The Annals of Applied Statistics
2012, Vol. 6, No. 4, 1641–1663
DOI: 10.1214/12-AOAS560

ASSESSING TRANSIENT CARRYOVER EFFECTS IN RECURRENT EVENT PROCESSES, WITH APPLICATION TO CHRONIC HEALTH CONDITIONS

BY CANDEMIR ÇIĞŞAR[1] AND JERALD F. LAWLESS[2]

Women's College Hospital and Princess Margaret Hospital, and University of Waterloo

In some settings involving recurrent events, the occurrence of one event may produce a temporary increase in the event intensity; we refer to this phenomenon as a transient carryover effect. This paper provides models and tests for carryover effect. Motivation for our work comes from events associated with chronic health conditions, and we consider two studies involving asthma attacks in children in some detail. We consider how carryover effects can be modeled and assessed, and note some difficulties in the context of heterogeneous groups of individuals. We give a simple intuitive test for no carryover effect and examine its properties. In addition, we demonstrate the need for detailed modeling in trying to deconstruct the dynamics of recurrent events.

1. Introduction. Recurrent events experienced by individuals, units or systems occur in many fields [Cook and Lawless (2007)]. For example, repeated failures can occur for equipment or for software systems [Ascher and Feingold (1984), Baker (2001), Lindqvist (2006)]. In medical contexts, individuals may experience multiple episodes of hospitalization, recurrent infections or children may suffer repeated attacks of asthma [Duchateau et al. (2003)]. Models for recurrent events are discussed in books such as Cox and Isham (1980) and Daley and Vere-Jones (2003). Cox and Lewis (1966), Karr (1991) and Cook and Lawless (2007) discuss related methods of analysis.

In certain settings an event intensity is temporarily increased (or in some cases, decreased) after some condition or event occurs. Such transient effects may be due to factors that are either external or internal to the individuals or systems in question. Transient effects due to external factors have received considerable recent attention. For example, Farrington and Whitaker (2006) and Farrington and Hocine (2010) have examined potential adverse health effects following administration of the mumps, measles and rubella (MMR) vaccine to children. Farrington, Whitaker and Hocine (2009) consider adverse effects related to drug treatments.

Received July 2011; revised April 2012.

[1]Supported by Cancer Care Ontario and the Ontario Institute for Cancer Research (through funding provided by the Ministry of Health and Long-Term Care and the Ministry of Research and Innovation of the Government of Ontario).

[2]Supported by the Natural Sciences and Engineering Research Council of Canada.

Key words and phrases. Internal covariates, Poisson processes, renewal processes, score tests.

Our focus in this paper is on internal factors. These are not usually observable, and evidence for their existence is sought by examining whether event intensities are temporarily increased soon after an event occurs. We call such effects carryover effects, which is also a term used to describe transient effects due to external factors such as vaccinations or residual effects of drugs [Cook and Lawless (2007), Section 3.8.2]. This phenomenon has also often been discussed for hardware or software systems, where repairs or modifications undertaken to deal with a failure may not resolve the problem or may even create new problems [e.g., Baker (1996, 2001), Peña (2006)].

The motivation for the present paper is from attempts to identify potential carryover effects related to events occurring in subjects with chronic medical conditions. Such effects are inherently difficult to assess because of complex factors that may influence event occurrence. These include unobservable covariates that can produce wide heterogeneity in event rates across individuals and the presence of temporal trends that may be related to the age of a process or to external factors such as seasonal effects. In addition, clinical events are often related to unobservable processes concerning a person's health and fluctuations in such processes can produce clustering of events. This paper is motivated specifically by studies of adverse events in children. Two studies that we consider here involve randomized treatment trials for the prevention of asthma attacks; a third study that will be discussed more briefly later in the paper involves failures associated with shunts which are used to drain excess cerebrospinal fluid in children with hydrocephalus [Tuli et al. (2000)].

In the first asthma prevention trial, infants who were considered at high risk for asthma were randomized at 6 months of age to receive either a placebo or drug treatment [Duchateau et al. (2003)]. They then were followed for 18 months, and occurrences of any asthma attacks (according to specified symptoms) were recorded. In addition to the assessment of any drug effect, other points of interest are the evolution of the asthma recurrent event rate over time and how the occurrence of an event influences the event rate [Duchateau et al. (2003), page 356]. In the second study [Verona et al. (2003)], children aged 4–11 years were randomized to receive either 200 or 400 μg per day of fluticasone propionate (FP) for the prevent of asthma exacerbations. The original protocol called for 3 months of follow-up per child, but this was later amended to 12 months. Most of the exacerbations in question were defined as "moderate;" these were defined as occurring if a child experienced a period of two consecutive days on which either (i) their morning percentage predicted expiratory flow (PEF, a measure of lung function) fell more than 20% below their baseline value measured at randomization, or (ii) they had an increase in inhaler (β_2-agonist) usage.

In each of these studies we will examine whether there is an indication that individuals are temporarily at a higher risk of a new event (exacerbation) following the resolution of a previous exacerbation. Insights into this can affect strategies for the prevention and treatment of exacerbations. As an illustration we show a simple

synopsis of data from the first asthma trial, in which 119 children were randomized to the placebo control group and 113 were randomized to the treatment group. The total numbers of asthma attacks were 483 (control group) and 336 (treatment group). The total observed and expected (calculated under a hypothesis of no carryover effect, as described in Section 5.1) number of attacks which occurred within two weeks of the preceding attack are as follows:

$$\begin{aligned} &\textit{Control group}: \quad \textit{Observed} = 121, \qquad \textit{Expected} = 80.3, \\ &\textit{Treatment group}: \quad \textit{Observed} = 76, \qquad \textit{Expected} = 40.5. \end{aligned}$$

The data show an excessive number of events soon after the preceding event.

The presence of a carryover effect can be assessed fairly readily in single systems which experience large numbers of events [e.g., Baker (2001)]. However, in medical contexts we typically have a large number of individuals, each with a small number of events. The purpose of this paper is to discuss models through which the presence of a carryover effect can be assessed in settings involving multiple heterogeneous individuals, as seen in the preceding studies. We make three novel contributions. First, we show that internal carryover effects can be difficult to distinguish from subject heterogeneity in settings where the average number of events per subject is fairly small. Second, we show that the data often have limited information about the duration of an effect, so reliance on background information is crucial. Finally, we provide tests for no carryover effect which are simple to interpret and reasonably robust.

In Section 2 we consider models for transient carryover effects, discuss their connection to the concept of event clustering, and show how heterogeneity makes the assessment of transient effects more difficult. Section 3 considers some simple tests and Section 4 presents simulation results on their properties. Section 5 examines the studies on asthma in infants. Section 6 contains concluding remarks and discusses a study of cerebrospinal fluid shunt failures in pediatric patients. In the interests of exposition, some technical derivations are placed in the Appendix.

2. Models for carryover effects. We use standard notation for recurrent events. We assume that an individual process is observed over time interval $[0, \tau]$, and let $N(t)$ denote the number of events in $(0, t]$. The history of events over $[0, t)$ is denoted by $\mathcal{H}(t)$ and the event intensity function [Cook and Lawless (2007), page 10] is given by

$$\lambda(t|\mathcal{H}(t)) = \lim_{\Delta t \downarrow 0} \frac{\Pr\{N(t + \Delta t^-) - N(t^-) = 1|\mathcal{H}(t)\}}{\Delta t}. \tag{2.1}$$

The intensity fully specifies continuous time processes where at most one event can occur at a given time. The times of events are denoted $T_1 < T_2 < \cdots$, and $B(t) = t - T_{N(t^-)}$ is the elapsed time since the most recent event prior to t. Familiar

models include Poisson processes, where $\lambda(t|\mathcal{H}(t)) = \rho(t)$ for some function ρ, and renewal processes, where $\lambda(t|\mathcal{H}(t)) = h(B(t))$ for some function h.

