^{\text{fwd}}$ in distribution $\iff f_{Y^{\text{bwd}}, Y^{\text{fwd}}}(x, y) = f_{Y^{\text{bwd}}, Y^{\text{fwd}}}(y, x)$.

Proof See Appendix.

In order to address power, we now turn to more specific hypotheses than those specified earlier under the alternative to (3).

4 Characterizations of departures from stationarity

We extend Theorem 1 of Asgharian et al. by characterizing a certain class of alternatives to stationarity in terms of the backward and forward recurrence time distributions.

The most common alternative to stationarity is that the intensity of the incidence process is non-decreasing, and non-constant (or alternatively, non-increasing and non-constant) as a function of time. A two-sided alternative is less common. For definiteness, we focus on,

$$H_0 : \lambda(t) \text{ is constant vs. } H_1 : \lambda(t) \text{ is non-decreasing and non-constant,} \quad (6)$$

noting that analogous results hold under a non-increasing alternative. Under the alternative hypothesis in (6), from Asgharian et al., we have that for $x > 0$,

$$f_{\text{bwd}}(x) = \frac{g(x)S(x)}{\int_0^\infty g(u)S(u)du} \quad \text{and} \quad (7)$$

$$f_{\text{fwd}}(x) = \frac{\int_0^\infty g(u)f(u+x)du}{\int_0^\infty g(u)S(u)du}. \quad (8)$$

We write $X >^{SL} Y$ if X is stochastically larger than Y , that is, if $F_X(x) \leq F_Y(x) \forall x \in \mathbb{R}$ with strict inequality for at least one x .

Theorem 3 *Suppose that $\lambda(t)$ is non-decreasing and non-constant. Then $Y^{\text{fwd}} >^{SL} Y^{\text{bwd}}$.*

Proof See Appendix.

In the case of a non-decreasing intensity, the equivalence of the backward and forward recurrence time distributions is upset by observing more individuals closer to recruitment than expected under stationarity. Analogously to Theorem 3, a similar argument shows that if we suppose $\lambda(t)$ is non-increasing and non-constant (i.e. $g(t)$ is non-decreasing and non-constant) then $Y^{\text{bwd}} >^{SL} Y^{\text{fwd}}$.

4.1 Restricting the class of intensities

The converse to Theorem 3 does not hold, and in the Appendix we provide a counterexample where $\lambda(t)$ is strictly decreasing over a portion of its support, but where $Y^{\text{fwd}} >^{SL} Y^{\text{bwd}}$. Hence, there is no simple characterization of a non-decreasing, non-constant intensity, $\lambda(t)$, in terms of $F_{\text{bwd}}(x)$ and $F_{\text{fwd}}(x)$. However, if we restrict the class of admissible intensities, then the following theorem provides a characterization that, for practical purposes, covers a large number of cases.

Theorem 4 *Suppose that $\lambda(t)$ belongs to the class of monotone intensity functions (i.e. $\lambda(t)$ is non-decreasing or non-increasing). Then $Y^{\text{fwd}} >^{SL} Y^{\text{bwd}} \Leftrightarrow \lambda(t)$ is non-decreasing, and non-constant.*

Proof See Appendix.

Clearly, Theorem 4 also provides the characterization, $Y^{\text{bwd}} >^{SL} Y^{\text{fwd}} \Leftrightarrow \lambda(t)$ is non-increasing, and non-constant, if we again restrict $\lambda(t)$ to the class of monotone intensity functions. In view of Theorems 3 and 4, the hypotheses in (6) are equivalent to,

$$H_0^* : Y^{\text{fwd}} = Y^{\text{bwd}} \text{ in distribution vs. } H_1^* : Y^{\text{fwd}} >^{SL} Y^{\text{bwd}},$$

within the restricted class of monotone intensity functions.

Wei (1980) provided results from a small power study which showed that his test outperformed the sign test when the alternative hypothesis was stochastic ordering of the random variables. In Sect. 5, we provide a more extensive examination of the power of Wei's test for detecting departure from stationarity in favor of the types of alternatives we have just characterized.

