JIRSS (2013)
Vol. 12, No. 1, pp 35-69

Concepts and Tests for Trend in Recurrent Event Processes

R. J. Cook, J. F. Lawless

Department of Statistics and Actuarial Science, University of Waterloo, Ontario, Canada.

Abstract. Interest in the presence and nature of trend arises frequently in science, public health, technology, and many other areas. In this article we discuss the notion of trend in the context of recurrent event processes. We discuss different frameworks within which one can investigate trend and consider various ways in which trends may be manifest. Tests for trend are discussed in detail and the utility of intensity-based models is emphasized for characterizing event processes and understanding trends. Simulation studies are conducted to study the effect of heterogeneity in the investigation of trend. Data from a study of hospitalization patterns in patients with affective disorder are analysed for illustration.

Keywords. Intensity function, recurrent events, robust methods, stochastic modeling, trend.

MSC: 62N01, 62N02, 62N05, 62P10, 62P30, 62M05, 62M09.

1 Introduction

Processes of recurrent events are of interest in many settings, including medicine, sociology, economics and reliability (Cook and Lawless, 2007). In analysing such data, we are generally interested in understanding

R.J.Cook(✉)(rjcook@uwaterloo.ca), J.F.Lawless(jlawless@uwaterloo.ca)
Received: September 23, 2011; Accepted: June 29, 2012

aspects of the stochastic mechanism generating the events. This leads to consideration of models for the process intensity function, which specifies the probability of a new event at time t, given the previous history of event occurrences. In many situations the question arises about the presence and nature of trend. We discuss definitions of trend in what follows but, broadly speaking, the term refers to systematic variation in either event occurrence rates or times between events. Trends, as we discuss them, are monotonic (increasing or decreasing), which is in line with both the dictionary definition of trend as a general direction or tendency, and the common view of trend in time series (White and Granger, 2011).

Trends in recurrent event processes were discussed at some length by Cox and Lewis (1966), who gave a number of tests for trend in settings where there were no covariates. These and subsequent tests proposed in the literature assume a specific family of models within which certain sub-models are said to be trend-free. However, for models of recurrent events there is no universally adopted definition of a trend-free process or a process with trend (Ascher and Feingold, 1984, Section 9A). This is also true in other settings; in a review of trends in time series, White and Granger (2011) noted that "there is no generally accepted definition" of trend. Despite this, when certain types of trends are present they are often readily apparent in plots. In medical research, studies typically involve multiple individuals with covariates, however, and in that case simple plots may not provide clear evidence concerning trend. For example, a tendency for times between hospitalizations for psychiatric patients to decrease has been proposed (Kvist et al., 2008), but as we discuss later, this is difficult to establish.

The purpose of this paper is to review concepts of trend in recurrent event processes and the contexts in which they arise. We will focus on problems arising in chronic disease processes and on trends that manifest themselves at the individual subject level. Typical data sets involve many subjects but perhaps a rather small average number of events per subject. Standard notation and terminology for recurrent events will be used, as discussed by Cook and Lawless (2007).

Consider an individual process which starts at a designated time $t = 0$, and let $N(t)$ denote the number of events in $(0, t]$. The history of event occurrence at time t is denoted by $\mathcal{H}(t)$ and includes the number of events, $N(t) = n$, and their respective times $0 < T_1 < \cdots < T_n < t$. The times between events are called gap times and are denoted by $W_j = T_j - T_{j-1}$ $(j = 1, \ldots, n)$, where $T_0 = 0$. Unless stated otherwise

we consider processes in continuous time in which case the intensity function

$$\lambda\left(t|\mathcal{H}(t)\right) = \lim_{\Delta t \downarrow 0} \frac{\Pr\left\{N(t+\Delta t) - N(t) = 1|\mathcal{H}(t)\right\}}{\Delta t} \tag{1.1}$$

fully specifies the recurrent event process (Cook and Lawless, 2007, p. 10). Other features of a process that we consider are the marginal mean and rate functions, defined as

$$\mu(t) = E\{N(t)\}, \qquad \rho(t) = d\mu(t)/dt,$$

respectively.

As noted above, there is no universal definition of trend, or of its absence. The most common definitions of "no trend" are that either (a) the gap times W_j are identically distributed for $j = 1, 2, \ldots$ or (b) the rate function $\rho(t)$ is a constant α. The special case of (a) in which the W_j are also assumed to be independent gives a renewal process. Then, $\rho(t)$ approaches the value $\alpha = E(W_j)^{-1}$ as $t \to \infty$ although it may vary substantially for small values of t. Another special process is the equilibrium renewal process; this is a "delayed" renewal process for which W_j $(j = 2, 3, \ldots)$ are i.i.d. with distribution function $F(w)$ and survivor function $S(w)$, but with W_1 having density function $f_1(w) = \mu^{-1}S(w)$, where $\mu = E(W_j)$, $j = 1, 2, \ldots$ (Cook and Lawless, 2007, p. 148). The equilibrium renewal process has constant rate function $\rho(t) = \alpha = \mu^{-1}$. Such processes are sometimes useful when the time $t = 0$ indicates when observation of an individual begins, but their recurrent event process has already been operating for some time.

With the "no trend" frameworks (a) and (b) in mind, definitions of trend usually either involve (a) a model where the gap times W_j are stochastically increasing (or decreasing) in some way for $j = 1, 2, \ldots$, or (b) a model in which $\rho(t)$ is either monotonically increasing or decreasing. Our purpose in this paper is to consider models and tests for trend and we operate mainly within these frameworks. If individuals are observed over long periods so that each experiences many events, then checking for trend is fairly straightforward. As we will see, however, the assessment of trends involving rather large numbers of individuals who experience rather small numbers of event is often difficult. Moreover, this can be exacerbated by heterogeneity across individuals (see Section 3), the existence of systematic temporal factors, and event-related end-of-followup times τ_i.

The remainder of this paper is organized as follows. In Section 2 we review the canonical frameworks for analysing recurrent events and

associated tests for trend. Tests which accommodate heterogeneity between individuals are discussed along with robust tests. In Section 3 we discuss the possible impact of heterogeneity on inferences regarding trends in gap time analyses which ignore this feature. This is done analytically and through simulation studies. Intensity-based approaches for assessing trend are then discussed and these are used in Section 4 to model and study different types of trend in the context of a study of recurrent hospitalization among patients with affective disorder. Concluding remarks and topics for further research are discussed in Section 5.

2 Review of Models and Tests for Trend

We assume that m independent processes are under study, and focus on settings where m is reasonably large and the numbers of events n_i for individuals $i = 1, \ldots, m$ are small to moderate. As discussed in Section 1, we consider two main types of model that represent absence of trend: (a) models where the times W_j $(j = 1, 2, \ldots)$ between successive events are identically distributed for a given individual, and (b) models where the rate function $\rho(t)$ for an individual is constant. Case (a) typically requires a full specification of the recurrent event processes for reasons indicated below. Case (b) is often considered when we wish to focus on rate and mean functions for the recurrent events, and avoid full specification of the process.

We provide a brief review of previous work on testing trend; each setting involves an assumed family of models that includes sub-models of type (a) or (b) for absence of trend, as well as alternatives that incorporate trend.

2.1 Tests Based on Poisson Processes

Many early tests of trend were based on Poisson processes with rate and intensity functions

$$\lambda_i\left(t|\mathcal{H}_i(t)\right) = \rho_i(t) = \alpha_i \exp(\beta g(t)), \qquad i = 1, \ldots, m\ , \tag{2.1}$$

where $g(t)$ is a specified function, β is a real-valued parameter and $\alpha_1, \ldots, \alpha_m$ are positive-valued parameters. When $\beta = 0$ there is no trend in the processes. If data consist of event times t_{ij} $(j = 1, \ldots, n_i)$ observed over specified time intervals $(0, \tau_i]$ for individuals $i = 1, \ldots, m$,

a simple test of $\beta = 0$ can be obtained from the conditional distributions of the event times given that $N_i(\tau_i) = n_i$. The nuisance parameters $\alpha_1, \ldots, \alpha_m$ can be eliminated by this conditioning, leading to the conditional likelihood function (Cox and Lewis, 1966, Sec. 3.3)

$$L_c(\beta) = \prod_{i=1}^{m} \prod_{j=1}^{n_i} \left\{ \frac{\exp(\beta g(t_{ij}))}{\int_0^{\tau_i} \exp(\beta g(t)) dt} \right\} , \tag{2.2}$$

which can be used to test $\beta = 0$. A particularly simple test is obtained from the score statistic $U_c(0) = (\partial \log L_c(\beta)/\partial\beta)|_{\beta=0}$, which reduces to

$$U_c(0) = \sum_{i=1}^{m} U_{ci}(0) = \sum_{i=1}^{m} \left\{ \sum_{j=1}^{n_i} g(t_{ij}) - \frac{n_i}{\tau_i} \int_0^{\tau_i} g(t) dt \right\} , \tag{2.3}$$

where $U_{ci}(0)$ is the contribution from individual i to $U_c(0)$. Under H_0, conditional on $n_1, \ldots, n_m$, the variance of $U_c(0)$ is (Cox and Lewis, 1966, Sec. 3.3; Cook and Lawless, 2007, Sec. 3.7)

$$\text{var}\{U_c(0)\} = \sum_{i=1}^{m} n_i \left\{ \frac{1}{\tau_i} \int_0^{\tau_i} g^2(t) dt - \left[\frac{1}{\tau_i} \int_0^{\tau_i} g^2(t) dt \right]^2 \right\} , \tag{2.4}$$

and under mild conditions, $Z = U_c(0)/\text{var}\{U_c(0)\}^{1/2}$ is asymptotically standard normal as $m \to \infty$.

Versions of this test with $g(t) = t$ and $g(t) = \log t$ are especially well known and widely used (e.g. see Ascher and Feingold, 1984, Sec. 5B; Bain et al., 1985; Cohen and Sackrowitz, 1993). These tests can be extended in a number of directions. First, while the α_i allow for heterogeneity in event rates across individuals, in some cases there may also be fixed covariate vectors x_i $(i = 1, \ldots, m)$. If these affect the intensity multiplicatively, so that $\rho_i(t)$ becomes $\alpha_i \exp(\beta g(t)) f(x_i)$ for some positive-valued function f, then once again we are led to the conditional likelihood (2.2), with the $f(x_i)$ eliminated. This fact has been used by Kvist et al. (2008) and others. Second, if extra flexibility is needed for representing trends, $g(t)$ and β in (2.1) can be vectors (e.g. Agustin and Pena, 2005). Third, in settings where the n_i tend to be small, tests based on (2.2) may lack power. In this case we may assume that the α_i are identically distributed random variables whose distribution involves a small number of parameters; alternatively we may opt to model heterogeneity solely in terms of external observable covariates, and use the unconditional likelihood function (2.11) instead of (2.2). In both cases

however, the assessment of trend becomes more dependent on explicit modeling assumptions than when (2.2) is used.