Carryover effects can be modeled in a number of ways. A model that is very useful when events may display time trends is a modulated Poisson process. In this case, (2.1) takes the form

$$\lambda(t|\mathcal{H}(t)) = \rho_0(t)\exp(\beta' Z(t)), \tag{2.2}$$

where $Z(t)$ is a $q \times 1$ vector of time-varying covariates that is allowed to contain functions of the event history $\mathcal{H}(t)$ as well as external covariates. More specifically, we can consider models for which $Z(t)$ includes terms that are zero except for a limited time period following the occurrence of an event; such terms specify the carryover effects. A simple but very useful model is one where $Z(t)$ in (2.2) takes the form

$$Z(t) = I\big(N(t^-) > 0\big)I\big(B(t) \le \Delta\big), \tag{2.3}$$

where $\Delta > 0$ is a specified value. In that case the intensity function following an event temporarily changes from $\rho_0(t)$ to $e^{\beta}\rho_0(t)$. Tests of the null hypothesis $H_0 : \beta = 0$, developed below, provide simple and intuitive tests of no carryover effect.

Other similar models with carryover effects can also be specified. For example, a model (2.2) with $Z(t) = I(N(t^-) > 0)\exp(-\gamma B(t))$ or an additive linear self-exciting process [Cox and Isham (1980), Section 3.3; Ogata (1983)] with $\lambda(t|\mathcal{H}(t)) = \rho_0(t) + \beta\sum_{j=1}^{N(t^-)} e^{-\gamma(t-t_j)}$ also produces transient effects following events, while allowing possible time trends as in (2.2). Such models are more difficult to handle than (2.2) and (2.3), and do not impose a time limit on the duration of an effect, but have been found useful in areas such as seismology [Ogata (1983)].

There is a close connection between what we term carryover effects and cluster processes [Cox and Isham (1980), Section 3.4]. In a cluster process the events occur in clusters, or groups of events that are close together in time. Carryover effects in essence produce a type of clustering, and models such as (2.3) or the linear self-exciting process can be viewed as cluster processes in which each new event produces a subprocess going forward in time, with a decreasing rate function [e.g., Cox and Isham (1980), pages 69, 77]. On the basis of observed events alone, it is impossible to say what produces observed clustering, and we must rely on context-specific background to suggest plausible mechanisms. We view internal carryover effects as arising when a "remedy" for an adverse effect is unsuccessful or partially successful, and consider models that facilitate interpretation within that framework. Many models for cluster processes are harder to handle [e.g., Cox and Isham (1980), Section 3.4; Xie, Sun and Naus (2009)], especially when the rate of events is not stationary, and there is heterogeneity. Standard clustering models do not address these points. Our models are straightforward to handle and provide insight, but as always, models should be checked, and other approaches may be

needed in some situations. We note as well that although we focus on the case where the intensity temporarily increases following an event, in some contexts it could decrease, with β in (2.2) being negative in that case.

Another way to consider carryover effects is through the distribution of gap times $W_j = T_j - T_{j-1}$ (with $T_0 = 0$) between successive events. Gap time models [Cook and Lawless (2007), Chapter 4] are particularly useful in settings where an adverse event results in some corrective action which ideally returns an individual to a "good as new" state [e.g., Peña (2006)]. Gap time models in which the times between successive events have distributions with substantial mass near zero could be considered as suggesting a carryover effect [e.g., see Baker (2001), Lindqvist (2006), Peña (2006)]. They contain more parameters and are more difficult to handle than (2.2) and (2.3), and do not accommodate calendar time trends as readily, but are often useful. In the special stationary case where $\rho_0(t)$ in (2.2) is a constant α, the model with $Z(t)$ given by (2.3) is a delayed renewal process where the times W_j $(j = 2, 3, \ldots)$ between successive events are independent random variables with a hazard function of the form $h(w) = \alpha e^{\beta} I(w \le \Delta) + \alpha I(w > \Delta)$.

In applications involving multiple systems or individuals, heterogeneity is often apparent [e.g., Lawless (1987), Baker (2001), Lindqvist (2006), Cook and Lawless (2007), Section 3.5]. For example, individual processes may be (approximately) Poisson, but their rate functions may vary. Such variation is typically due to unmeasured differences in the individuals or the environment in which the processes operate. It is imperative to consider the possibility of heterogeneity because, as we show below, it can create an appearance of a carryover effect when no such exists.

The simplest and most useful extension of modulated Poisson process models (2.2) to include heterogeneity is where independent processes $i = 1, \ldots, m$ have rate functions

$$\rho_i(t|\mathcal{H}_i(t)) = \alpha_i \rho_0(t) \exp(\beta Z_i(t)), \tag{2.4}$$

where $\alpha_1, \ldots, \alpha_m$ are positive-valued variables. Models for which the α_i are fixed parameters can be problematic because the α_i cannot be estimated consistently. An alternative is to assume the α_i are independent and identically distributed random effects with some distribution function $G(\alpha; \phi)$, where ϕ is a vector of parameters [Cook and Lawless (2007), Section 3.5], and we consider this for most analyses.

We now show why heterogeneity that is not taken into account can misleadingly suggest a carryover effect. Suppose for illustration that the model (2.4) with $\beta = 0$ and α_i following a gamma distribution describes a situation. Without loss of generality, we take the α_i to have mean 1 and variance ϕ, and then [Cook and Lawless (2007), page 79] we find that the intensity function for the process with the unobservable α_i integrated out is

$$\begin{aligned} \lambda_i(t|\mathcal{H}_i(t)) &= E(\alpha_i|\mathcal{H}_i(t))\rho_0(t) \\ &= \left\{ \frac{\phi^{-1} + N_i(t^-)}{\phi^{-1} + \int_0^t \rho_0(u)\,du} \right\} \rho_0(t). \end{aligned} \tag{2.5}$$

Note that when an event occurs, the numerator term in brackets in (2.5) increases by one, thus increasing the intensity. As t increases up to the next event, the denominator in brackets increases, so the overall effect is that the intensity increases immediately after an event occurs and then decreases. This is the type of behavior we associate with a carryover effect. The larger the degree of heterogeneity across the individuals (i.e., the larger ϕ is), the larger is the increase following an event. As t becomes arbitrarily large, the term in brackets converges in probability to α_i so the appearance of a carryover effect is mainly in the earlier events. However, failure to incorporate heterogeneity in models can produce spurious evidence of an effect. To demonstrate, we ran a small simulation by generating 1000 realizations of a random effect model without carryover effects; we used model (2.4) where the α_i have a gamma distribution with mean 1 and variance ϕ and parameters $\rho_0(t) = \gamma$ and $\beta = 0$ (no carryover effect). We considered eight scenarios with various combinations of γ, ϕ and m ($\gamma = 2, 5$, $\phi = 0.2, 0.5$ and $m = 100, 500$). Observation periods were $(0, \tau_i)$ and the τ_i times were generated from a uniform distribution over $(0.8, 1.2)$. For $\tau_i = 1$ the expected number of events per individual is 2 or 5 when $\gamma = 2$ or 5, respectively. For each sample we obtained the maximum likelihood estimates of parameters and their standard errors in the carryover effect model (2.2) with (2.3), without incorporating heterogeneity. We found that $\hat{\beta}$ was positively biased across the 1000 simulations for each scenario, with mean to standard deviation ratios varying from 0.7 to over 10. Correspondingly, tests of the null hypothesis $H_0 : \beta = 0$ incorrectly reject H_0 with high probability. Using the same data sets, we also fitted the carryover effect model with random effects (2.4) with $\rho_0(t) = \gamma$, and in this case the means of $\hat{\beta}$ were close to zero for all scenarios. The results can be found in the supplementary material [Çıığşar and Lawless (2012)].