5 Detecting departure from stationarity: a power study

We examine the power of Wei's test to detect departure from stationarity assuming varying degrees of non-stationarity. To carry out this assessment, we generated onsets, over a pre-specified interval $(0, \tau^*)$, according to a parametric intensity that is strictly monotone. This corresponds to a situation where the forward and backward recurrence time distributions are stochastically ordered. For each onset, we generated a corresponding survival time. The survival times which extended beyond τ^* , originating from their respective onset times, constituted the observed prevalent cohort data. We constructed an α -level critical region using Wei's test for equality of the forward and backward recurrence time distributions.

5.1 Details of the simulations

The following issues need to be considered: the specification of the underlying survival density, $f(x)$, the residual censoring time distribution, the form of the onset intensity, $\lambda(t)$, and the specific parameters used in $\lambda(t)$, the support of the truncation distribution $(0, \tau^*)$, where τ^* represents the calendar time of recruitment, the sample size, n , and the number of replications of the entire procedure, M .

For the underlying survival distribution, three different scenarios were simulated: **(A)** Weibull($\gamma = 2, \beta = 10$), **(B)** Weibull($\gamma = 0.75, \beta = 1.5$), and **(C)** Lognormal($\mu = 1.75, \sigma = 0.4$), where the parametrization of the Weibull(γ, β) distribution is:

$$f(x) = \frac{\gamma}{\beta} x^{\gamma-1} e^{-\frac{x^\gamma}{\beta}} \mathbf{1}[x > 0],$$

and the parametrization of the Lognormal(μ, σ^2) distribution is:

$$f(x) = \frac{e^{-\frac{(\log x - \mu)^2}{2\sigma^2}}}{\sqrt{2\pi\sigma x}} \mathbf{1}[x > 0].$$

The (γ, β) pair in **(A)** represents a Weibull distribution with an increasing hazard function, whereas the (γ, β) pair in **(B)** represents a Weibull distribution with a decreasing hazard function.

For each of **(A)**, **(B)**, and **(C)**, we examined the power of Wei's test using sample sizes of $n \approx 200$, $n \approx 500$, and $n \approx 1000$.

For each survival distribution, the residual censoring time distribution was chosen to be Exponential with mean, β , selected so that approximately 25% of the forward recurrence times were censored. Specifically, we employed: **(A)** Exponential($\beta = 6$), **(B)** Exponential($\beta = 8$), and **(C)** Exponential($\beta = 12$), where the parametrization of the Exponential(β) distribution is:

$$f(x) = \frac{1}{\beta} e^{-\frac{x}{\beta}} \mathbf{1}[x > 0].$$

The onset process was chosen to be a Poisson process with log-linear intensity function,

$$\lambda(t) = \lambda e^{\alpha_1 t} \quad (9)$$

The intensity function in (9) is strictly increasing when α_1 is positive, and strictly decreasing when α_1 is negative. Stationarity holds if and only if $\alpha_1 = 0$. For the cases representing non-homogeneous Poisson processes, we simulated the onsets over an interval, $(0, \tau^*)$, corresponding to the support of the truncation distribution, by following Lewis and Shedler (1976). For simulation of the stationary onsets, Uniform($0, \tau^*$) random variables were generated and ordered. The support of the truncation distribution was chosen to be sufficiently large to ensure that the backward recurrence time distribution was not "clipped" beyond a certain value: clearly, if onsets are generated on $(0, \tau^*)$ then no backward recurrence time can be observed larger than τ^* . If τ^* were chosen too small, we would have artificially reduced the support of the backward recurrence time distribution. Of course, since the survival distributions in **(A)**, **(B)**, and **(C)** have their support on $(0, \infty)$, we can never select τ^* large enough to guarantee that there is absolutely no "clipping" of the backward recurrence time distribution. We can, however, choose τ^* large enough to ensure that this "clipping" is negligible. The values of τ^* that were chosen for the present power study were: **(A)** $\tau^* = 6$, **(B)** $\tau^* = 12$, and **(C)** $\tau^* = 16$.