Finally, we note that the tests considered here depend critically on the null (no trend) and alternative models being Poisson processes. The null homogeneous Poisson process is the only process which is an ordinary renewal process (with exponential gap times W_j) and at the same time has a constant rate function. We now consider more general tests based on rate functions and on renewal processes.

2.2 Tests Based on Rate Functions

More robust tests of trend can be based on robust methods for estimation of rate and mean functions (Lawless and Nadeau, 1995; Cook and Lawless, 2007, Ch. 3). One family of tests employs the same type of score statistic $U_c(0)$ as in (2.3), since it can be shown that $E\{U_c(0)\} = 0$ as long as the τ_i are specified followup times and the rate functions $\rho_i(t)$ are of the form (2.1) with $\beta = 0$. However, the variance estimate (2.4) does not apply unless the process is Poisson, so it is replaced by the robust estimate

$$\widehat{\text{var}}\{U_c(0)\} = \sum_{i=1}^{m} U_{ci}(0)^2 \, . \tag{2.5}$$

Under mild conditions, the statistic $Z = U_c(0)/\widehat{\text{var}}\{U_c(0)\}^{1/2}$ is asymptotically standard normal as $m \to \infty$. Robust tests of this sort have been considered in more detail by Cook et al. (1996), Cook and Lawless (2007, Sec. 3.7) and Lawless et al. (2012).

2.3 Tests Based on Renewal Processes

There is also a substantial literature on tests of a "no trend" null hypothesis H_0 for a renewal process, where for each individual process $i = 1, \ldots, m$, the gap times W_{ij} $(j = 1, 2, \ldots)$ are independent and identically distributed (e.g. see Cox and Lewis, 1966, Sec. 3.4; Lewis and Robinson, 1974; Wang and Chen, 2000; Kvaloy and Lindqvist, 2003; Lawless et al., 2012). The tests for renewal processes are typically based on the assumption that the number of events n_i observed for each individual is fixed, rather than the followup time. The simplest procedure is, for the ith process, to use a rank statistic that tests for no association between the gap times W_{ij} and a specified covariate x_{ij} that reflects the type of trend of interest (Cox and Lewis, 1966, Sec. 3.4). In using a rank test we replace the W_{ij} with scores; for illustration we take $x_{ij} = j$

and consider ordered exponential (log rank) scores

$$\psi_{ij} = \frac{1}{n_i} + \frac{1}{n_i - 1} + \ldots + \frac{1}{n_i - r_{ij} + 1}\,, \tag{2.6}$$

where r_{ij} is the rank of W_{ij} among $W_{i1}, \ldots, W_{in_i}$. The rank statistic is then

$$U_i = \sum_{j=1}^{n_i} \psi_{ij} \left(x_{ij} - \bar{x}_i\right) \tag{2.7}$$

and under the null (no trend) hypothesis that the W_{ij} $(j = 1, \ldots, n_i)$ are i.i.d., U_i has mean zero and variance (Kalbfleisch and Prentice, 2002, Sec 7.2)

$$\mathrm{var}(U_i) = \left\{ \sum_{j=1}^{n_i} \left(x_{ij} - \bar{x}_i\right)^2 \right\} \left\{ \sum_{j=1}^{n_i} \frac{\left(\psi_{ij} - \bar{\psi}_i\right)^2}{n_i - 1} \right\}. \tag{2.8}$$

Combining across individuals $i = 1, \ldots, m$, a test of no trend can be based on the statistic

$$Z_R = \sum_{i=1}^{m} U_i \Big/ \left\{ \sum_{i=1}^{m} \mathrm{var}(U_i) \right\}^{1/2}, \tag{2.9}$$

which under the null hypothesis of no trend is asymptotically normal as $m \to \infty$ and, for fixed m, as the $n_i \to \infty$. The statistic Z_R will be effective against alternatives for which the W_{ij} are stochastically increasing or decreasing as j increases.

In most biomedical applications the followup time τ_i is fixed or can be treated as fixed, rather than the number of events n_i, $i = 1, \ldots, m$. However, the rank statistic (2.7) is still suitable under the null renewal hypothesis since the W_{ij} $(j = 1, \ldots, n_i)$ are exchangeable given that $N_i(\tau_i) = n_i$. The permutation variance estimate (2.8) is also valid, so a test of no trend can be carried out using (2.7) - (2.9). This test accounts for heterogeneity by summing rank statistics across the m processes. A normal approximation or a permutation (resampling) approach can be used to get p-values when m is small and the normal approximation is inadequate. Any resampling method must obey the null hypothesis H_0, and bootstrap procedures that have been proposed for point processes (e.g. Loh, 2010) do not do this. The procedure whereby we permute the gap times W_{ij} $(j = 1, \ldots, n_i)$ within each individual, thus generating new event times t_{ij} $(j = 1, \ldots, n_i)$, is valid however, because under H_0

the gap times are identically distributed and exchangeable. By generating B new sets of data this way we can approximate the null distribution of Z_R. The null distribution of the pseudoscore statistic can also be approximated this way.

We remark that the practice of plotting separate Kaplan-Meier estimates of the survival function for each successive gap (i.e. first gap, second gap, etc.) is inappropriate when there is heterogeneity, due to a form of induced dependent censoring (Cook and Lawless, 2007, Sec. 4.4), and can lead to a false indication of trend; see Section 3.2.

The rank tests above allow for heterogeneity across individuals, in fact in a much more general way than the tests of Sections 2.1 and 2.2. They are also simpler than the nonparametric tests proposed by Wang and Chen (2000), who consider a statistic based on pairwise comparisons of gap times that is similar in spirit to a Wilcoxon rank test. However, we note two situations that cause problems for the rank tests, as well as the tests in Sections 2.1 and 2.2. First, if the gap times W_{ij} $(j = 1, 2, \ldots)$ for an individual form a stationary but not i.i.d. series, the variance estimates (2.8) are no longer valid. Second, if the stopping time τ_i is determined adaptively based on the event history (e.g. if an individual is more likely to be lost to followup if they experience many events), then the distributions of the test statistics considered so far are not as stated. To deal with such complications, we have to consider models for the process intensities in more detail, and we turn to this next.

2.4 Tests Based on Intensity Specifications

Time trends may occur because of factors related to the age of a process or to the occurrence of previous events. A model that allows for both types of trend is one with intensity function

$$\lambda\left(t|\mathcal{H}(t)\right) = h_0\left(B(t)\right) e^{\beta_1 g_1(t) + \beta_2 g_2(N(t^-))} \tag{2.10}$$

where $h_0(\cdot)$ is a non-negative function and $B(t) = t - t_{N(t^-)}$ is the time since the last event. Such a model is often called a modulated renewal process (Cook and Lawless, 2007, p. 132). If $\beta_1 = \beta_2 = 0$ the process is a renewal process but otherwise can be said to involve a trend. The model (2.10) can be extended to allow for heterogeneity across individuals, for example by adding a multiplicative random effect in front of $h_0(B(t))$. Doing this is sometimes necessary for the model to adequately represent the processes of interest, but comes at a price, because dependence of the intensity on the number of previous events

is confounded with heterogeneity. See, for example, Cook and Lawless (2007, Sec. 3.5.3), Aalen et al. (2008, Sec. 7.3), and the next section. Thus, it is usually of interest to account for heterogeneity as much as possible through a vector of covariates x on the individuals, say by incorporating a multiplicative term $\exp(\gamma' x)$ in (2.10). In some cases we might wish to allow time-varying covariates in order to look for a trend after adjusting for other temporal effects such as seasonality.

If (2.10) or some other model adequately describes the intensity for the process, then the likelihood function for m independent individuals is

$$L = \prod_{i=1}^{m} \left\{ \prod_{j=1}^{n_i} \lambda_i \left(t|\mathcal{H}_i(t_{ij})\right) \right\} \exp\left\{ -\int_0^{\tau_i} \lambda_i \left(t|\mathcal{H}_i(t)\right) dt \right\} , \qquad (2.11)$$

and it remains valid under event-dependent stopping times τ_i (Cook and Lawless, 2007, p. 30). Parametric models are easily handled by ordinary maximum likelihood, as we illustrate later. Semiparametric models in which $h_0(w)$ in (2.10) is allowed to be an arbitrary positive-valued function can also be handled using the Cox partial likelihood approach in many cases (Dabrowska et al., 1994; Lawless et al., 2001).

3 Heterogeneity and Trend in Gap Time Analyses

3.1 Impact of Heterogeneity on Gap Time Analyses

Here we explore the effect of model misspecification in the context of recurrent event gap time analyses directed at the investigation of trend. We consider a simple model with a conditional intensity

$$\lambda_i(t|\alpha_i, \mathcal{H}_i(t)) = \alpha_i h_0(B_i(t)) \exp(\beta N_i(t^-)) \qquad (3.1)$$

where α_i is a gamma distributed random effect for individual i with $E(\alpha_i) = 1$ and $\text{var}(\alpha_i) = \phi$, and $B_i(t) = t - t_{iN_i(t^-)}$. Such "modulated renewal" models have been considered by various authors (e.g. Pena, 2006). This is a conditional modulated renewal model which incorporates a trend in the gap time distributions when $\beta \neq 0$; when $\beta > 0$ the gap times tend to get shorter as the cumulative number of events increases. By averaging over the unobservable random effect, we obtain the intensity of the form

$$\lambda_i(t|\mathcal{H}_i(t)) = E(\alpha_i|\mathcal{H}_i(t)) h_0(B_i(t)) \exp(\beta N_i(t^-)) \, . \qquad (3.2)$$

If $H_0(w) = \int_0^w h_0(u)du$ and $N_i(t^-) = n-1$, following the derivation in the Appendix we obtain

$$E(\alpha_i|\mathcal{H}_i(t)) = \frac{1+\phi(n-1)}{1+\phi[\sum_{j=1}^{n-1} H_0(w_{ij})\exp(\beta(j-1)) + H_0(B_i(t))\exp(\beta(n-1))]}. \tag{3.3}$$

Upon the occurrence of the nth event, the exponential term in (3.2) introduces a persistent multiplicative factor $\exp(\beta)$ to the intensity function. The intensity also increases at the nth event, by a multiplicative factor $(1+\phi n)/(1+\phi(n-1))$ from (3.3), but (3.3) also decreases with increasing time $B_i(t)$ since the most recent event, due to the $H_0(B_i(t))$ term in the denominator which increases with t; the intensity is also influenced by the preceding gap times, through their presence in the denominator of (3.3).