We remark also that heterogeneity is to some extent confounded with a carryover effect even with a proper model specification. With (2.4), for example, and gamma distributed α_i, the intensity function is

$$\lambda_i(t|\mathcal{H}_i(t)) = \left\{ \frac{\phi^{-1} + N_i(t^-)}{\phi^{-1} + \int_0^t \rho_0(u) e^{\beta Z_i(u)}\, du} \right\} \rho_0(t) e^{\beta Z_i(t)}.$$

As t becomes large the term in brackets once again converges in probability to α_i, so a carryover effect represented by $\beta \neq 0$ can be readily assessed. When t and $N_i(t^-)$ are small, however, the carryover effect and the expression in brackets can both produce substantial temporary increases in the event intensity. In many of the applications we consider, there are many individuals but relatively few events for most individuals, and therefore a process of careful modeling and model-checking is warranted. Next, we consider some tests of no carryover effect based on (2.4). These are reasonably robust and have a simple interpretation in terms of the observed data.

3. Tests based on Poisson processes. We consider tests of no carryover effect based on the Poisson model (2.4), and testing that $\beta = 0$. This can be done either using a parametric model for $\rho_0(t)$ or by using a nonparametric specification, in which case (2.4) is a modulated Andersen–Gill model with frailty [Cook and Lawless (2007), page 81]. We describe the parametric setting in detail, so as to show the intuitive form of the test statistics, and then discuss the semiparametric case. The tests use a specified value for Δ in (2.3) and (2.4). This is consistent with common practice and the resulting tests have the nice form of a difference between observed and expected numbers of events in the window of length Δ following an event. However, we later consider the effects of misspecifying Δ and in Section 5 we consider estimation of Δ.

3.1. *Fixed effects model.* We consider first the fixed effects model (2.4), for which the α_i are treated as unknown parameters. This can be useful when the number of individual processes m is small but there are many events per process. The follow-up (censoring) times τ_i throughout the paper are assumed to be stopping times [Cook and Lawless (2007), page 48]. The follow-up times are therefore allowed to be random and to depend on previous event history. In this case, data on m independent processes give the log likelihood function

$$\ell(\alpha, \gamma, \beta) = \sum_{i=1}^{m} \left\{ n_i \log \alpha_i + \sum_{j=1}^{n_i} [\log \rho_0(t_{ij}; \gamma) + \beta Z_i(t_{ij})] - \alpha_i R_i(\gamma, \beta) \right\}, \tag{3.1}$$

where $\alpha = (\alpha_1, \ldots, \alpha_m)'$ and

$$R_i(\gamma, \beta) = \int_0^{\tau_i} \rho_0(t; \gamma) e^{\beta Z_i(t)}\, dt. \tag{3.2}$$

For given γ and β, (3.1) is maximized by $\tilde{\alpha}_i(\gamma, \beta) = n_i / R_i(\gamma, \beta)$, and substitution of this into (3.1) gives the profile log likelihood for γ and β as a constant plus

$$\ell_p(\gamma, \beta) = \sum_{i=1}^{m} \left\{ \sum_{j=1}^{n_i} [\log \rho_0(t_{ij}; \gamma) + \beta Z_i(t_{ij})] - n_i \log R_i(\gamma, \beta) \right\}. \tag{3.3}$$

A likelihood ratio test of $H_0 : \beta = 0$ requires estimates $\hat{\gamma}, \hat{\beta}$ that maximize (3.3) and the estimate $\tilde{\gamma}$ that maximizes $\ell_p(\gamma, 0)$; the estimates are found easily by general optimization software.

A score test can be based on $U_\beta(\tilde{\gamma}, 0)$, where $U_\beta(\gamma, \beta) = \partial \ell_p(\gamma, \beta)/\partial \beta$. The standardized score statistic

$$\begin{aligned} U_\beta(\tilde{\gamma}, 0) &= \sum_{i=1}^{m} \left\{ \sum_{j=1}^{n_i} Z_i(t_{ij}) - \frac{n_i \int_0^{\tau_i} Z_i(t) \rho_0(t; \tilde{\gamma})\, dt}{\int_0^{\tau_i} \rho_0(t; \tilde{\gamma})\, dt} \right\} \\ &= \mathrm{Obs}(\Delta) - \mathrm{Exp}(\Delta), \end{aligned}$$

where $\mathrm{Obs}(\Delta) = \sum_{i=1}^{m} \sum_{j=1}^{n_i} Z_i(t_{ij})$ is the observed number of events that occur within time Δ of a preceding event, and $\mathrm{Exp}(\Delta)$ is an estimate of the expected number of such occurrences under the hypothesis of no carryover effect. For the simple case of a homogeneous Poisson process, $\rho_0(t;\gamma)$ is one, and we find $\mathrm{Exp}(\Delta) = (n_i/\tau_i)\sum_{j=2}^{n_i+1} \min(w_{ij}, \Delta)$, where $w_{ij} = t_{ij} - t_{i,j-1}$ $(j = 1, \ldots, n_i)$ and $w_{i,n_i+1} = \tau_i - t_{in_i}$. The form "observed minus expected" for $U_\beta(\tilde{\gamma}, 0)$ is easily understood and useful. The standardized form of $U_\beta(\tilde{\gamma}, 0)$ is [Peña (1998)]

$$S = \frac{U_\beta(\tilde{\gamma}, 0)}{\widehat{\mathrm{Var}}[U_\beta(\tilde{\gamma}, 0)]^{1/2}}, \tag{3.4}$$

where $\widehat{\mathrm{Var}}[U_\beta(\tilde{\gamma}, 0)] = \tilde{I}_{\beta\beta} - \tilde{I}_{\gamma\beta}\tilde{I}_{\gamma\gamma}^{-1}\tilde{I}_{\beta\gamma}$ is obtained from the observed information matrix for β and γ based on (3.3), evaluated at $(\gamma, \beta) = (\tilde{\gamma}, 0)$.

A problem with S, and with the likelihood ratio statistic, is that if $m \to \infty$ but the τ_i are fixed, the limiting distributions are not standard normal and $\chi^2_{(1)}$, respectively, due to the fact that the α_i are not estimated consistently. The normal and χ^2 approximations may be adequate in cases where m is not too large and the numbers of events per process are fairly large, but simulations in Section 4 show they are inadequate in settings like those in Section 5. However, we can use a simulation (parametric bootstrap) approach to get p-values. Under H_0, the event times $T_{i1}, \ldots, T_{in_i}$, given $N_i(\tau_i) = n_i$, are the order statistics for a random sample of size n_i from the truncated distribution with density function [Cox and Lewis (1966), Section 3.3]

$$f_i(t;\gamma) = \frac{\rho_0(t;\gamma)}{\int_0^{\tau_i} \rho_0(s;\gamma)\,ds}, \qquad 0 \le t \le \tau_i.$$

Thus, we can generate random samples from each $f_i(t;\tilde{\gamma})$, $i = 1, \ldots, m$, and use these to obtain values of the test statistic in question. For the HPP case, $f_i(t;\gamma)$ is the uniform distribution on $[0, \tau_i]$. It should be noted that p-values obtained from this approach are conditional on the observed values $n_1, \ldots, n_m$ and so are not strictly comparable to the unconditional p-values provided by a normal or χ^2 approximation, or to p-values for the random effects model in Section 3.2.

3.2. *Random effects model.* Random effects models employ a distribution for the α_i in (2.4), which are assumed independent. We assume for discussion that the α_i have a gamma distribution with mean 1 and variance ϕ, which is a widely used model; similar developments can be given for other distributions. In this case the log likelihood function is [Cook and Lawless (2007), Section 3.5.3]

$$\begin{aligned}\ell(\gamma, \beta, \phi) = \sum_{i=1}^{m} \Bigg\{ & \sum_{j=1}^{n_i} [\log \rho_0(t_{ij};\gamma) + \beta Z_i(t_{ij})] \\ & + \log\Gamma(n_i + \phi^{-1}) - \log\Gamma(\phi^{-1}) \\ & + n_i \log\phi - (n_i + \phi^{-1}) \log[1 + \phi R_i(\gamma, \beta)] \Bigg\}.\end{aligned} \tag{3.5}$$

Likelihood ratio tests of $H_0 : \beta = 0$ require maximum likelihood estimates $\hat{\gamma}, \hat{\beta}, \hat{\phi}$ and $\tilde{\gamma}, \tilde{\phi}$ (when $\beta = 0$); these are readily obtained with general optimization software. The $R_i(\gamma, \beta)$ in (3.5) are as defined in (3.2).