The parameter that controls the "degree" of non-stationarity is α_1 , and the larger the magnitude of α_1 , the greater the departure from stationarity. The α_1 values that were chosen for the three survival distributions are as follows,

$$\mathbf{(A)} : (1)0.05 (2)0.10 (3)0.15 (4) - 0.05 (5) - 0.10 (6) - 0.15$$

$$\mathbf{(B)} : (1)0.05 (2)0.10 (3)0.15 (4) - 0.05 (5) - 0.10 (6) - 0.15$$

$$\mathbf{(C)} : (1)0.03 (2)0.06 (3)0.09 (4) - 0.03 (5) - 0.06 (6) - 0.09$$

These α_1 values cover a range of circumstances from "mild" non-stationarity, to "moderate" non-stationarity, to "severe" non-stationarity. The specific values of α_1 were chosen by considering the expected number of onsets that

would occur over $(0, \tau^*)$ under the non-stationary onset intensity relative to the same number under stationarity. The values of λ were chosen to attain the desired sample size (either $n \approx 200$, $n \approx 500$, or $n \approx 1000$), and are not reported. For each of the three survival distributions, we simulated stationary onsets by generating a sufficient number of Uniform random variables over the interval $(0, \tau^*)$ in order to attain the desired sample size.

For $n \approx 200$, simulations were only performed for the “severe” departures from stationarity, along with the null situation. For $n \approx 500$ and $n \approx 1000$, all seven situations (six non-stationary intensities and a constant intensity representing stationarity) were simulated. This led to 51 different simulation scenarios in total. For each of these 51 scenarios, we performed $M = 500$ replicates, and we recorded the number of rejections at the $\alpha = 0.05$ level. We used a two-sided rejection region in assessing the properties of this test, although in practice, a researcher would typically have in mind, a priori, exactly one of the alternatives (either a non-decreasing or a non-increasing onset intensity).

5.2 Results of the simulation study

The percentages of rejections of the null hypothesis of stationarity are presented in Tables 1, 2, and 3 for **(A)**, **(B)**, and **(C)**, respectively. When $n \approx 200$, the power for the most severely non-stationary onset intensities ranged from 32.0% to 47.0% in the increasing intensity cases and from 35.6% to 66.4% in the decreasing intensity cases. By comparison, for $n \approx 500$, the power for the most severely non-stationary onset intensities ranged from 80.2% to 90.6% in the increasing intensity cases and from 82.2% to 97.0% in the decreasing intensity cases. The effect of doubling the sample size from 500 to 1000 was substantial in all three cases in that it considerably increased the power for the “moderate” and “mild” departures from stationarity. When $n \approx 1000$, the power for the “moderate” departure from non-stationarity ranged from 77.8% to 92.4% when the onset intensity was increasing, and from 70.0% to 94.2% when the onset intensity was decreasing. It is important to observe that under stationarity, the rejection rate was much lower than the nominal 5% value.

5.3 Discussion of the results of the power study

This power study demonstrated some general properties of Wei’s test. First, the power is not adequate for sample sizes of $n \approx 200$. Severe departures from stationarity will be detected for sample sizes close to $n \approx 500$. We feel that a

Table 1 Percentages of rejections for Weibull($\gamma = 2, \beta = 10$)

α_1	0.05	0.10	0.15	- 0.05	-0.10	-0.15	0 (Stationarity)
$n \approx 200$	–	–	33.2	–	–	35.6	1.6
$n \approx 500$	9.0	42.2	80.2	5.6	37.2	82.2	0.8
$n \approx 1000$	24.0	77.8	99.0	12.6	70.0	99.2	1.0

Table 2 Percentages of rejections for Weibull($\gamma = 0.75, \beta = 1.5$)

α_1	0.05	0.10	0.15	-0.05	-0.10	-0.15	0 (Stationarity)
$n \approx 200$	–	–	32.0	–	–	53.8	0.2
$n \approx 500$	9.8	49.6	89.4	2.8	43.8	95.8	0.4
$n \approx 1000$	32.4	90.4	100.0	6.8	85.0	100.0	0.6

Table 3 Percentages of rejections for Lognormal($\mu = 1.75, \sigma = 0.4$)

α_1	0.03	0.06	0.09	-0.03	-0.06	-0.09	0 (Stationarity)
$n \approx 200$	–	–	47.0	–	–	66.4	0.6
$n \approx 500$	15.6	65.0	90.6	15.4	63.2	97.0	2.0
$n \approx 1000$	30.6	92.4	99.8	31.8	94.2	100.0	1.6

sample size of $n \approx 500$ is very attainable for most prevalent cohort studies. Also, even for relatively large sample sizes, “mild” departures from stationarity will not be detected.