The introduction of random effects therefore accommodates heterogeneity in the respective gap time distributions between individuals, but their introduction can also be viewed as a device to generate an intensity which reflects transient changes in risk following event occurrence. In contrast, the multiplicative term $\exp(\beta N_i(t^-))$ reflects a persistent effect of event occurrence on risk of the next event. The ability to distinguish between these two types of event dependence is related to the magnitude of ϕ and β, the distribution of the number of events per individual, and the number of individuals.

3.2 Simulation Studies

The purpose of these simulation studies is to explore the issue of heterogeneity and model misspecification in the assessment of trend based on gap time analyses. In the first scenario we generate data from a model with $\beta = 0$ in (3.1) so there is no renewal trend, but set $\phi > 0$ so there is heterogeneity. In the second scenario we take $\phi = 0$ in (3.1) and consider values of β greater than or equal to zero. Here we use $\bar{N}_i(t^-)$ in place of $N_i(t^-)$ in (3.1) where $\bar{N}_i(t^-) = N_i(t^-)$ if $N_i(t^-) \leq 20$ and $\bar{N}_i(t^-) = 20$ otherwise. This upper limit is adopted to simplify computation of expected numbers of events and minimize the probability of generating unduly large numbers of events for some individuals. In the third scenario we consider both heterogeneity ($\phi > 0$) and a renewal trend ($\beta > 0$) with $\bar{N}_i(t^-)$. Data are simulated under a constant baseline hazard so $h_0(w) = h_0$, and events are generated over the time interval $(0, 1]$. In each of the settings studied, h_0 was determined so that the expected number of events per individual was 1, 2 or 4. Two

thousand data sets of $m = 1,000$ individuals are simulated for each of the parameter configurations.

For each scenario we fit three semiparametric models in which the form of $h_0(w)$ is left unspecified : a semiparametric *modulated renewal* model based on (3.1) without the random effect, a semiparametric *mixed renewal* model based on (3.1) with the constraint $\beta = 0$, and a semiparametric *hybrid* model based on (3.1) with no constraints. The empirical mean and empirical standard error (ESE) of the parameter estimates from each of the three fitted models are reported. We also implemented the pseudo-score test for trend based on (2.3) with $g(t) = t$ and using the robust variance in (2.5). As discussed in Section 2.1, this test is geared towards trends in the rate function and accommodates latent individual-specific effects and fixed covariates. Finally we carried out the rank-based test for trend using the statistic (2.9) with scores given by (2.6). For each trend test we report the empirical rejection rates as the proportion of simulated datasets for which the test statistic exceeded the nominal 5% critical value.

Table 1 contains the results of the simulation studies for the first scenario ($\beta = 0$). We considered values of ϕ from 0.1 for minimal heterogeneity, 0.2 for mild heterogeneity to 0.4 for moderate heterogeneity. In all cases in Table 1 the mean estimate of β from the fit of the modulated renewal model is positive reflecting the phenomenon that the heterogeneity between individuals leads to an apparent trend in the gap time distributions. That is, even though each individual has no trend in their event process, when the heterogeneity between individuals is not accounted for, this omission creates an apparent trend in which the mean gap times decrease as the number of events increases. The magnitude of $E(\widehat{\beta})$ increases as ϕ increases, but for a given ϕ, is decreasing as $E\{N(\tau)\}$ increases. The latter phenomenon reflects the fact that in (3.3), larger $E\{N(\tau)\}$ comes from scenarios with larger $H_0(w) = h_0 w$, so a smaller value of β produces a given value of (3.3).

The estimates of ϕ from the mixed renewal model are generally good but slightly conservative. For the hybrid model the empirical bias of the estimator of β is much smaller than for the modulated renewal model since the heterogeneity is adequately accounted for. The empirical standard error for $\widehat{\beta}$ from the hybrid model tends to be larger than it is for the misspecified modulated renewal model. Moreover the empirical standard errors for $\widehat{\beta}$ are much larger than the corresponding average model-based standard error (ASE) under the hybrid model. This is due to inadequacy of the usual asymptotic normal approximations, which we

discuss below. The empirical rejection rates of the robust pseudo-score test and the rank test for trend are compatible with the nominal 5% level for all parameter configurations in Table 1.

Table 2 contains the results from the second scenario in which data were generated with no heterogeneity (i.e. $\phi = 0$). We set $\beta = 0$, $\log 1.10 \approx 0.095$ and $\log 1.25 \approx 0.223$ to reflect a renewal model, as well as modulated renewal models with mild and moderate increase in risk with each event occurrence; again we consider the case where the expected number of events was 1, 2 or 4. When the (correct) modulated renewal model is fit, as expected the empirical biases in $\widehat{\beta}$ are negligible, there is good agreement between the empirical and average model-based standard error. The mean variance estimate (average of $\widehat{\phi}$) from the mixed renewal model is close to zero when $\beta = 0$, but increases with larger values of $\beta > 0$. This indicates that trends at the individual level must be adequately modeled to ensure estimates of ϕ reflect only heterogeneity between individuals and not other types of trend or model misspecification. The estimator of β from the semiparametric hybrid model has considerably greater empirical standard error than average model-based standard error when $E\{N(\tau)\} = 1$, as in Table 1, and there is also a non-negligible positive bias in the estimate of ϕ. The empirical performance of the estimators of β and ϕ from the hybrid model are considerably improved for $E\{N(\tau)\} = 2$ or 4.

The empirical rejection rates of the two trend tests under $\beta = 0$ represent empirical type I error rates and these are again compatible with the nominal level. For $\beta > 0$ these rejection rates give the empirical power and we see greater power with larger β and larger $E\{N(\tau)\}$. Among the two trend tests the robust pseudo-score test appears to be the most powerful for trends of this type.

Table 3 contains the results of fitting the three models to data generated with combinations of $\phi = 0.2$ and 0.4 and $\beta = \log 1.1 \approx 0.095$ and $\log 1.25 \approx 0.223$. Here we see that estimates of β from the modulated renewal model are inflated. Likewise, the mixed renewal model gives estimates of ϕ which, on average, are considerably larger than the true values. The hybrid model yields estimates which are somewhat positively biased for β and negatively biased for ϕ; both biases decrease as $E\{N(\tau)\}$ increases. There is considerably greater variability in the estimate of β from the hybrid model than is accounted for by the model-based standard errors (ASE); the relative over-estimation decreases as $E\{N(\tau)\}$ increases. Again, as one would expect, the empirical power of the trend tests is greater for larger values of β, but interestingly for

Table 1: Empirical frequency properties of parameter estimators based on modulated renewal, mixed renewal and hybrid models, and rejection rates (%) of the robust pseudo-score test and rank test for trend; data generated over $(0, \tau = 1]$ according to intensity (3.1) with $\beta = 0$ and $\phi > 0$; $m = 1,000$; $2,000$ simulations.

			$\phi = 0.1$			$\phi = 0.2$			$\phi = 0.4$		
$E\{N(\tau)\}$	PARAMETER	MODEL/TEST	MEAN	ESE†	ASE‡	MEAN	ESE†	ASE‡	MEAN	ESE†	ASE‡
1	β	Modulated Renewal	0.067	0.039	0.040	0.122	0.036	0.037	0.197	0.029	0.032
		Hybrid	0.013	0.100	0.040	0.040	0.110	0.038	0.032	0.111	0.034
	ϕ	Mixed Renewal	0.063	0.074	-	0.184	0.086	-	0.391	0.094	-
		Hybrid	0.093	0.167	-	0.151	0.201	-	0.352	0.251	-
	% Rejection	Robust Pseudo-Score		5.5			5.5			4.5	
	% Rejection	Rank Test		4.7			5.6			4.8	
2	β	Modulated Renewal	0.055	0.018	0.018	0.096	0.016	0.017	0.143	0.012	0.014
		Hybrid	0.036	0.041	0.018	0.033	0.055	0.017	0.009	0.034	0.015
	ϕ	Mixed Renewal	0.073	0.055	-	0.190	0.047	-	0.341	0.053	-
		Hybrid	0.037	0.072	-	0.132	0.118	-	0.372	0.113	-
	% Rejection	Robust Pseudo-Score		5.5			5.5			5.2	
	% Rejection	Rank Test		5.5			4.8			5.7	

Table 1: Continued

			$\phi = 0.1$			$\phi = 0.2$			$\phi = 0.4$		
$E\{N(\tau)\}$	PARAMETER	MODEL/TEST	MEAN	ESE†	ASE‡	MEAN	ESE†	ASE‡	MEAN	ESE†	ASE‡
4	β	Modulated Renewal	0.042	0.008	0.008	0.068	0.007	0.007	0.094	0.005	0.006
		Hybrid	0.021	0.023	0.008	0.004	0.014	0.008	0.004	0.012	0.007
	ϕ	Mixed Renewal	0.105	0.031	-	0.165	0.024	-	0.392	0.041	-
		Hybrid	0.050	0.053	-	0.182	0.046	-	0.374	0.075	-
	% Rejection	Robust Pseudo-Score		5.0			5.9			5.1	
	% Rejection	Rank Test		4.5			5.2			5.1	

†ESE is the standard deviation of the estimate across the 2000 simulations.
‡ASE is the average of the estimated model-based standard deviations for the estimate over the 2000 simulations.

Table 2: Empirical frequency properties of parameter estimators based on modulated renewal, mixed renewal and hybrid models, and rejection rates (%) of the robust pseudo-score test and rank test for trend; data generated over $(0, \tau = 1]$ according to intensity (3.1) with $\beta \geq 0$ and $\phi = 0$; $m = 1,000$; $2,000$ simulations.