Score tests of $\beta = 0$ require only $\tilde{\gamma}$ and $\tilde{\phi}$. Appendix B gives the score statistic

$$S = U_\beta(\tilde{\gamma}, 0, \tilde{\phi}) / \widehat{\mathrm{Var}}[U_\beta(\tilde{\gamma}, 0, \tilde{\phi})]^{1/2} \tag{3.6}$$

corresponding to (3.4). It is instructive to consider the numerators of (3.4) and (3.6); the numerator of (3.6) is (see Appendix B)

$$U_\beta(\tilde{\gamma}, 0, \tilde{\phi}) = \mathrm{Obs}(\Delta) - \sum_{i=1}^{m} \frac{(1 + n_i \tilde{\phi}) \int_0^{\tau_i} Z_i(t) \rho_0(t; \tilde{\gamma})\, dt}{1 + \tilde{\phi} \int_0^{\tau_i} \rho_0(t; \tilde{\gamma})\, dt}. \tag{3.7}$$

Equation (3.7) differs from the numerator of (3.4) in the calculation of the second term, $\mathrm{Exp}(\Delta)$. The fixed effects case (3.4) corresponds to the limit of (3.7) as the estimated variance $\tilde{\phi}$ of the α_i becomes arbitrarily large. Assuming that the gamma distribution for the α_i is correct, the statistic S in (3.6) is asymptotically $N(0, 1)$ as $m \to \infty$, unlike the fixed effects statistic. In Section 4 we examine the adequacy of the normal approximation in practical settings. In situations where it is inadequate we can use simulation (parametric bootstrap) to obtain p-values. In addition, the gamma distribution will never be exactly correct in practice, so we consider the performance of (3.6) under departures from the gamma in Section 4.

The Andersen–Gill model with random effects α_i [Cook and Lawless (2007), page 81] can also be used. This model places no parametric restrictions on $\rho_0(t)$ in (2.4). The R/S-Plus function `coxph` with the frailty option implements this, but some work is needed to extract *Observed–Expected* components analogous to (3.7); see Appendix B.

3.3. *Power of tests and choice of* Δ. The tests of no carryover effect in the preceding section are based on a specified value of Δ and a family of alternative hypotheses, but are robust in the sense that the tests of the null Poisson processes are, under some conditions, consistent against carryover alternatives that are not in the family represented by (2.2) and (2.3). That is, as $m \to \infty$, the probability H_0 is rejected approaches one under the alternative. We illustrate this property via simulation in Section 4, where we show that the tests in Sections 3.1 and 3.2 retain good power when Δ is misspecified and when the random effects distribution (Section 3.2) is misspecified. Simulation results (not shown) also indicate the tests retain power against alternatives where the true form of the process intensity is additive $[\lambda_0(t; \gamma) + \beta Z_i(t)]$ rather than multiplicative. The model (2.4) should, however, be checked for consistency with the data; ways to do this are discussed by Cook and Lawless (2007), Chapters 3 and 5. Carryover effects can also be tested within alternative modulated Poisson process models such as the preceding additive model. We also note that assessment of the dynamics of individual processes

with rather few events is inherently difficult, and we have found it useful to consider models based on gap times as well as Poisson models. This is illustrated in Sections 5.1 and 6.

In choosing a value of Δ, we must rely on background information that suggests how long a carryover effect might last for the process under study. Typically Δ would be fairly small relative to the average time between events across individuals. The use of specified durations Δ for carryover effects is common [e.g., Farrington and Whitaker (2006), Cook and Lawless (2007), Section 3.8.2], but there is generally some uncertainty concerning Δ and it is best to consider a few separate values. Xu et al. (2011) have recently considered uncertainty about Δ for external carryover effects but do not discuss estimation of Δ. If we treat Δ as an unknown parameter, there is often an estimability issue, because the profile likelihood for Δ supports quite a wide range of values. We examine this in Section 5, where we find that the asthma data sets do not rule out fairly large values of Δ, due partly to the fact that a carryover effect is partially confounded with heterogeneity. In addition, as Δ becomes sufficiently large all events after the first will lie within the carryover period and an effect β as in (2.2) is confounded with the scale parameter in $\rho_0(t)$.

4. Simulation studies. In this section we present the results of simulation studies conducted to assess when asymptotic normal approximations for parametric score test statistics are satisfactory, to investigate the tests' power and to evaluate their robustness with respect to model misspecification. Because of space limitations, we provide figures and tables for selected scenarios, and briefly discuss other scenarios. Additional results are given in Çığşar (2010) and in the supplementary material [Çığşar and Lawless (2012)]. We focus on cases where the null models are homogeneous Poisson processes; results for nonhomogeneous processes are similar.

We first consider the fixed effects model (2.4) where $\rho_0(t;\gamma) = \gamma$, and the hypothesis of no carryover effect is tested by using the statistic (3.4). In simulations we took $\gamma = 1$, and generated the α_i from the gamma distribution with mean 1 and variance $\phi = 0.3$. This variance represents a degree of heterogenity often seen in medical data. Similar results were obtained for $\phi = 0.6$. The α_i were generated once for each scenario, so that $\alpha_1, \ldots, \alpha_m$ are fixed across the repeated samples. To examine the asymptotic normal approximation for the null distribution of (3.4), we generated 10,000 realizations of the m homogeneous Poisson processes. In simulations reported below, scenarios with various combinations of m, τ, Δ were considered, with $m = 10$, 20, 50, 100 and $\tau_i = \tau = 10$. Results are similar if the τ_i vary, with mean equal to 10. In practice, we would be interested in small values of Δ, and we consider $\Delta = 0.0202$, 0.0513 and 0.1054. The inter-event times satisfy $\Pr(W_{ij} \le \Delta) = 1 - e^{-\gamma\Delta} = c$ (say), and with $\gamma = 1$, the preceding values of Δ give $c = 0.02$, 0.05 and 0.10, respectively. Table 1 presents empirical pth quantiles, $\hat{Q}_p$, of the 10,000 score statistics S as well as the estimates $\hat{P}(S > Q_p)$,

TABLE 1

$\hat{Q}_p$ is the empirical pth quantile of S in (3.4) computed from 10,000 samples when $\tau = 10$. $\hat{P}(S > Q_p)$ is the proportion of the values of S in 10,000 samples which are larger than the pth quantile of a standard normal distribution. The null model (2.4) has $\rho_0(t;\gamma) = 1$ and $\alpha_i \sim$ gamma (mean $= 1$, variance $= 0.3$)

Δ	m	$\hat{Q}_{0.950}$	$\hat{Q}_{0.975}$	$\hat{Q}_{0.990}$	$\hat{P}(S > 1.645)$	$\hat{P}(S > 1.960)$	$\hat{P}(S > 2.326)$
0.0202	10	1.658	2.090	2.632	0.0515	0.0301	0.0174
	20	1.569	1.950	2.426	0.0433	0.0248	0.0115
	50	1.362	1.720	2.095	0.0292	0.0148	0.0067
	100	1.243	1.591	1.990	0.0226	0.0107	0.0049
0.0513	10	1.469	1.873	2.289	0.0367	0.0206	0.0090
	20	1.418	1.781	2.166	0.0319	0.0168	0.0072
	50	1.234	1.511	1.932	0.0192	0.0096	0.0024
	100	0.988	1.265	1.622	0.0094	0.0045	0.0017
0.1054	10	1.361	1.685	2.139	0.0276	0.0142	0.0074
	20	1.242	1.599	1.981	0.0220	0.0104	0.0045
	50	1.013	1.365	1.703	0.0117	0.0059	0.0026
	100	0.751	1.047	1.417	0.0062	0.0027	0.0008

where Q_p are the standard normal p-quantiles for $p = 0.950$, 0.975 and 0.990. The results indicate that as m increases the standard normal approximation significantly underestimates right tail probabilities 0.05, 0.025 and 0.01. As the discussion in Section 3.1 indicates, this inaccuracy reflects the fact that, for fixed τ and increasing m, the α_i are not estimated consistently and (3.4) is not asymptotically normal. Most applications of the type considered here involve fairly large m and rather small numbers of events per individual, so we need an alternative way to get "honest" p-values. We recommend the use of simulation to obtain conditional (on $n_1, \ldots, n_m$) p-values, as described at the end of Section 3.1.