Good power was observed for detecting “moderate” departures from stationarity only when $n \approx 1000$. Reasons for this result include the possibility that using a high censoring proportion ($\approx 25\%$) compromised the power of the test when $n \approx 500$. The power was also slightly reduced by the use of a two-sided rejection region. Another factor which most certainly affected the power of Wei’s test is that its rejection rate under the null hypothesis did not reach the nominal 5% value, but was much closer to 1%. Cheng (1984) addresses the reason for this occurrence.

He points out that Wei’s $\hat{\sigma}_0^2$ is not a good estimator of σ_0^2 . Cheng provides an alternative estimator of σ_0^2 which, as he demonstrates through examples, performs better than $\hat{\sigma}_0^2$. Unfortunately, Cheng’s estimator is only valid under the assumption of identical within pair censoring distributions (i.e. $J(s) \equiv K(s)$). Thus, we cannot utilize Cheng’s estimator for our purposes since the backward recurrence times, by definition, cannot be censored whereas the forward recurrence times may be censored. Cheng showed that using $\hat{\sigma}_0^2$ yields percentiles of the test statistic ($\sqrt{n}W_n/\hat{\sigma}_0$) which are not consistent with those of a Normal(0,1) random variable. The results reported in Cheng’s paper are in accordance with those obtained here. That is, under H_0 , the test statistic ($\sqrt{n}W_n/\hat{\sigma}_0$) will fall outside of the interval ($-z_{0.025}, z_{0.025}$) (where $z_\alpha = 100(1 - \alpha)$ percentile of a standard Normal distribution) with a probability of roughly 0.01. This conservative feature of Wei’s test surely affected its power in this simulation study.

6 Analysis of CSHA data

As an application of the formal test for stationarity we assess the stationarity of the incidence of dementia using data collected as part of the Canadian

Study of Health and Aging (CSHA). In 1991, over 10,000 elderly Canadians (65 years or older) living at home or in institutions were screened for dementia (CSHA working group 1994). This phase of the study was known as CSHA-1. At the time of CSHA-1, 821 subjects were identified as having either possible Alzheimer's disease, probable Alzheimer's disease, or vascular dementia. Henceforth, the definition of dementia is restricted to exactly one of these three conditions since they constitute the vast majority of dementias. The approximate dates of onset were derived in a hierarchical fashion from the answers to three questions (Wolfson et al. 2001). In 1996, the second phase of the study, CSHA-2, was completed. CSHA-2 included the ascertainment of the date of death or right censoring for those cases identified at CSHA-1. Moreover, those subjects who were deemed not to have dementia at CSHA-1 were re-evaluated for dementia at CSHA-2.

Amongst the objectives of the CSHA was the estimation of the incidence rate of dementia in Canada in an elderly cohort. Assuming that this incidence rate had remained constant, it was estimated using the incident cases observed between CSHA-1 and CSHA-2 (CSHA working group 2000). Since the estimation procedure used for the incidence rate was predicated on the assumption that it was constant it is appropriate to formally examine this assumption, using *only* the prevalent cohort data obtained at CSHA-1, and their follow-up failure/censoring times until CSHA-2.

Using data obtained on the 821 subjects identified at CSHA-1, that is, the approximate dates of onset of dementia and the dates of death or censoring, we computed $\sqrt{n}W_n/\hat{\sigma}_0$. The observed value of this test statistic was approximately 0.98, which is to be compared to a standard Normal distribution, yielding a two-sided p-value of roughly 0.33. This is consistent with stationarity. This conclusion agrees with the informal checks of stationarity given in Asgharian et al. (2006) and Asgharian et al. (2002). Asgharian et al. (2006) address the issue of stationarity of the incidence of dementia in greater detail.