			$\beta = 0.0$			$\beta = 0.095$			$\beta = 0.223$		
$E\{N(\tau)\}$	PARAMETER	MODEL/TEST	MEAN	ESE†	ASE‡	MEAN	ESE†	ASE‡	MEAN	ESE†	ASE‡
1	β	Modulated Renewal	-0.002	0.044	0.043	0.093	0.040	0.039	0.220	0.033	0.033
		Hybrid	-0.031	0.083	0.043	0.066	0.078	0.040	0.193	0.072	0.034
	ϕ	Mixed Renewal	0.004	0.019	-	0.104	0.085	-	0.384	0.097	-
		Hybrid	0.047	0.120	-	0.047	0.121	-	0.051	0.128	-
	% Rejection	Robust Pseudo-Score		5.4			15.5			66.0	
	% Rejection	Rank Test		5.2			5.9			17.7	
2	β	Modulated Renewal	-0.001	0.020	0.020	0.094	0.018	0.018	0.221	0.012	0.011
		Hybrid	-0.003	0.023	0.020	0.092	0.021	0.018	0.219	0.013	0.011
	ϕ	Mixed Renewal	0.001	0.003	-	0.142	0.048	-	0.522	0.104	-
		Hybrid	0.004	0.019	-	0.004	0.021	-	0.003	0.017	-
	% Rejection	Robust Pseudo-Score		4.9			73.0			100.0	
	% Rejection	Rank Test		5.4			27.2			99.8	

Table 2: Continued

			$\beta = 0.0$			$\beta = 0.095$			$\beta = 0.223$		
$E\{N(\tau)\}$	PARAMETER	MODEL/TEST	MEAN	ESE†	ASE‡	MEAN	ESE†	ASE‡	MEAN	ESE†	ASE‡
4	β	Modulated Renewal	-0.000	0.009	0.009	0.095	0.007	0.007	0.223	0.004	0.004
		Hybrid	-0.001	0.009	0.009	0.094	0.008	0.007	0.223	0.004	0.004
	ϕ	Mixed Renewal	0.001	0.004	-	0.150	0.016	-	0.907	0.076	-
		Hybrid	0.001	0.001	-	0.001	0.001	-	0.001	0.004	-
	% Rejection	Robust Pseudo-Score		5.4			100.0			100.0	
	% Rejection	Rank Test		4.9			100.0			100.0	

†ESE is the standard deviation of the estimate across the 2000 simulations.
‡ASE is the average of the estimated model-based standard deviations for the estimate over the 2000 simulations.

a given value of β the powers of the two tests are greater for larger values of ϕ. This might be due to the occurrence of more individuals with larger n_i when ϕ is bigger, though the mean n_i is the same. The robust pseudo-score test often has higher power, but the exception is when $\beta = 0.223$ and $\phi = 0.4$ where the rank test has 99.5% empirical power compared to 97% seen for the robust pseudo-score test. The higher power of the pseudo-score test may be due to the fact that all individuals with $n_i \geq 1$ contribute to it, whereas only those with $n_i \geq 2$ are used in the rank test.

The bias in $\widehat{\beta}$ and $\widehat{\phi}$ and the poor agreement between the ASE and ESE of $\widehat{\beta}$ seen in the empirical results for the hybrid model in Tables 1-3, are related to the fact that $\widehat{\beta}$ and $\widehat{\phi}$ are negatively correlated and that when $E\{N(\tau)\}$ is fairly small, the usual normal approximation to the sampling distribution of $(\widehat{\beta}, \widehat{\phi})$ is poor; the performance is particularly poor for smaller m. A further problem arises from the fact that $\phi \geq 0$ and when $\phi = 0$ the maximum likelihood estimate of (β, ϕ) has a non-standard distribution (Self and Liang, 1987; Barnabani, 2008). When ϕ is positive but small, sample sizes (in terms of m and n_i) must be very large for the standard asymptotic normal approximations to be valid.

For illustration we examine six simulated datasets with $m = 100$ from the settings where $(\beta, \phi) = (\log 1.25, 0)$, $(0, 1.0)$ and $(\log 1.25, 1.0)$ for $E\{N(\tau)\} = 1$ and 4. Figure 1 contains contour plots for the profile relative likelihood function for (β, ϕ) denoted $RL_p(\beta, \phi)$, defined by setting $-2 \log RL_p(\beta, \phi)$ equal to the 50th, 80th, 90th and 95th percentiles of the χ^2_1 distribution; these would be relevant for obtaining profile likelihood-based confidence intervals for the individual parameters. For all datasets the negative association between the estimates is apparent from the orientation of the contours. For the cases in which $E\{N(\tau)\} = 1$ the non-elliptical shape of the contours are most evident, as is the impact of the boundary constraint for ϕ when there is no heterogeneity. When $E\{N(\tau)\} = 4$ the precision is much greater but the impact of the boundary constraint is still evident.

We also consider two larger simulated datasets of $m = 1000$ and $E\{N(\tau)\} = 4$; one is simulated with $\phi = 0$ and $\beta = \log 1.25$ and another with $\phi = 1.0$ and $\beta = 0$. For each of these datasets we fitted a semiparametric modulated renewal model and a hybrid model with

$$\lambda(t|\mathcal{H}_i(t)) = h_n(B_i(t)) \,, \qquad n = N_i(t^-) \tag{3.4}$$

and

$$\lambda(t|\alpha_i, \mathcal{H}_i(t)) = \alpha_i h_n(B_i(t)) \,, \qquad n = N_i(t^-) \tag{3.5}$$

Table 3: Empirical frequency properties of parameter estimators based on modulated renewal, mixed renewal and hybrid models, and rejection rates (%) of the robust pseudo-score test and rank test for trend; data generated over $(0, \tau = 1]$ according to intensity (3.1) with $\beta > 0$ and $\phi > 0$; $m = 1,000$; $2,000$ simulations.

		$E\{N(\tau)\} = 1$			$E\{N(\tau)\} = 2$			$E\{N(\tau)\} = 4$		
PARAMETER	MODEL/TEST	MEAN	ESE†	ASE‡	MEAN	ESE†	ASE‡	MEAN	ESE†	ASE‡
					$\phi = 0.2$ and $\beta = 0.095$					
β	Modulated Renewal	0.208	0.030	0.032	0.175	0.011	0.012	0.142	0.004	0.004
	Hybrid	0.131	0.102	0.033	0.122	0.045	0.013	0.098	0.008	0.005
ϕ	Mixed Renewal	0.387	0.093	-	0.407	0.070	-	0.588	0.053	-
	Hybrid	0.154	0.204	-	0.137	0.120	-	0.184	0.046	-
% Rejection	Robust Pseudo-Score		20.9			94.0			100.0	
% Rejection	Rank Test		9.4			71.7			100.0	
					$\phi = 0.2$ and $\beta = 0.223$					
β	Modulated Renewal	0.296	0.023	0.019	0.258	0.007	0.005	0.245	0.004	0.003
	Hybrid	0.235	0.064	0.021	0.226	0.011	0.006	0.221	0.004	0.003
ϕ	Mixed Renewal	0.860	0.177	-	1.515	0.161	-	1.813	0.090	-
	Hybrid	0.180	0.169	-	0.182	0.064	-	0.244	0.042	-
% Rejection	Robust Pseudo-Score		91.5			100.0			100.0	
% Rejection	Rank Test		75.5			100.0			100.0	

Table 3: Continued

		$E\{N(\tau)\}=1$			$E\{N(\tau)\}=2$			$E\{N(\tau)\}=4$		
PARAMETER	MODEL/TEST	MEAN	ESE†	ASE‡	MEAN	ESE†	ASE‡	MEAN	ESE†	ASE‡
					$\phi=0.4$ and $\beta=0.095$					
β	Modulated Renewal	0.264	0.022	0.025	0.191	0.010	0.008	0.150	0.004	0.003
	Hybrid	0.115	0.088	0.029	0.098	0.019	0.010	0.099	0.006	0.004
ϕ	Mixed Renewal	0.660	0.109	-	0.776	0.094	-	1.000	0.065	-
	Hybrid	0.368	0.229	-	0.386	0.097	-	0.345	0.068	-
% Rejection	Robust Pseudo-Score		28.6			99.3			100.0	
% Rejection	Rank Test		15.2			98.2			100.0	
					$\phi=0.4$ and $\beta=0.223$					
β	Modulated Renewal	0.297	0.019	0.011	0.263	0.006	0.004	0.249	0.003	0.003
	Hybrid	0.224	0.028	0.013	0.223	0.006	0.005	0.220	0.004	0.003
ϕ	Mixed Renewal	1.514	0.281	-	2.239	0.172	-	2.341	0.124	-
	Hybrid	0.396	0.136	-	0.415	0.077	-	0.518	0.073	-
% Rejection	Robust Pseudo-Score		97.0			100.0			100.0	
% Rejection	Rank Test		99.5			100.0			100.0	

†ESE is the standard deviation of the estimate across the 2000 simulations.
‡ASE is the average of the estimated model-based standard deviations for the estimate over the 2000 simulations.