We next examine the power of (3.4) for tests with size 0.05. In each scenario described below we used the 10,000 realizations of the m processes represented in Table 1 to estimate 5% critical values, so as to have (approximately) correct type 1 error 0.05. We then estimated the power of (3.4) by generating 1000 samples in each scenario from the following model:

$$\lambda_i(t|H_i(t)) = \alpha_i \exp\{\beta I(N_i(t^-) > 1) I(B_i(t) \le \Delta_0)\}, \qquad i = 1, \ldots, m, \tag{4.1}$$

where the α_i $(i = 1, \ldots, m)$ are generated from a gamma distribution with mean 1 and variance ϕ. We allow Δ_0 to differ from Δ used in (3.4) in order to check on the effect of misspecifying Δ. We report here only the results under the model in (4.1) when $m = 20$. Table 2 and further simulation results confirm that power increases as τ and m increase. There is some loss of power if the assumed value of Δ is too large (i.e., if $\Delta > \Delta_0$), but little loss if it is too small. We also examined the effect of using the statistic (3.4) when the α_i in (2.4) are actually equal [model (4.1) with

TABLE 2
Proportion of times in 1000 *samples that test statistic* (3.4) *exceeded its* 0.05 *critical value for the alternative model* (4.1) *under various scenarios when* $m = 20$ *and* $\phi = 0.3$. *Critical values were estimated from* 10,000 *simulated samples*

		$\tau = 5$		$\tau = 10$	
Δ	Δ_0	$e^\beta = 2$	$e^\beta = 4$	$e^\beta = 2$	$e^\beta = 4$
	$\frac{2}{3}\Delta$	0.174	0.675	0.290	0.908
0.0202	Δ	0.294	0.874	0.481	0.983
	$\frac{4}{3}\Delta$	0.298	0.889	0.473	0.988
	$\frac{2}{3}\Delta$	0.317	0.945	0.531	0.998
0.0513	Δ	0.543	0.994	0.821	1.000
	$\frac{4}{3}\Delta$	0.509	0.991	0.794	1.000
	$\frac{2}{3}\Delta$	0.505	0.998	0.779	1.000
0.1054	Δ	0.794	1.000	0.973	1.000
	$\frac{4}{3}\Delta$	0.720	0.999	0.940	1.000

$\alpha_i = \alpha$], so that there is no heterogeneity. There is a slight loss of power relative to the test based on homogeneous Poisson processes [Çığşar (2010)], due to the fact that m values $\alpha_1, \ldots, \alpha_m$ are estimated instead of a single common value α. However, since failure to recognize heterogeneity can lead to incorrect rejection of the hypothesis of no carryover effect, the statistic (3.4) is preferable to the test statistic based on homogeneous processes.

The fixed effects tests are primarily of interest when m is small. We recommend the random effects tests more generally, and the remaining discussion concerns them. We first investigated the random effects test statistic (3.6) for the case where $\rho_0(t; \gamma) = \gamma$ in (2.4), and the α_i were independent gamma random variables with mean 1 and variance $\phi = 0.3$. We generated 10,000 replicates of m homogeneous Poisson processes for $\gamma = 1$ and different combinations of (Δ, m, τ) to evaluate the null distribution and critical values of (3.6). Normal quantile–quantile plots indicate that the standard normal approximation underestimates small p-values slightly for m less than 50 but is quite good at $m = 100$. Table 3 shows empirical type 1 errors corresponding to normal errors of 0.01, 0.025 and 0.05 for $\tau = 10$ and $m = 10, 20, 50, 100$. We also generated 1000 samples from versions of model (4.1) to estimate the power of the test. In each simulation run, we generated a new set of α_i from the gamma distribution with mean 1 and variance ϕ. Table 4 shows the results for different $(\Delta_0, e^\beta, m, \tau)$ combinations and $\phi = 0.3$. The power is generally high when $e^\beta = 3$, with a little decrease when Δ is chosen too large. The power values are higher than those for the fixed effects test in Table 2, in comparable scenarios. A simulation study for the power of the statistic (3.6) when

TABLE 3

$\hat{Q}_p$ is the empirical pth quantile of S in (3.6) computed from 10,000 samples when $m > 1$ and $\tau = 10$. $\hat{P}(S > Q_p)$ is the proportion of the values of S in 10,000 samples which are larger than the pth quantile of a standard normal distribution. The null model (2.4) has $\rho_0(t;\gamma) = 1$ and $\alpha_i \sim$ gamma (mean $= 1$, variance $= 0.3$)

Δ	m	$\hat{Q}_{0.950}$	$\hat{Q}_{0.975}$	$\hat{Q}_{0.990}$	$\hat{P}(S > 1.645)$	$\hat{P}(S > 1.960)$	$\hat{P}(S > 2.326)$
0.0202	10	1.835	2.263	2.735	0.0479	0.0303	0.0171
	20	1.785	2.177	2.707	0.0625	0.0370	0.0196
	50	1.725	2.099	2.589	0.0573	0.0326	0.0159
	100	1.703	2.020	2.434	0.0561	0.0284	0.0124
0.0513	10	1.779	2.179	2.656	0.0627	0.0357	0.0192
	20	1.694	2.080	2.458	0.0562	0.0312	0.0146
	50	1.691	2.027	2.404	0.0554	0.0289	0.0120
	100	1.665	1.997	2.361	0.0515	0.0268	0.0111
0.1054	10	1.682	2.049	2.456	0.0534	0.0291	0.0126
	20	1.669	2.016	2.366	0.0523	0.0285	0.0110
	50	1.642	2.008	2.345	0.0497	0.0280	0.0105
	100	1.631	1.942	2.359	0.0479	0.0238	0.0107

$\phi = 0.6$ gave similar results [Çığşar (2010) and supplementary file, Çığşar and Lawless (2012), Table S. 4].

In applications like the ones we consider, the number of individuals m is usually large but the expected number of events per individual is small. We next generated 10,000 realizations of m processes under the model (4.1) with $\phi = 0.6$ and $\beta = 0$,

TABLE 4

Proportion of times in 1000 samples that test statistic (3.6) exceeded its 0.05 critical value for the alternative model (4.1) under various scenarios when $\phi = 0.3$