7 Concluding remarks

Although the test statistic calculated from the CSHA data was not statistically significant, its positive value suggests slightly longer backward recurrence times in comparison to the forward recurrence times. This result agrees with the observation made in Asgharian et al. (2006) that the Kaplan–Meier estimate of the backward recurrence time survivor function lies slightly above that of the forward recurrence time survivor function for times close to the origin. Asgharian et al. suggest the screening tests may not have been sensitive enough to detect dementia when onset occurred close to recruitment at CSHA-1. We add another potential explanation for this result.

As noted in Wolfson et al. (2001), stratified sampling was carried out with over 9,000 subjects living in the community and over 1,200 living in institutions. This represented an oversampling of subjects living in institutions.

Moreover, it is likely that the cases of dementia for those living in institutions were more advanced than for those living at home. This provides another reason for the observation of an excess of longer backward recurrence times near the origin.

Finally, we return to the observation that Wei’s $\hat{\sigma}_0^2$ is not a good estimator of σ_0^2 . Obtaining a better estimator of σ_0^2 , which is applicable when the data are backward and forward recurrence times arising from a prevalent cohort study with follow-up, remains a topic for further investigation.

Acknowledgments This research was supported in part by the Natural Sciences and Engineering Research Council of Canada. The data reported in this article were collected as part of the Canadian Study of Health and Aging. The core study was funded by the Seniors’ Independence Research Program, through the National Health Research and Development Program (NHRDP) of Health Canada Project 6606-3954-MC(S). Additional funding was provided by Pfizer Canada Incorporated through the Medical Research Council/Pharmaceutical Manufacturers Association of Canada Health Activity Program, NHRDP Project 6603-1417-302(R), Bayer Incorporated, and the British Columbia Health Research Foundation Projects 38 (93–2) and 34 (96–1). The study was coordinated through the University of Ottawa and the Division of Aging and Seniors, Health Canada. We thank the referees and Associate Editor for their suggestions and comments which have enhanced our paper.

Appendix

Proof of Lemma 1 We assume that the support of the underlying lifetime distribution is infinite. The result continues to hold when the support of the underlying lifetime distribution is finite, but we omit the proof since it is similar to that of the infinite case. It is not difficult to find the joint distribution of the backward and forward recurrence times. If Y represent the observed left truncated lifetime, then

$$f_{Y^{bwd}|Y}(x|l) = \frac{g(x)\mathbf{1}[0 \leq x \leq l]}{G(l)}$$

where g and G represent the truncation time density and distribution function, respectively. Also,

$$f_Y(l) = \frac{f(l)G(l)\mathbf{1}[l \geq 0]}{\int_0^\infty g(u)S(u)du}$$

where f and S represent the density and survivor function, respectively, of the underlying lifetime. Therefore,

$$\begin{aligned} f_{Y^{bwd}, Y}(x, l) &= f_{Y^{bwd}|Y}(x|l)f_Y(l) \\ &= \frac{g(x)f(l)\mathbf{1}[0 \leq x \leq l]\mathbf{1}[l \geq 0]}{\int_0^\infty g(u)S(u)du} \end{aligned}$$

Now, performing the bivariate transformation $Y^{fwd} = Y - Y^{bwd}$, $Y^{bwd} \equiv Y^{bwd}$, we have that:

$$f_{Y^{\text{bwd}}, Y^{\text{fwd}}}(x, y) = f_{Y^{\text{bwd}}, Y}(x, x + y) \\ = \frac{g(x)f(x+y)\mathbf{1}[0 \leq x \leq \infty]\mathbf{1}[0 \leq y \leq \infty]}{\int_0^\infty g(u)S(u)du} \quad (10)$$

$$= \frac{g(x)f(x+y)\mathbf{1}[0 \leq x \leq \infty]\mathbf{1}[0 \leq y \leq \infty]}{\int_0^\infty \int_0^\infty g(u)f(u+v)dvdu} \quad (11)$$

Under stationarity, $g(x)$ is constant so that from (11) the joint density of the backward and forward recurrence times depends only on their sum, $x + y$. Hence, it is clear that stationarity implies that $f_{Y^{\text{bwd}}, Y^{\text{fwd}}}(x, y) = f_{Y^{\text{bwd}}, Y^{\text{fwd}}}(y, x)$, that is, that the joint density of the backward and forward recurrence times is symmetric in its arguments. However, stationarity is equivalent to the hypothesis in (4) by the characterization of Asgharian et al. (2006), and so we have that (4) \Rightarrow (5). \square

Proof of Theorem 3 We assume that the support of the underlying lifetime distribution is infinite. The result continues to hold when the support of the underlying lifetime distribution is finite, but we omit the proof since it is similar to that of the infinite case.