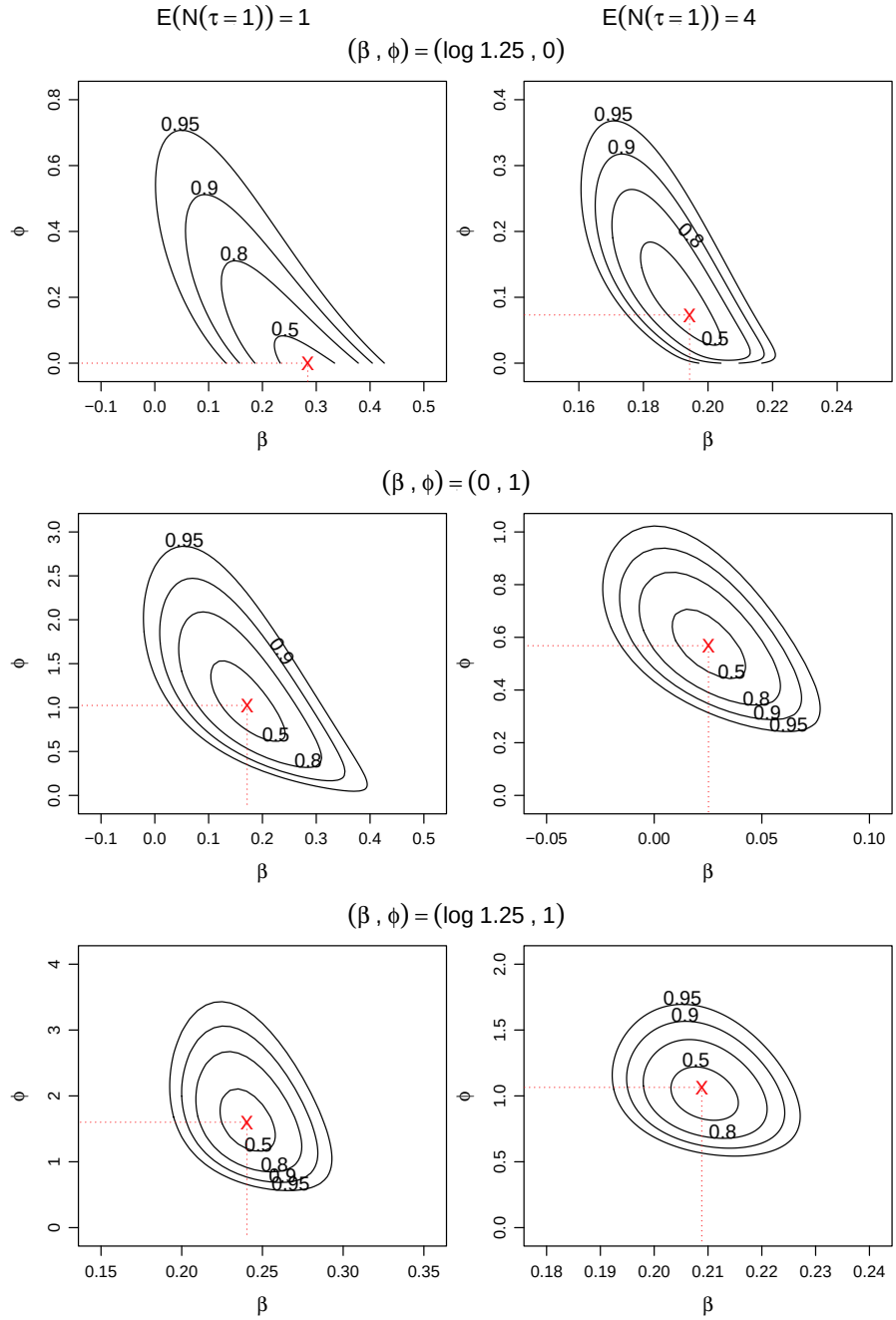

Figure 1: Relative profile likelihood contour plots for (β, ϕ) from the hybrid model (3.1) for several simulated datasets; $E\{N(\tau)\} = 1$ (left column) and $E\{N(\tau)\} = 4$ (right column); $m = 100$; p contours are obtained by setting $-2 \log RL_p(\beta, \phi)$ equal to the $100p$th percentile of the χ^2_1 distribution, $p = 0.50, 0.80, 0.90$ and 0.95.

respectively, where the $h_n(w)$, $n = 0, 1, 2, \ldots$ are arbitrary hazard functions. Estimates of the cumulative baseline hazard functions $\widehat{H}_n(w)$, $n = 0, 1, 2, \ldots$, are plotted for both fitted models for each dataset in Figure 2. The top two figures correspond to the data set generated with $(\beta, \phi) = (\log 1.25, 0)$ under models (3.4) and (3.5) and the bottom two are for the dataset with $(\beta, \phi) = (0, 1)$. When there is no heterogeneity present, the estimated cumulative baseline hazards are comparable for the Markov renewal model and mixed Markov renewal (hybrid) models. However, when the true process is a mixed renewal process, the modulated renewal model incorrectly suggests a trend towards shorter gap times with increasing number of events. When the heterogeneity is adequately dealt with through the hybrid model, no trend is seen.

4 Assessment of Trend in Psychiatric Admissions

We consider here the analysis of data from a study on hospitalization patterns for psychiatric patients with affective disorder (Kessing et al., 2004). We begin with a description of the data, which deal with recurrent hospitalizations over the years 1994 to 1999. It is of considerable interest whether the time gaps between successive episodes of hospitalization tend to increase with the number of episodes. Earlier, Kessing et al. (1999) considered models like (3.1) with some additional covariates for data similar to these here, collected over 1971-1993. Kvist et al. (2010) provide a discussion of the current study along with a statistical analysis aiming to provide insight into aspects of trend regarding recurrent hospitalizations. We report here on some models fitted to investigate additional aspects of trend in these data.

We restrict attention to 10,523 individuals with a discharge from a first hospitalization due to affective disorder between January 1, 1994 and December 31, 1999. Among these individuals, 1106 were diagnosed as bipolar at the time of first admission and an additional 1295 were classified as bipolar later during the course of follow-up. Follow-up was terminated at December 31, 1999 or upon diagnosis of schizophrenia or an organic disorder, and the hospitalization process itself was terminated by death. There was a total of $6,498$ readmissions and the number of readmissions per patient ranged from 0 to 89 (mean = 0.62, SD = 1.72).

Figure 3 displays a timeline for a hypothetical individual where $\mathcal{B}$ denotes the calendar time of birth, and $\mathcal{A}_k$ and $\mathcal{D}_k$ denote the calendar

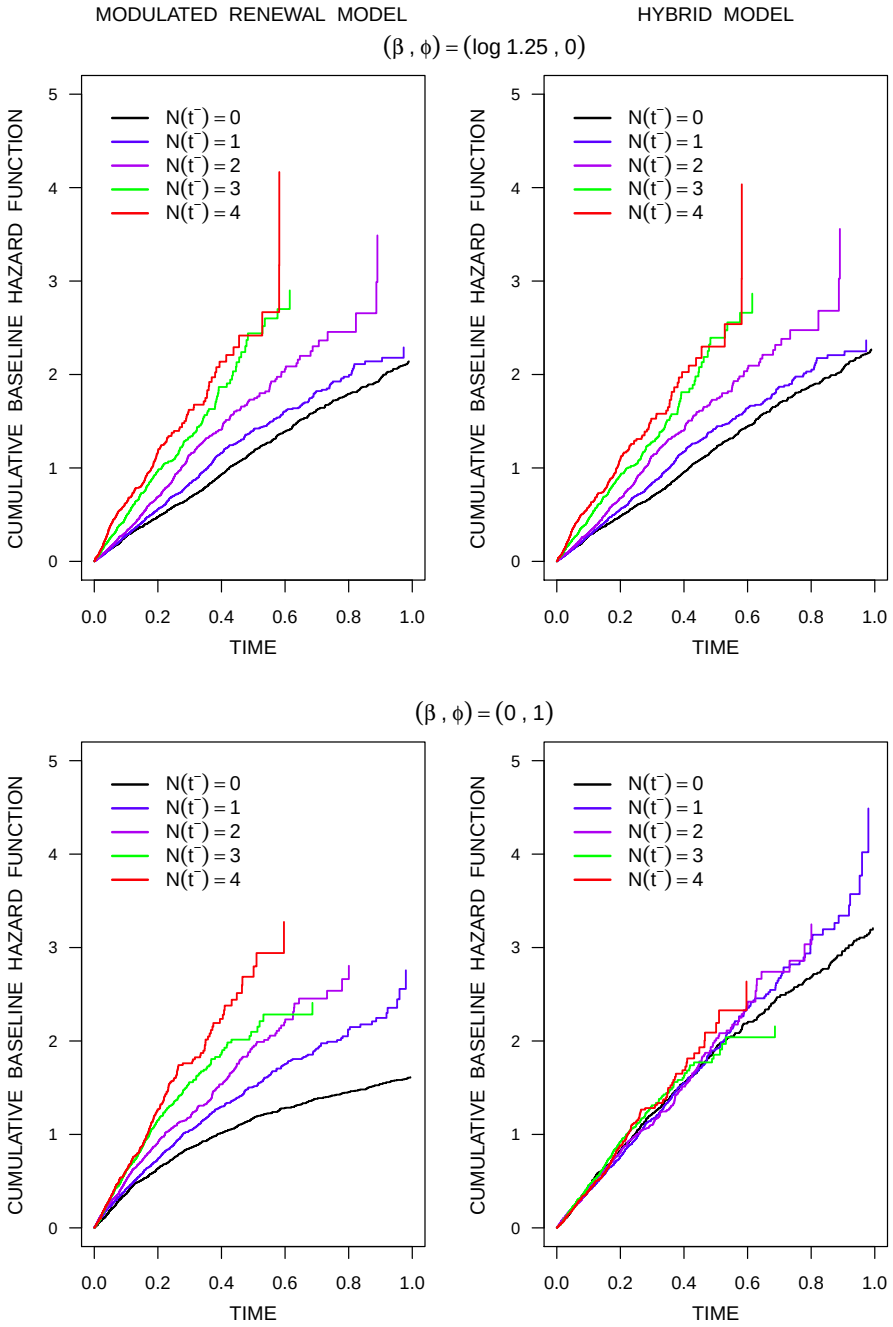

Figure 2: Estimated cumulative baseline hazard functions $\widehat{H}_n(w)$ for modulated renewal and hybrid models fitted to simulated datasets with $(\beta, \phi) = (\log 1.25, 0)$ (top row) and $(\beta, \phi) = (0, 1)$ (bottom row) ; $m = 1000$

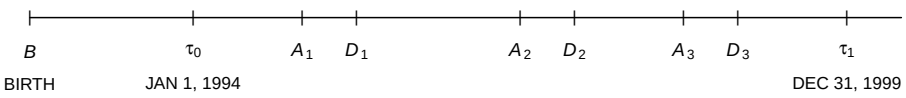

Figure 3: Timeline diagram for a hypothetical individual with recurrent hospitalizations

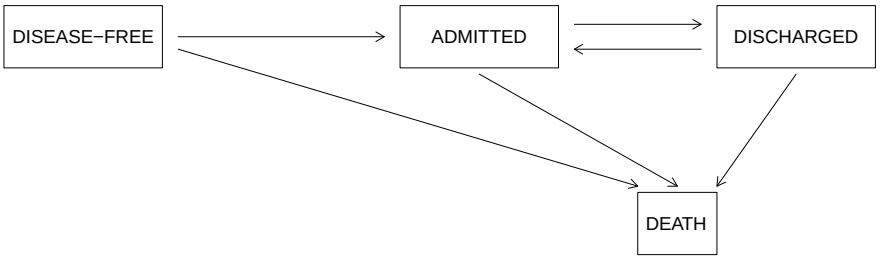

Figure 4: Multistate diagram for disease onset, recurrent hospitalization and death

times of admission and discharge from the kth hospitalization. We use script letters to indicate measurement on calendar time. We let $(\tau_0, \tau_1]$ denote the calendar interval over which events are observed. To be included in the sample, individuals must have had a first hospitalization over the calendar interval $(\tau_0, \tau_1]$, and so $L = \tau_0 - \mathcal{B}$ is the left truncation time for the age of disease onset (taken to be the age at first hospitalization). The disease onset and recurrent hospitalization process can be characterized by the multistate diagram in Figure 4 where we distinguish the first admission from subsequent admissions by having a disease-free state which is occupied prior to disease onset.