		$m = 20, \tau = 10$		$m = 40, \tau = 5$		$m = 40, \tau = 10$	
Δ	Δ_0	$e^\beta = 2$	$e^\beta = 3$	$e^\beta = 2$	$e^\beta = 3$	$e^\beta = 2$	$e^\beta = 3$
	$\frac{2}{3}\Delta$	0.282	0.693	0.316	0.692	0.493	0.936
0.0202	Δ	0.437	0.912	0.496	0.924	0.781	0.994
	$\frac{4}{3}\Delta$	0.460	0.886	0.498	0.914	0.776	0.994
	$\frac{2}{3}\Delta$	0.565	0.959	0.527	0.949	0.805	0.999
0.0513	Δ	0.828	0.997	0.809	0.998	0.979	1.000
	$\frac{4}{3}\Delta$	0.776	0.997	0.806	0.996	0.972	1.000
	$\frac{2}{3}\Delta$	0.785	0.999	0.808	0.996	0.959	1.000
0.1054	Δ	0.968	1.000	0.968	1.000	1.000	1.000
	$\frac{4}{3}\Delta$	0.961	1.000	0.942	1.000	0.997	1.000

for the cases $m = 100, 200, 500$ and $E\{N_i(\tau_i)\}$ made equal to 1, 2, 5 by generating τ_i from a uniform distribution over (0.8, 1.2), (1.6, 2.4) or (4.0, 6.0), respectively. We calculated test statistic (3.6) for the values of $\Delta = 0.0513$, 0.1054 and 0.2231. The larger Δ values reflect features of the data considered in Section 5 and $\phi = 0.6$ is between plausible values in the two data sets there. Normal probability plots of (3.6) and Tables S. 5, S. 6 and S. 7 in the supplementary material [Çığşar and Lawless (2012)] show the standard normal approximation to be quite good except when $\Delta = 0.0513$, $m = 100$ and $E\{N_i(\tau_i)\} = 1, 2$. Once again, we recommend using simulation (parametric bootstrap) to get "honest" p-values for such cases. We also conducted a simulation study to investigate the power of the score statistic (3.6). We used the 10,000 realizations of the null model discussed above to estimate 5% critical values. We considered $m = 100, 200, 500$ and $\phi = 0.6$, and generated 1000 realizations of processes with the intensity function (4.1) for $\exp(\beta) = 1, 2$ and 3. Table 5 shows power of (3.6) for the combinations of $[\Delta, \Delta_0, \exp(\beta), E\{N_i(\tau_i)\}]$ when $m = 200$ (Tables S. 8 and S. 9 in the supplementary material give the results when $m = 100$ and 500, resp.). Overall, test statistic (3.6) maintains high power in these settings, and is robust with respect to mild misspecification of Δ.

Finally, simulation studies were conducted to examine the performance of the test statistic (3.6) when the assumption that the α_i have a gamma distribution is not true. To do that, we generated the α_i from a lognormal distribution with mean 1 and variance ϕ. We then generated 1000 realizations of m processes when $\Delta = 0.0202$ and $e^\beta = 1, 2, 3, 4$, and calculated the proportion of the time that (3.6) exceeded the 0.05 critical value. Results are given in Supplementary Table S. 10, for scenarios with $\tau = 10$ and $m = 20, 40$. The column $e^\beta = 1$ shows the empirical type 1 errors based on the 1000 samples; they are close to the nominal significance level 0.05.

TABLE 5

Proportion of times in 1000 *samples that test statistic* (3.6) *exceeded its* 0.05 *critical value for the alternative model* (4.1) *under various scenarios when* $\phi = 0.6$ *and* $m = 200$

		$E\{N_i(\tau_i)\} = 1$		$E\{N_i(\tau_i)\} = 2$		$E\{N_i(\tau_i)\} = 5$	
Δ	Δ_0	$e^\beta = 2$	$e^\beta = 3$	$e^\beta = 2$	$e^\beta = 3$	$e^\beta = 2$	$e^\beta = 3$
	$\frac{2}{3}\Delta$	0.585	0.975	0.858	1.000	0.995	1.000
0.0513	Δ	0.843	0.999	0.990	1.000	1.000	1.000
	$\frac{4}{3}\Delta$	0.809	0.998	0.985	1.000	1.000	1.000
	$\frac{2}{3}\Delta$	0.756	1.000	0.984	1.000	1.000	1.000
0.1054	Δ	0.952	1.000	1.000	1.000	1.000	1.000
	$\frac{4}{3}\Delta$	0.873	1.000	0.997	1.000	1.000	1.000
	$\frac{2}{3}\Delta$	0.803	1.000	0.999	1.000	1.000	1.000
0.2231	Δ	0.951	1.000	1.000	1.000	1.000	1.000
	$\frac{4}{3}\Delta$	0.842	0.999	0.999	1.000	1.000	1.000

In addition, (3.6) maintains high power in this case, and we conclude that mild misspecification of the distribution of random effects is not a problem; this agrees with similar results for estimation of rate functions in mixed Poisson processes without carryover effects [Lawless (1987)].

5. Applications.

5.1. *Recurrent asthma attacks in children* (*I*). Duchateau et al. (2003) discussed data from a prevention trial in infants with a high risk of asthma, but without a prior attack. The subjects were 6 months of age on entry to the study. The follow-up period for each subject was approximately 18 months, and started after random allocation to a placebo control group or an active drug treatment group. The main aim of the study was to assess the effect of the drug on the occurrence of asthma attacks, but an interesting secondary question was whether the occurrence of an event (asthma attack) influences the future event rate. There were 483 asthma attacks among 119 children in the control group and 336 asthma attacks among 113 children in the treatment group, during the 18 month follow-up.

The Nelson–Aalen estimates of the mean function [Cook and Lawless (2007), Section 3.4] for each treatment group are close to linear but that does not in itself show that the possibly heterogeneous individual rate functions are constant. Therefore, we fitted models (2.4) in which $\rho_0(t)$ took the power law form $\gamma_1\gamma_2 t^{\gamma_2-1}$. We found no evidence against the constancy of $\rho_0(t)$, and so the following details are based on constant rates which may vary across individuals. A caveat concerning the data is that Duchateau et al. (2003) do not provide the trial entry dates for each subject, so it is not possible to assess whether there might be a seasonal effect. However, for the second asthma data set considered in Section 5.2, such information was available and no seasonal effect was seen. An asthma attack lasts an average of 6–7 days, and a patient is not considered at-risk for a new attack over that time. The at-risk indicator $Y_i(t)$ takes value 1 if subject i is at risk of an asthma attack at time t, and the intensity model for subject i that we consider is therefore

$$\lambda_i(t|\mathcal{H}_i(t)) = Y_i(t)\alpha_i\gamma\exp\{\beta Z_i(t)\}, \qquad t \geq 0, \tag{5.1}$$

where $Z_i(t) = I\{N_i(t^-) > 0\}I\{B_i(t) \leq \Delta\}$, and $B_i(t)$ is the time since the subject i started their current at-risk period.

We will consider the treatment and control groups separately. To allow for heterogeneity, we use the tests of Section 3 with the random effects model (5.1), where $\alpha_i \sim \text{Gamma}(1, \phi)$, for testing $H_0 : \beta = 0$. Results obtained by fitting models with a range of values for Δ are shown in Table 6; to conserve space, standard errors for estimates $\hat{\phi}$ are not given, but in every model heterogeneity ($\phi > 0$) is strongly significant.

Table 6 gives, for each value of Δ, the estimates of γ, β and ϕ in model (5.1), along with the squared score statistic S^2 [with S given by (3.6)] and a corresponding Wald statistic for testing $\beta = 0$, defined as $Z^2 = \hat{\beta}^2/\widehat{\text{Var}}(\hat{\beta})$. The models were

TABLE 6

The results of the no carryover test based on (3.6) for various Δ values. Exp(Δ) *is the second term on the right-hand side of (3.7).* Z^2 *is the square of* $\hat{\beta}/se(\hat{\beta})$, *and* $\ell_{\max} = \ell(\hat{\gamma}, \hat{\beta}, \hat{\phi})$