From (7) and (8), we have that,

$$F_{\text{bwd}}(x) = \frac{\int_0^x g(t)S(t)dt}{\int_0^\infty g(u)S(u)du} \quad \text{and} \quad (12)$$

$$F_{\text{fwd}}(x) = \frac{\int_0^x \int_0^\infty g(u)f(u+t)du dt}{\int_0^\infty g(u)S(u)du} \quad (13)$$

However, (12) and (13) can be expressed differently, and in a more convenient manner, for the purposes of this proof. Let X represent the true lifetime, and T the left truncation time:

$$F_{\text{bwd}}(x) = P(Y^{\text{bwd}} \leq x | X \geq T) \\ = \frac{\int_0^\infty \int_0^a P(Y^{\text{bwd}} \leq x, X \geq T | X = a, T = t)g(t|X = a)f(a)dt da}{\int_0^\infty P(X \geq T | T = u)g(u)du} \\ = \frac{\int_0^\infty \int_0^a P(Y^{\text{bwd}} \leq x, X \geq T | X = a, T = t)g(t)f(a)dt da}{\int_0^\infty P(X \geq u | T = u)g(u)du} \\ = \frac{\int_0^\infty \int_0^a P(Y^{\text{bwd}} \leq x, X \geq T | X = a, T = t)g(t)f(a)dt da}{\int_0^\infty S(u)g(u)du}$$

Also, $P(Y^{\text{bwd}} \leq x, X \geq T | X = a, T = t) = \mathbf{1}[t \leq x]$. Therefore, we have,

$$F_{\text{bwd}}(x) = \frac{\int_0^x \int_0^a g(t)f(a)dt da + \int_x^\infty \int_0^x g(t)f(a)dt da}{\int_0^\infty S(u)g(u)du} \quad (14)$$

Similarly, for $F_{\text{fwd}}(x)$, we have that,

$$\begin{aligned}
 F_{\text{fwd}}(x) &= P(Y^{\text{fwd}} \leq x | X \geq T) \\
 &= \frac{\int_0^\infty \int_0^a P(Y^{\text{fwd}} \leq x, X \geq T | X = a, T = t) g(t) f(a) dt da}{\int_0^\infty S(u) g(u) du}
 \end{aligned}$$

Since $P(Y^{\text{fwd}} \leq x, X \geq T | X = a, T = t) = \mathbf{1}[t \geq a - x]$ we have,

$$F_{\text{fwd}}(x) = \frac{\int_0^x \int_0^a g(t) f(a) dt da + \int_x^\infty \int_{a-x}^a g(t) f(a) dt da}{\int_0^\infty S(u) g(u) du} \tag{15}$$

Therefore,

$$\begin{aligned}
 F_{\text{bwd}}(x) - F_{\text{fwd}}(x) &\propto \int_x^\infty \int_0^x g(t) f(a) dt da - \int_x^\infty \int_{a-x}^a g(t) f(a) dt da \\
 &= \int_x^\infty \left[\int_0^x g(t) dt - \int_{a-x}^a g(t) dt \right] f(a) da \\
 &= \int_x^\infty \int_0^x \left[g(t) - g(t + (a - x)) \right] f(a) dt da \\
 &\geq 0 \quad \forall \quad x \geq 0,
 \end{aligned} \tag{16}$$

since when $\lambda(t)$ is non-decreasing, $g(t)$ is non-increasing (as seen from Eq. (2)) and from the fact that $a > x$. This implies that $Y^{\text{fwd}} >^{SL} Y^{\text{bwd}}$, as required. \square

Counterexample: First, if any two density functions, $f_1(x)$ and $f_2(x)$, having common support cross each other only once over their support, then the two distributions must be stochastically ordered. We provide a more precise statement of this result in the following lemma.