We let s index time measured in terms of age for an individual, and use nonscripted letters to denote times measured in age; so $A_k = \mathcal{A}_k - \mathcal{B}$, $D_k = \mathcal{D}_k - \mathcal{B}$, etc. Let D denote the age of death, C denote the age at end of followup, $X = \min(D, C)$ and $\delta(s) = I(s \leq X)$. We let $N(s) = \sum_{k=1}^{\infty} I(s \geq A_k)$ count the number of admissions over the age interval $(0, s]$ and $Y_k(s) = I(N(s) = k - 1)$ indicate that the individual has had their $(k-1)$st but not their kth hospitalization by age s. We will also use notation $dN(s) = N(s) - N(s^-)$ to indicate that an admission occurs at time s; a similar notation is used for other counting processes. It is helpful to define counting processes for each hospitalization and so

we let $N_k(s) = I(A_k \leq s)$ indicate that the kth admission has occurred by age s and $N_k^D(s) = I(D \leq s, N(s) = k)$ indicate death at age s with a history of k prior admissions. If $C(s)$ indicates that the subject is not in hospital at age s, then $\bar{Y}_k(s) = \delta(s)C(s)Y_k(s)$ indicates they are under observation and alive at age s and at risk for their kth hospitalization.

The selection condition that subjects must have their first admission for affective disorder following January 1, 1994 leads to left truncation and $L = \tau_0 - \mathcal{B}$ is the left truncation time (age) for the first admission; we let $L(s) = I(L \leq s)$ indicate that age s is beyond the left truncation time for a subject.

We begin with descriptive graphical analyses of the admission rates as a function of age and admission history. Let $Q_k(s)$ denote the cumulative intensity for the kth admission under a working Markov model in which the admission intensity is a function of age alone, $k = 2, \ldots,$. Here $Q_k'(s)$ is the rate of admission at age s for a kth hospitalization episode, given that a person is at risk. If we add subscripts to indicate individual i, this may be estimated nonparametrically by

$$d\widehat{Q}_k(s) = \frac{\sum_{i=1}^m L_i(s)\bar{Y}_{ik}(s)\ dN_{ik}(s)}{\sum_{i=1}^m L_i(s)\bar{Y}_{ik}(s)} ,$$

$\widehat{Q}_k(s) = \int_0^s d\widehat{Q}_k(u)$. The corresponding nonparametric estimates of the cumulative intensities for death at age s are obtained similarly as

$$d\widehat{\Gamma}_k(s) = \frac{\sum_{i=1}^m L_i(s)\ \delta_i(s)Y_{i,k+1}(s)\ dN_{ik}^D(s)}{\sum_{i=1}^m L_i(s)\ \delta_i(s)Y_{i,k+1}(s)} ,$$

giving $\widehat{\Gamma}_k(s) = \int_0^s d\widehat{\Gamma}_k(u)$. Although these estimates are not in general directly connected to modulated renewal models like (3.1), they provide insight into hospitalization patterns.

Plots of these estimates are given in Figure 5 in the left panel for the admission process and right panel for death. The estimates of the cumulative transition rates for second and subsequent hospitalizations are increasingly steep, indicating that at any given age there is an increase in risk of hospitalization with an increased history of prior hospitalization; the slope of the cumulative rate for the second admission is quite steep among young children, in part because of the small number of individuals at risk in this age range.

The right hand panel reveals little evidence of an association between hospitalization and death since there is no apparent trend in the mortality rate by hospitalization history.

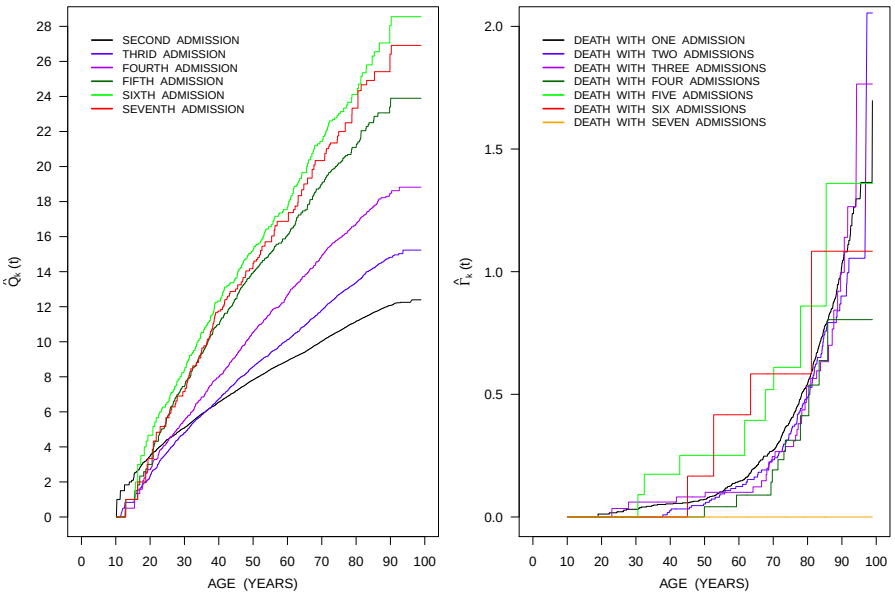

Figure 5: Plots of cumulative transition rates for re-admission and death in the study of affective disorder, stratified by the number of prior admissions with age as the basic time scale.

The robust pseudo-score test and the rank test were applied to this data using the gaps between discharge and admission for successive hospitalizations. This ignores the effect of durations of hospitalizations, but given these tend to be short relative to the times between hospitalizations this may be expected to have little impact on the inferences regarding trend. The robust pseudo-score statistic was 27.37 giving $p < 0.0001$, and the rank test with log rank scores ψ_{ij} as given in (2.6) was -2.49 with $p = 0.0128$, so both lead to rejection of the null hypothesis and a conclusion that successive gap times tend to decrease. This does not, however, provide a clear picture of the nature of the phenomenon. The much smaller p−value from the pseudo-score test may be due to the fact that it uses all individuals with $n_i \geq 1$ whereas the rank statistic only uses persons with $n_i \geq 2$.

We next consider gap time analyses with conditional modulated renewal models incorporating information on age at disease onset, calendar time effects, time since disease onset, and the cumulative number of events. To this end we specify models of the form

$$\lambda(t|\alpha_i, \mathcal{H}_i(t)) = \alpha_i h_0(B_i(t)) \exp(z_i'(t)\beta) \tag{4.1}$$

where $z_i(t)$ is a $p \times 1$ vector of time-dependent covariates, β is the corresponding $p \times 1$ vector of regression coefficients, and α_i is a random effect. We define time t in this analysis as time since disease onset, which is defined as $\mathcal{A}_{i1}$, so $t = s - O_i$ where $O_i = \mathcal{A}_{i1} - \mathcal{B}$ is the age at disease onset. The vector $z_i(t)$ can contain functions of age, calendar time, the number of previous episodes, and so on. This allows us to examine possible sources of the trend seen in Figure 5.

Let $r = 1, \ldots, 6$ denote the six years from 1994 to 1999 covered by the data, and let $\mathcal{V}_{r-1}$ denote the start of calendar year r, $r = 1, \ldots, 6$. We define $P_{ir} = I(\mathcal{V}_{r-1} \leq \mathcal{A}_{i1} < \mathcal{V}_r)$ for $r = 1, \ldots, 6$, where $\mathcal{V}_6$ refers to the end of 1999. This variable is useful for assessing whether there are trends in the hospitalization patterns with the year of disease onset. We can also define a related time-dependent covariate which indicates which period of time t is in: $P_{ir}(t) = I(\mathcal{V}_{r-1} \leq \mathcal{A}_{i1} + t < \mathcal{V}_r)$, $r = 1, \ldots, 6$. This variable enables us to examine whether there are patterns in hospitalization over the calendar time period of observation.

Interest also lies in trends related to the age at disease onset. We consider a piecewise model to allow for various trends. Let $0 = s_0 < s_1 < \cdots < s_{r-1} < s_{R_a} = \infty$ denote cut-points for age at disease onset, and let $G_{ir} = I(s_{r-1} < \mathcal{A}_{i1} - \mathcal{B}_i < s_r)$ indicate that individual i had disease onset during age interval $(s_{r-1}, s_r]$, $r = 1, \ldots, R_a$. Trends in admissions related to current age can also be examined by defining time-dependent covariates $G_{ir}(t) = I(s_{r-1} < \mathcal{A}_{i1} - \mathcal{B}_i + t < s_r)$, $r = 1, \ldots, R_a$.

One can also examine the effect of time since disease onset which is by its nature time-varying. To do this we let $O_{ir}(t) = I(b_{r-1} < t < b_r)$ where $0 = b_0 < b_1 < \cdots < b_{R_o-1} < b_{R_o} = \infty$ denote cut-points based, for example on years. Figure 6 contains a Lexis diagram to indicate how these time-dependent covariates are defined.

The results of fitting several models of the form (4.1) are reported in Tables 4 and 5. The first model (1A) reported on in Table 4 contains a covariate for sex, a categorical covariate for the cumulative number of prior admissions, which is $N_i(t^-)$ for $N_i(t^-) \leq 7$ and 8 for $N_i(t^-) \geq 8$, and a six-category time-dependent covariate for the time since disease onset. The results reveal a significantly higher rate of admission for women compared to men. We also see a highly significant trend towards increased risk of admission with increasing numbers of prior hospitalizations. There is, however, also a significant trend indicating a lower admission rate with increasing time since first admission (disease onset). There is evidence of substantial heterogeneity, as represented by the estimate of ϕ.