Group	Δ	Obs(Δ)	Exp(Δ)	$\hat{\gamma}$	$\hat{\beta}$	$\hat{\phi}$	S^2	Z^2	$\ell_{\max}$
Treatment	7	40	22.858	0.006	0.681	0.476	14.900	14.314	−2009.41
	14	76	40.464	0.005	0.904	0.388	40.513	33.338	−1998.52
	28	119	67.099	0.005	1.017	0.305	61.968	59.360	−1988.08
	42	143	86.213	0.004	1.015	0.284	65.206	62.880	−1985.84
	56	162	101.774	0.004	1.029	0.270	68.857	66.694	−1983.75
	70	171	114.660	0.004	0.942	0.288	57.882	56.791	−1988.47
Control	7	68	47.173	0.008	0.486	0.521	11.751	11.551	−2726.18
	14	121	80.302	0.007	0.637	0.455	29.921	29.142	−2717.95
	28	185	130.457	0.007	0.678	0.399	40.284	39.411	−2712.53
	42	227	167.050	0.006	0.699	0.373	43.944	43.485	−2710.27
	56	260	195.336	0.006	0.745	0.350	49.393	48.478	−2707.26
	70	272	218.287	0.006	0.622	0.383	33.698	33.169	−2714.75

fitted using R function `nlm`, which automatically provides variance estimates via numerical differentiation. The score statistic S is more easily obtained since only restricted estimates $\tilde{\gamma}$ and $\tilde{\phi}$ are needed, but computational differences are unimportant here. The two statistics agree closely and strongly contradict the hypothesis ($\beta = 0$) of no carryover effect for every value of Δ shown. The p-values obtained from $\chi^2_{(1)}$ approximations for S^2 and Z^2 are virtually zero. As a check on this we also obtained p-values for S^2 by simulating 1000 samples under the null model with parameter values $\tilde{\gamma}, \tilde{\phi}$. For each value of Δ, there were no samples out of the 1000 generated in which S^2 exceeded its observed value in the data set. We also show observed and expected numbers [Obs(Δ), Exp(Δ)] of events in carryover periods, assuming no carryover effect; these are given in (3.7). This provides a nice summary of the excess events observed within time Δ of a preceding event.

Table 6 indicates that a wide range of values for Δ is plausible. We show the maximum values $\ell_{\max}$ of the log likelihood for each model, and see that the value of Δ (among those shown) best supported by the data is Δ = 56 days in both the treatment and control groups. It is also seen in Table 6 that estimates $\hat{\beta}$ and $\hat{\phi}$ are negatively correlated, as our discussion in Section 2 suggests. As Δ increases further beyond 70 days, the values of $\ell_{\max}$ continue to decrease, and Δ = 100 days still gives values that are about the same as Δ = 14 days. The values of $\tilde{\gamma}$ in the treatment and control groups, respectively, are 0.00608 and 0.00822, indicating an average of about one event every 165 days per subject in the treatment group, and one event every 122 days for the control group. The evidence indicates that events tend to occur closer to the previous event more often than is expected under a homogeneous Poisson process.

Our results can also be interpreted as indicating that the gap times between successive asthma attacks do not follow exponential distributions for individual subjects. Duchateau et al. (2003) fitted models in which gap times are assumed to be independent Weibull random variables within individuals, and heterogeneity is incorporated through individual-level gamma-distributed random effects. They found strong evidence of a decreasing hazard function for gap times, which is consistent with a carryover effect. The Duchateau et al. model has $p = 4$ parameters and ours has three, but AIC values ($-2\ell_{\max} + 2p$) are very close. For example, in the control group we find the AIC for (5.1) with $\Delta = 14$ days as 5441.9 ($p = 3$) and the AIC for the Duchateau et al. model as 5437.6 ($p = 4$). Smaller AIC values are obtained for models (5.1) with larger values of Δ; for example, when $\Delta = 56$ the AIC for the control group is 5420.5, the smallest for the models considered here.

Thus, all models indicate that the probability of a new asthma attack is highest soon after a preceding attack, and then decreases. Whether a delayed renewal process or a modulated Poisson process best describe the situation is not clear, nor whether there is a carryover effect of limited duration or a smooth decreasing hazard function for gap times. Without additional information concerning the asthma attacks and their treatment, we also cannot know the basis of the perceived effect.

5.2. *Recurrent asthma attacks in children (II).* We now briefly consider the randomized trial on the effects of 200 versus 400 μg per day of fluticasone propionate (FP) in preventing asthma attacks in children, mentioned in Section 1. Verona et al. (2003) describe the study in detail, and the data have been reanalyzed by Cook and Lawless (2007), Section 5.5.2. None of the previous analyses has looked at the interesting secondary issue of whether there is any indication of a carryover effect; we consider this here.

Earlier analyses showed that age and predicted expiratory flow (PEF) at enrollment had some predictive power for asthma exacerbations and we included them in our models. Seasonal effects and covariates such as sex and weight were examined but were not found significant and are excluded here. We ran analyses based on the modulated Poisson model (2.4) with gamma random effects and different values of Δ for the duration of carryover. In the interest of brevity we focus here on the FP200 group, which had 267 subjects. About one-third had approximately 3 months follow-up, with two-thirds followed for approximately 12 months. Cook and Lawless [(2007), page 195] show the numbers of asthma attacks per subject; there were a total of 359 in the FP200 group. As an illustration of the semiparametric approach we used the Andersen–Gill version of (2.4) with additional covariates, so no parametric assumption concerning $\rho_0(t)$ was made. According to the protocol for the trial, an exacerbation was counted only if it was not within 10 days of the start of a previous exacerbation, so the at-risk indicator $Y_i(t)$ introduced in (5.1) is defined so that $Y_i(t)$ equals 1 if and only if subject i is not within 10 days of a preceding exacerbation.

TABLE 7
Estimation results for Andersen–Gill models (2.4) with gamma random effects, fitted to FP200 asthma trial data

Δ^{a}	$\hat{\beta}$	$se(\hat{\beta})$	$\hat{\phi}$	Z^{2b}	$\ell_{\max}$
7	0.206	0.185	1.58	1.24	−1800.28
14	0.394	0.144	1.46	2.49	−1797.56
28	0.241	0.133	1.47	3.28	−1799.43
42	0.426	0.130	1.27	10.34	−1796.77
56	0.487	0.132	1.18	13.61	−1796.00
70	0.462	0.134	1.19	11.89	−1796.89
84	0.419	0.136	1.21	9.49	−1797.78

[a] Δ is in days.
[b] $Z^2 = \hat{\beta}^2/se(\hat{\beta})^2$.

Table 7 shows results for models fitted with various values of carryover duration Δ; models were fitted using R function `coxph`. As in the preceding example, there is strong evidence against the hypothesis of no carryover effect ($\beta = 0$), but a wide range of values for Δ is supported by the data. The best supported value is about 56 days (8 weeks), as in the study in Section 5.1. The average rate of events per subject in these data is about 1.8 per year, or about one asthma attack every 29 weeks. Therefore, there is once again an indication that the risk of a new attack is higher soon after a previous attack. As in the preceding case, there is also strong evidence of heterogeneity across subjects. This information, in conjunction with background medical information, may suggest that modifications to the prevention or treatment of attacks be considered.

6. Concluding remarks. We have considered modulated Poisson process models and tests for carryover effects, allowing for time trends and heterogeneity across processes. The random effects models and tests are recommended for general use; the tests have better power and are better approximated by asymptotic normal theory, especially when m is large. Fixed or time-varying covariates can be incorporated into our approach, as illustrated in Section 5.2.

It can be hard to deconstruct the dynamics of event occurrence when there are few events for most individuals, and the examination of alternative models is important. An alternative approach that is useful is to examine the distribution of "gap" times between successive events. The presence of a carryover effect is suggested by the density or hazard function for the gap times having substantial mass near zero. Such models do not produce definitions or tests for a carryover effect or handle time trends as readily as the models in Section 3. However, examination of gap time models as in Section 5.1 is often helpful, and in the absence of covariates, nonparametric estimates of hazard or density functions