Lemma 2 *Let $f_1(x)$ and $f_2(x)$ be two density functions sharing a common support and satisfying the following property: there exists a unique x^* such that $f_1(x) > f_2(x)$ for $x \in (0, x^*)$, $f_1(x) \leq f_2(x)$ for $x > x^*$, and $f_1(x^*) = f_2(x^*)$. Let $X \sim f_1$ and $Y \sim f_2$. Then $X >^{SL} Y$.*

Proof of Lemma 2 Let $F_1(x) = \int_0^x f_1(u) du$, $S_1(x) = \int_x^\infty f_1(u) du$, $F_2(x) = \int_0^x f_2(u) du$, and $S_2(x) = \int_x^\infty f_2(u) du$. Then $F_1(x) > F_2(x)$ for all $x \in (0, x^*)$ since for these x 's we have that $f_1(x) > f_2(x)$. Also, $S_2(x) \geq S_1(x)$ for all $x \in [x^*, \infty)$ since for these x 's we have that $f_2(x) \geq f_1(x)$. This implies that $F_1(x) \geq F_2(x)$ for all $x \in [x^*, \infty)$. Therefore, $X >^{SL} Y$, as required. \square
 Now, we select an intensity $\lambda(t)$ and an underlying survival density $f(x)$ which results in $f_{\text{bwd}}(x)$ and $f_{\text{fwd}}(x)$ crossing only once, such that $Y^{\text{fwd}} >^{SL} Y^{\text{bwd}}$. Let,

$$\lambda(t) = (0.01 - t)\mathbf{1}[0 < t \leq 0.01] + (t - 0.01)\mathbf{1}[0.01 < t < 1].$$

It is clear that $\lambda(t)$ is decreasing over $(0,0.01)$. Using equation (2), this implies that,

$$g(x) \propto (0.99 - x)\mathbf{1}[0 < x \leq 0.99] + (x - 0.99)\mathbf{1}[0.99 < x < 1].$$

Let $f(x) = \mathbf{1}[0 < x < 1]$. This implies that $S(x) = \mathbf{1}[x \leq 0] + (1 - x)\mathbf{1}[0 < x < 1]$. We define,

$$\begin{aligned} h(x) &= g(x)S(x) - \int_0^\infty g(u)f(u+x)du \\ &= \int_0^\infty (g(x) - g(u))f(u+x)du \\ &\propto f_{\text{bwd}}(x) - f_{\text{fwd}}(x). \end{aligned}$$

We now show that $\lim_{x \rightarrow 0^+} h(x) > 0$, and that $h(x)$ crosses zero only once on the interval $(0,1)$. This implies that $f_{\text{bwd}}(x)$ and $f_{\text{fwd}}(x)$ cross each other only once over $(0,1)$ and that $Y^{\text{fwd}} >^{SL} Y^{\text{bwd}}$. We have that for $0 < x < 1$,

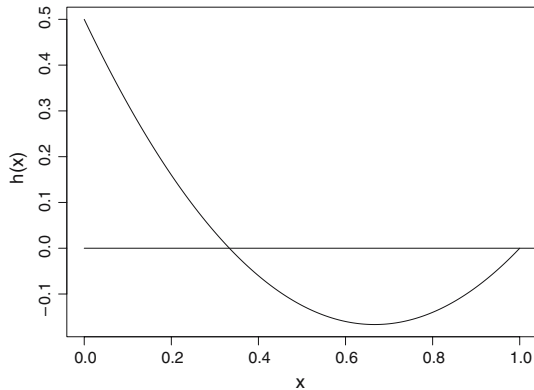
$$\begin{aligned} g(x)S(x) &\propto [(0.99 - x)\mathbf{1}[0 < x \leq 0.99] + (x - 0.99)\mathbf{1}[0.99 < x < 1]](1 - x) \\ &= (x^2 - 1.99x + 0.99)\mathbf{1}[0 < x \leq 0.99] - (x^2 - 1.99x + 0.99)\mathbf{1}[0.99 < x < 1] \end{aligned}$$