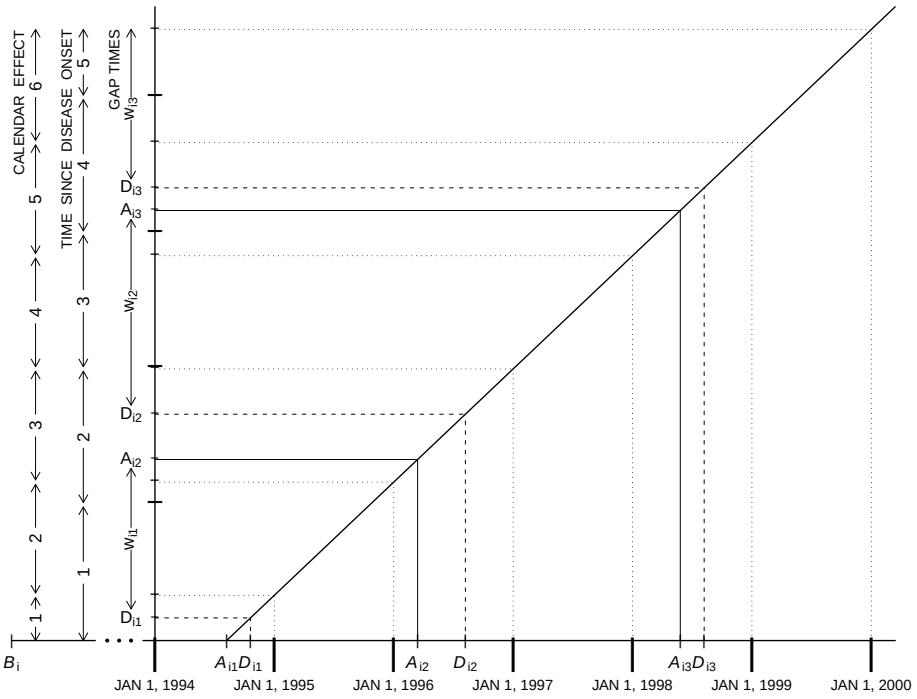

Figure 6: Diagram illustrating construction of time-dependent covariates for an individual in the analysis of hospitalizations for affective disorder

To illustrate the effect of omitting the random effect from this model, in Figure 7 we plot the estimated cumulative baseline hazards for the first few gap times based on the corresponding model excluding the random effect (left panel), as well as this model (right panel); here we stratify on the number of prior admissions. As in the simulated example, there is considerably more (spurious) evidence of a trend in the model excluding the random effect than in the model which appropriately accommodates heterogeneity.

The conclusions about the effects of these variables are unchanged when we also adjust for the age of onset (Model 2A). Here we see a significantly lower rate of admission for those individuals with later ages of onset. The third model (Model 3A) indicates that if we include period of onset but drop years since onset as a factor, the effect of prior admissions becomes inconsistent, with negative and then positive effects. This may be because the prior admissions parameters are compensating for the missing years since onset effects; models 1A and 2A indicate oppo-

site trends for years since onset (decreasing trend) and number of prior admissions (increasing trend). The final model (4A) includes an effect for both the period of onset and years since onset, and for this we find also a lower rate of re-admission for those individuals with disease onset in the latter part of the observation window, but a consistent pattern for the effect of prior admissions similar to those for models 1A and 2A. For each of these models there remains significant heterogeneity in the admission process across individuals.

Table 5 contains analogous results for models with other time-dependent covariates. In these models the age and period variables do not relate to disease onset but rather are time-dependent covariates indicating current age and current period. The age variable changes at most once during the period of observation given the wide age intervals used, but the period variable changes with each year. The conclusions are broadly similar to those of the analyses reported in Table 4 except that in this case the effect of prior admissions is consistent across all four models. This may be due to the fact that for the third model, the current age and period variables compensate for the missing years since onset, whereas the period of onset variable in Model 3A of Table 4 does not do this.

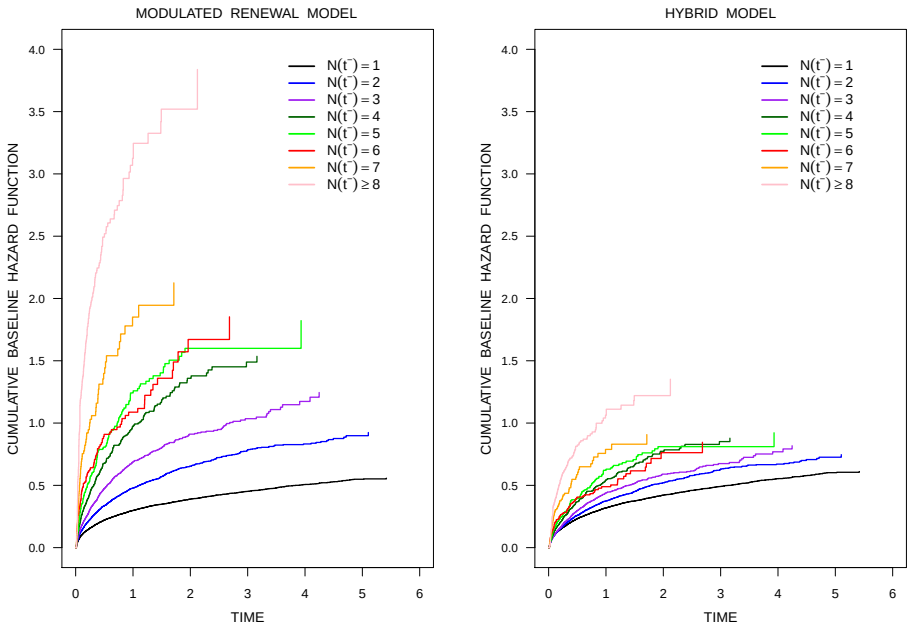

Figure 7: Estimates of cumulative baseline hazards from modulated renewal and hybrid models related to Model 1A of Table 4

Table 4: Results of intensity-based analyses of hospital re-admissions among patients with affective disorder.

	MODEL 1A			MODEL 2A			MODEL 3A			MODEL 4A		
	EST	SE	p	EST	SE	p	EST	SE	p	EST	SE	p
Sex: Female vs. Male	0.101	0.032	0.0018	0.113	0.033	0.0006	0.113	0.035	0.0011	0.111	0.033	0.0008
Prior Admissions (Ref: 1)			< 0.0001			< 0.0001			< 0.0001			< 0.0001
2	0.049	0.037	0.1933	0.012	0.038	0.7531	-0.133	0.035	0.0002	-0.004	0.038	0.9135
3	0.096	0.054	0.0755	0.042	0.054	0.4377	-0.181	0.047	0.0001	0.024	0.054	0.6629
4	0.273	0.071	0.0001	0.208	0.071	0.0033	-0.069	0.061	0.2598	0.188	0.071	0.0081
5	0.335	0.087	0.0001	0.258	0.087	0.0032	-0.063	0.078	0.4201	0.232	0.087	0.0081
6	0.294	0.111	0.0081	0.209	0.111	0.0607	-0.147	0.101	0.1471	0.189	0.111	0.0888
7	0.631	0.130	< 0.0001	0.532	0.131	< 0.0001	0.155	0.121	0.2030	0.510	0.131	< 0.0001
≥ 8	0.898	0.090	< 0.0001	0.821	0.090	< 0.0001	0.404	0.070	< 0.0001	0.802	0.091	< 0.0001
Years Since Disease Onset (Ref: $[0, 1)$)			0.0001			0.0004						< 0.0001
$[1, 2)$	-0.009	0.044	0.8407	0.005	0.044	0.9097				-0.023	0.044	0.6117
$[2, 3)$	-0.222	0.062	0.0004	-0.201	0.062	0.0013				-0.242	0.063	0.0001
$[3, 4)$	-0.207	0.076	0.0066	-0.180	0.077	0.0184				-0.229	0.077	0.0030
$[4, 5)$	-0.251	0.095	0.0080	-0.222	0.095	0.0197				-0.274	0.096	0.0044
$[5, \infty)$	-0.472	0.160	0.0031	-0.444	0.160	0.0055				-0.489	0.161	0.0024
Age of Onset (Ref: $[0, 20)$)						< 0.0001			< 0.0001			< 0.0001
$[20, 40)$				-0.218	0.096	0.0236	-0.245	0.102	0.0168	-0.234	0.097	0.0162
$[40, 60)$				-0.374	0.096	< 0.0001	-0.429	0.102	< 0.0001	-0.405	0.097	< 0.0001
$[60, 80)$				-0.454	0.096	< 0.0001	-0.518	0.102	< 0.0001	-0.492	0.098	< 0.0001
$[80, \infty)$				-0.602	0.109	< 0.0001	-0.665	0.115	< 0.0001	-0.639	0.110	< 0.0001
Period of Onset (Ref: $[1994, 1995)$)									< 0.0001			< 0.0001
$[1995, 1996)$							0.039	0.052	0.4547	0.028	0.050	0.5734
$[1996, 1997)$							-0.021	0.053	0.6971	-0.038	0.051	0.4615
$[1997, 1998)$							-0.028	0.053	0.6022	-0.056	0.052	0.2823
$[1998, 1999)$							-0.144	0.057	0.0121	-0.171	0.056	0.0022
$[1999, 2000)$							-0.369	0.073	< 0.0001	-0.375	0.072	< 0.0001
Frailty Variance	0.637			0.679			0.873			0.682		

Table 5: Results of intensity-based analyses of hospital re-admissions among patients with affective disorder with time-dependent age and period effects.

	MODEL 1B			MODEL 2B			MODEL 3B			MODEL 4B		
	EST	SE	p	EST	SE	p	EST	SE	p	EST	SE	p
Sex: Female vs. Male	0.101	0.032	0.0018	0.111	0.033	0.0007	0.111	0.033	0.0008	0.110	0.033	0.0008
Prior Admissions (Ref: 1)			< 0.0001			< 0.0001			< 0.0001			< 0.0001
2	0.049	0.037	0.1933	0.029	0.037	0.4418	0.003	0.035	0.9279	0.051	0.037	0.1760
3	0.096	0.054	0.0755	0.067	0.054	0.2135	0.024	0.047	0.6081	0.104	0.054	0.0562
4	0.273	0.071	0.0001	0.236	0.071	0.0009	0.176	0.062	0.0043	0.281	0.071	< 0.0001
5	0.335	0.087	0.0001	0.287	0.087	0.0010	0.205	0.078	0.0084	0.330	0.087	0.0002
6	0.294	0.111	0.0081	0.240	0.111	0.0306	0.156	0.102	0.1261	0.299	0.111	0.0071
7	0.631	0.130	< 0.0001	0.561	0.131	< 0.0001	0.468	0.122	0.0001	0.623	0.131	< 0.0001
≥8	0.898	0.090	< 0.0001	0.858	0.090	< 0.0001	0.762	0.072	< 0.0001	0.932	0.090	< 0.0001
Years Since Disease Onset (Ref: $[0, 1)$)			0.0001			0.0009						0.0204
$[1, 2)$	-0.009	0.044	0.8407	0.006	0.044	0.8874				0.018	0.044	0.6880
$[2, 3)$	-0.222	0.062	0.0004	-0.194	0.062	0.0019				-0.165	0.063	0.0092
$[3, 4)$	-0.207	0.076	0.0066	-0.169	0.077	0.0270				-0.104	0.078	0.1849
$[4, 5)$	-0.251	0.095	0.0080	-0.208	0.095	0.0287				-0.088	0.098	0.3676
$[5, \infty)$	-0.472	0.160	0.0031	-0.424	0.160	0.0081				-0.251	0.163	0.1236
Age (Ref: $[0, 20)$)						< 0.0001			< 0.0001			< 0.0001
$[20, 40)$				-0.159	0.105	0.1300	-0.176	0.106	0.0965	-0.172	0.105	0.1008
$[40, 60)$				-0.354	0.105	0.0008	-0.389	0.106	0.0002	-0.377	0.105	0.0003
$[60, 80)$				-0.431	0.106	< 0.0001	-0.473	0.107	< 0.0001	-0.459	0.106	< 0.0001
$[80, \infty)$				-0.497	0.114	< 0.0001	-0.540	0.115	< 0.0001	-0.525	0.114	< 0.0001
Period (Ref: $[1994, 1995)$)									< 0.0001			< 0.0001
$[1995, 1996)$							-0.021	0.071	0.7621	-0.031	0.071	0.6595
$[1996, 1997)$							-0.057	0.069	0.4119	-0.059	0.070	0.3928
$[1997, 1998)$							-0.065	0.069	0.3433	-0.059	0.069	0.3966
$[1998, 1999)$							-0.128	0.069	0.0660	-0.118	0.070	0.0903
$[1999, 2000)$							-0.269	0.070	0.0001	-0.254	0.071	0.0003
Frailty Variance	0.637			0.652			0.682			0.617		

The results in Tables 4 and 5 indicate that the gaps between admissions tend to decrease with the number of prior admissions, but that this is offset by a tendency for gaps between admissions to increase with calendar time, age and years since disease onset. To examine the differences between individuals with unipolar and bipolar affective disorder, we fitted models analogous to those in Tables 4 and 5 but including a binary time-dependent variable indicating a diagnosis of bipolar disorder. These models reveal a significantly higher re-admission rate following a bipolar diagnosis; for example, adding this variable to Model 1A gives $RR = 1.29$ (95% CI: 1.18, 1.41; $p < 0.0001$). The evidence of higher risk of re-admission following a diagnosis of bipolar disorder is present in all fitted models. Similar conclusions are reached from sensitivity analyses in which the same models were fit but on a data set excluding an individual with 89 hospital admissions, but these models generally had lower estimates of the frailty variance parameter.

A marginal analysis of the effect of prior admissions, as in Figure 5, is biased towards persons with earlier onset and age of onset, which gives longer periods of followup. Thus, the effects of prior admissions seen in Figure 5 are larger than the relative risks seen in Tables 4 and 5.

5 Discussion

There is a wide variety of models that can be fitted to recurrent event data, and different models lead to different characterizations of trends. It is important that models which allow an assessment of different aspects of trend be considered. As illustrated in Section 4, there may be several types of trend present in data and it can be challenging to distinguish them. We should add that model diagnostics are important but this is challenging when heterogeneity is present and when there is a large number of individuals but small numbers of events per individual. To a large extent, model checking will depend on comparisons with expanded models in which additional structure is present. We have illustrated this to some degree in the application.

An important point that we have not dealt with is selection effects. These arise when patients are included or withdrawn from studies for reasons related to their event processes. In the context of the psychiatry study, patients may die during the course of follow-up and if there is an association between death and the admission process the individuals remaining on study may be at lower risk of admission and re-admission.

This would help explain apparent lower re-admission rates with increasing time since disease onset or even calendar/period effects. The plot on the right-hand side of Figure 5 does not suggest a strong relation between admission and death, but this is examined in the context of a marginal Markov model with the time-scale based on age.

The model (3.1) has a parsimonious way of characterizing dependence on the event history for a given individual. Gjessing et al. (2010) discuss such models and point out that because of the exponential term $\exp(\beta N_i(t^-))$, the model "explodes" if t is allowed to become large. To avoid this problem, we adopt useful closely related models in the simulation studies and application in which the event count in the linear predictor is capped at some specified value. When this upper limit is reasonably large, trends of this sort can still be effectively studied.

Acknowledgements

Data on psychiatric hospitalizations were kindly provided by Professor Lars Vedel Kessing and Professor Per Kragh Andersen who have used the data in prior publications on recurrent events. This research was supported by the Natural Sciences and Engineering Research Council of Canada (RJC and JFL) and the Canadian Institutes for Health Research (RJC). Richard Cook is a Canada Research Chair in Statistical Methods for Health Research. We thank Ker-Ai Lee for programming assistance in the simulation studies.

Appendix: Derivation of (3.3)

Assume the history $\mathcal{H}_i(t)$ consists of $n-1$ events, at times $t_{i1}, \ldots, t_{i,n-1}$. Writing "Pr" to denote a probability density for convenience and using (2.11), we find for model (3.1) that

$$\begin{aligned}
\Pr(\mathcal{H}_i(t), \alpha_i) &= \left\{ \prod_{j=1}^{n-1} \alpha_i h_0(B_i(t_{ij})) e^{\beta(j-1)} \right\} \\
&\quad \times \exp\left\{ -\alpha_i \int_0^t h_0(B_i(u)) e^{\beta N_i(u^-)} du \right\} g(\alpha_i) \\
&= \alpha_i^{n-1} \left\{ \prod_{j=1}^{n-1} h_0(w_{ij}) e^{\beta(j-1)} \right\} \times
\end{aligned}$$

$$\exp\left\{-\alpha_i\left[\sum_{j=1}^{n-1} H_0(w_{ij})e^{\beta(j-1)} + H_0(B_i(t))e^{\beta(n-1)}\right]\right\} g(\alpha_i)$$

Noting that

$$E(\alpha_i|\mathcal{H}_i(t)) = \frac{\int_0^\infty \alpha_i \Pr(\mathcal{H}_i(t), \alpha_i) d\alpha_i}{\int_0^\infty \Pr(\mathcal{H}_i(t), \alpha_i) d\alpha_i}$$

and that

$$g(\alpha_i) = \frac{\phi^{\phi^{-1}-1}\exp(-\alpha_i\phi^{-1})}{\phi^{\phi^{-1}}\Gamma(\phi^{-1})}\ ,$$

we find (3.3) after a little algebra.

References

Aalen, O. O., Borgan, O., and Gjessing, H. K. (2008), Survival and Event History Analysis: A Point Process Point of View. New York: Springer.

Agustin, Z. and Peña, E. A. (2005), A basis approach to goodness-of-fit testing in recurrent event models. Journal of Statistical Planning and Inference, **133**, 285–303.

Ascher, H. and Feingold, H. (1984), Repairable Systems Reliability: Modeling, Inference, Misconceptions and Their Causes. New York: Marcel Dekker.

Bain, L. J., Engelhardt, M., and Wright, F. T. (1985), Test for an increasing trend in the intensity of a Poisson process: A power study. Journal of the American Statistical Association, **80**, 419–422.

Barnabani, M. (2008), Marginal distributions of maximum-likeilhood estimator when one or two components of the true parameter are on the boundary of the parameter space. Working Paper 2008/11, Universitá degli Studi di Firenze.

Cohen, A. and Sackrowitz, H. B. (1993), Evaluating tests for increasing intensity of a Poisson process. Technometrics, **35**, 446–448.

Cook, R. J., Lawless, J. F., and Nadeau, J. C. (1996), Robust tests for treatment comparisons based on recurrent event responses. Biometrics, **52**, 732-739.

Cook, R. J. and Lawless, J. F. (2007), The Statistical Analysis of Recurrent Events. New York: Springer.

Cox, D. R. and Lewis, P. A. W. (1966), The Statistical Analysis of Series of Events. London: Chapman and Hall.

Dabrowska, D. M., Sun, G. and Horowitz, M. M. (1994), Cox regression in a Markov renewal model: an application to the analysis of bone marrow transplant data. Journal of the American Statistical Association, **89**, 867–877.

Gjessing, H. K., Roysland K., Pena, E. A., and Aalen, O. O. (2010), Recurrent events and the exploding Cox model. Lifetime Data Analysis, **16**, 525–546.

Hájek, J. and Sidák, Z. (1967), Theory of Rank Tests. New York: Academic Press.

Kalbfleisch, J. D. and Prentice, R. L. (2002), The Statistical Analysis of Failure Time Data, 2nd edition. New York: John Wiley and Sons.

Kessing, L. V., Olsen, E. W., and Andersen, P. K. (1999), Recurrence in affective disorder: analyses with frailty models. American Journal of Epidemiology, **149**, 404-411.

Kessing, L. V., Hansen M. G. and Andersen P. K. (2004), Course of illness in depressive and bipolar disorders. British Journal of Psychiatry, **185**, 372–377.

Kvaløy, J. T. and Lindqvist, B. H. (2003), A class of tests for renewal process versus monotonic and nonmonotonic trend in repairable systems data. In Mathematical and Statistical Methods in Reliability, 401–414. Eds. B.H. Lindqvist and K.A. Doksum. World Scientific, River Edge, NJ.

Kvist, K., Andersen, P. K., Angst, J., and Kessing, L. V. (2008), Repeated events and total time on test. Statistics in Medicine, **27**, 3817–3832.

Kvist, K., Andersen, P. K., Angst, J., and Kessing, L. V. (2010), Event dependent sampling of recurrent events. Lifetime Data Analysis, **16**, 580–598.

Lawless, J. F. and Nadeau, J. C. (1995), Nonparametric estimation of cumulative mean functions for recurrent events. Technometrics, **37**, 158–168

Lawless, J. F., Wigg, M. B., Tuli, S., Drake, J. M., and Lamberti-Pasculli, M. (2001), Analysis of repeated failures or durations, with application to shunt failures for patients with pediatric hydrocephalus. Applied Statistics, **50**, 449–465.

Lawless, J. F., Cigsar, C. and Cook, R. J. (2012), Testing for monotone trend in recurrent event processes, Technometrics, **54**, 147–158.

Lewis, P. A. W. and Robinson, D. W. (1974), Testing for a monotone trend in a modulated renewal process. In Reliability and Biometry. Statistical Analysis of Life-length. Eds. F. Proschan and R.J. Serfling. SIAM, Philadelphia.

Loh, J. M. (2010), Bootstrapping an inhomogeneous point process. Journal of Statistical Planning and Inference, **140**, 734–749.

Pena, E. A. (2006), Dynamic modelling and statistical analysis of event times. Statistical Science, **21**, 487–500.

Self, S. G. and Liang, K.-Y. (1987), Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. Journal of the American Statistical Association, **82**, 605–610.

Wang, M.-C. and Chen, Y.-Q. (2000), Nonparametric and semiparametric trend analysis for stratified recurrence times. Biometrics, **56**, 789–794.

White, H. and Granger, C. W. J. (2011), Consideration of trends in time series. Journal of Time Series Econometrics Vol. 3: Iss. 1, Article 2.