Also,

$$\begin{aligned} \int_0^\infty g(u)f(u+x)du &= \int_0^\infty g(u)\mathbf{1}[0 < u+x < 1]du \\ &\propto \int_0^\infty (0.99 - u)\mathbf{1}[0 < u < 0.99]\mathbf{1}[0 < u+x < 1]du \\ &\quad + \int_0^\infty (u - 0.99)\mathbf{1}[0.99 < u < 1]\mathbf{1}[0 < u+x < 1]du \\ &= \left[\frac{x^2}{2} - 0.01x + 0.4901 \right] \mathbf{1}[0 < x \leq 0.01] \\ &\quad + \left[\frac{-x^2}{2} + 0.01x + 0.49 \right] \mathbf{1}[0.01 < x < 1] \end{aligned}$$

This implies that,

$$h(x) \propto \begin{cases} \frac{x^2}{2} - 1.98x + 0.4999 & \text{if } 0 < x \leq 0.01 \\ \frac{3x^2}{2} - 2x + 0.5 & \text{if } 0.01 < x \leq 0.99 \\ \frac{-x^2}{2} + 1.98x - 1.48 & \text{if } 0.99 < x < 1 \end{cases}$$

Fig. 1 Graph of $h(x)$ 

The graph of $h(x)$, up to a constant of proportionality, on the interval $(0,1)$ is presented in Fig. 1. We see that $h(x)$ is strictly positive on $(0, \frac{1}{3})$, and strictly negative on $(\frac{1}{3}, 1)$, with $h(\frac{1}{3}) = 0$. This proves that $Y^{\text{fwd}} >^{SL} Y^{\text{bwd}}$, in spite of the fact that $\lambda(t)$ is strictly decreasing on $(0,0.01)$, which completes the counterexample.

Proof of Theorem 4 (\Leftarrow) See Theorem 3. (\Rightarrow) Since $Y^{\text{fwd}} >^{SL} Y^{\text{bwd}}$, we cannot have $\lambda(t)$ constant as this would violate the characterization of stationarity given by Asgharian et al. (2006). It is also not possible that $\lambda(t)$ is non-increasing since this would imply that $Y^{\text{bwd}} >^{SL} Y^{\text{fwd}}$. Hence, $\lambda(t)$ must be non-decreasing, and non-constant since, by assumption, $\lambda(t)$ belongs to the class of monotone intensity functions. \square

References

- Asgharian M, M'Lan CE, Wolfson DB (2002) Length-biased sampling with right censoring: an unconditional approach. *JASA* 97(457):201–209
- Asgharian M, Wolfson DB, Zhang X (2006) Checking stationarity of the incidence rate using prevalent cohort survival data. *Stat Med* 25:1751–1767
- Cheng KF (1984) Asymptotically nonparametric tests with censored paired data. *Commun Stat Theor Meth* 13(12):1453–1470
- The CSHA working group (1994) Canadian study of health and aging: study methods and prevalence of dementia. *J Can Med Assoc* 150:899–913
- The CSHA working group (2000) The incidence of dementia in Canada. *Neurology* 55:66–73
- Gehan EA (1965) A generalized Wilcoxon test for comparing arbitrarily singly-censored samples. *Biometrika* 52(1, 2):203–223
- Keiding N (1991) Age-specific incidence and prevalence: a statistical perspective. With discussion. *JRSS Ser A* 154(3):371–412
- Lancaster T (1990) *The econometric analysis of transition data*. Cambridge University Press, Cambridge
- Lewis PAW, Shedler GS (1976) Simulation of nonhomogeneous Poisson processes with log linear rate function. *Biometrika* 63(3):501–505
- Mantel N (1966) Evaluation of survival data and two new rank order statistics arising in its consideration. *Cancer Chemother Rep* 50(3):163–170
- Ross SM (2003) *Introduction to probability models*. Academic, San Diego
- Vardi Y (1989) Multiplicative censoring, renewal processes, deconvolution and decreasing density: nonparametric estimation. *Biometrika* 76(4):751–761

- Wang M-C (1991) Nonparametric estimation from cross-sectional survival data. *JASA* 86(413):130–143
- Wei LJ (1980) A generalized Gehan and Gilbert test for paired observations that are subject to arbitrary right censorship. *JASA* 75(371):634–637
- Wolfson C, Wolfson DB, Asgharian M, M'LAN CE, Østbye T, Rockwood K, Hogan DB (2001) A reevaluation of the duration of survival after the onset of dementia. *NEJM* 344(15):1111–1116