for gap times are useful. As an additional illustration, we consider data on children with hydrocephalus, who have shunts inserted to drain excess cerebrospinal fluid. In the study mentioned in Section 1 [Tuli et al. (2000)], data on 839 children who had initial shunts inserted during the years 1989–1996 at one Canadian hospital were analyzed. Such shunts can "fail" due to blockages, infections and other conditions, necessitating full or partial replacement of the shunt. The data in question were analyzed previously by Lawless et al. (2001) and Cook and Lawless (2007), Section 5.4.2. Gap time models are a natural approach in this case: the occurrence of a failure results in a new shunt, and it makes sense to examine the lifetime of each subsequent shunt. The previous analyses were based on Cox models fitted to the survival times of successive shunts, and they showed that there were several important covariates, including the cause of a child's hydrocephalus and the age of the child at the time a shunt was (surgically) inserted. They also showed a tendency for second or third shunts to fail sooner than initial shunts. Plots of estimated baseline cumulative hazard functions $\tilde{H}_{0j}(t)$ for shunts $j = 1, 2, \ldots$ [e.g., Cook and Lawless (2007), Figure 5.9] suggested that the risk of failure was high soon after a new shunt was inserted, but this was not examined further. Table 8 shows a discretized estimate of the baseline hazard functions $h_{02}(w)$ and $h_{03}(w)$ for second and third shunts for a model involving adjustment for important covariates and additional allowance for heterogeneity. The covariates are coded for the two models such that the baseline hazard functions $h_{02}(w)$ and $h_{03}(w)$ represent the same vector of covariate values. The estimates are piecewise-constant, with $\tilde{h}_{0j}(w) = [\tilde{H}_{0j}(a_j) - \tilde{H}_{0j}(a_{j-1})]/(a_j - a_{j-1})$ for $a_{j-1} < w \le a_j$ and $a_j = 0, 60, 120, \ldots$ (days) for $j = 0, 1, 2, \ldots$. It is seen that the hazard functions are sharply decreasing. The time to failure of the initial shunt also shows a decreasing hazard function, but with overall smaller values. This indicates the risk

TABLE 8
Estimates of baseline cumulative hazard and piecewise-constant hazard functions for cerebrospinal fluid shunts

	Second shunts		Third shunts	
a	$\hat{H}_{02}(a)$	$\hat{h}_{02}(a)$[a]	$\hat{H}_{03}(a)$	$\hat{h}_{03}(a)$[a]
0	0		0	
60	0.19263	0.00321	0.30316	0.00505
120	0.26584	0.00122	0.39280	0.00149
180	0.29106	0.00042	0.44324	0.00084
240	0.32343	0.00054	0.47100	0.00046
300	0.35529	0.00053	0.48560	0.00024
360	0.38951	0.00057	0.51582	0.00050

[a] $\hat{h}_{0j}(a) = [\hat{H}_{0j}(a) - \hat{H}_{0j}(a - 60)]/60$, $j = 2, 3$.

of shunt failure is highest soon after it is inserted, and one explanation is that problems leading to a shunt failure may in some cases persist and create problems for the new shunt.

Finally, in many settings events of different types may occur. For example, that is the case with shunt failures, which can be due to obstruction, infection or other causes. In this context we can specify separate covariates to represent carryover effects related to the different event types. This is readily done in either the modulated Poisson process framework or the gap time framework. Table 8 is from a combined-causes analysis of the shunt failures, but separate causes could be considered similarly.

APPENDIX A: ANDERSEN–GILL MODEL

For the modulated Andersen–Gill model (2.2) for recurrent events, the Cox partial likelihood function for β gives the score function [Cook and Lawless (2007), page 71]

$$U_\beta(\beta) = \sum_{i=1}^{m} \left\{ \int_0^{\tau_i} Z_i(t) \left[dN_i(t) - \frac{d\bar{N}_\cdot(t) e^{\beta' Z_i(t)}}{\sum_{l=1}^{m} Y_l(t) e^{\beta' Z_l(t)}} \right] \right\}, \tag{A.1}$$

where $dN_i(t) = I$ (process i has an event at time t), $Y_l(t) = I(\tau_l \geq t)$, and $d\bar{N}_\cdot(t) = \sum_{l=1}^{m} Y_l(t)\, dN_l(t)$. The score statistic at $\beta = 0$ is

$$U_\beta(0) = \sum_{i=1}^{m} \left\{ \int_0^{\tau_i} Z_i(t) [dN_i(t) - \tilde{\rho}_0(t)\, dt] \right\}, \tag{A.2}$$

where, taking liberties with notation,

$$\tilde{\rho}_0(t)\, dt = \frac{d\bar{N}_\cdot(t)}{\sum_{l=1}^{m} Y_l(t)} \tag{A.3}$$

is the estimated baseline rate function at time t. Thus, (A.1) can be rewritten in "*Observed–Expected*" form as

$$U_\beta(0) = \sum_{i=1}^{m} \sum_{j=1}^{n_i} Z_i(t_{ij}) - \sum_{r=1}^{R} Z_\cdot(t_r^*) \frac{d\bar{N}_\cdot(t_r^*)}{Y_\cdot(t_r^*)}, \tag{A.4}$$

where $t_1^*, \ldots, t_R^*$ are the distinct event times across all processes, and $Z_\cdot(t) = \sum_{i=1}^{m} Z_i(t)$, $Y_\cdot(t) = \sum_{i=1}^{m} Y_i(t)$, and $d\bar{N}_\cdot(t)$ is defined following (A.1). This approach can be used if there is no evidence of heterogeneity across individuals. Usually this is not the case and then we should use the approach described at the end of Appendix B.

APPENDIX B: SCORE STATISTICS FOR GAMMA RANDOM EFFECTS MODELS

We consider here the score statistic (3.6) arising from the log likelihood (3.5). The numerator is easily shown to be

$$\begin{aligned} U_\beta(\tilde\gamma, 0, \tilde\phi) &= \left(\frac{\partial \ell(\gamma, 0, \phi)}{\partial \beta}\right)_{(\tilde\gamma, 0, \tilde\phi)} \\ &= \sum_{i=1}^{m}\left\{\sum_{j=1}^{n_i} Z_i(t_{ij}) - \frac{(1+\tilde\phi n_i)\,\partial R_i(\tilde\gamma, 0)/\partial\beta}{1+\tilde\phi R_i(\tilde\gamma, 0)}\right\}, \end{aligned} \tag{B.1}$$

where $R_i(\gamma, \beta)$ is given by (3.2). A variance estimate for $U_\beta(\tilde\gamma, 0, \tilde\phi)$ under H_0 is given by asymptotic theory for counting processes in the case where $m \to \infty$ [Andersen et al. (1993), Chapter 6, Peña (1998)]. This takes the standard form

$$\widehat{\mathrm{Var}}\{U_\beta(\tilde\gamma, 0, \tilde\phi)\} = \tilde I_{\beta\beta} - (\,\tilde I_{\beta\gamma} \quad \tilde I_{\beta\phi}\,)\begin{pmatrix} \tilde I_{\gamma\gamma} & \tilde I_{\gamma\phi} \\ \tilde I_{\phi\gamma} & \tilde I_{\gamma\gamma} \end{pmatrix}^{-1}\begin{pmatrix} \tilde I_{\gamma\beta} \\ \tilde I_{\phi\beta} \end{pmatrix}. \tag{B.2}$$

The 2×2 matrix in (B.2) is the inverse of the negative Hessian matrix for the log likelihood $\ell(\gamma, 0, \phi)$ evaluated at $\tilde\gamma$, $\tilde\phi$, and the terms $\tilde I_{\beta\beta}$, $\tilde I_{\beta\gamma}$ and $\tilde I_{\beta\phi}$ are based on the following, evaluated at $\tilde\gamma, \beta = 0, \tilde\phi$:

$$\begin{aligned} I_{\beta\beta} &= \frac{-\partial^2 \ell(\gamma, \beta, \phi)}{\partial \beta^2} \\ &= \sum_{i=1}^{m}(n_i + \phi^{-1})\left\{\frac{[\phi^{-1} + R_i(\gamma, \beta)][\partial R_i/\partial\beta] - [\partial R_i/\partial\beta]^2}{[\phi^{-1} + R_i(\gamma, \beta)]^2}\right\}, \\ I_{\beta\gamma} &= \frac{-\partial^2 \ell(\gamma, \beta, \phi)}{\partial\beta\,\partial\gamma'} \\ &= \sum_{i=1}^{m}(n_i + \phi^{-1})\left\{\frac{[\phi^{-1} + R_i(\gamma, \beta)][\partial^2 R_i/\partial\beta\,\partial\gamma'] - [\partial R_i/\partial\beta][\partial R_i/\partial\gamma']}{[\phi^{-1} + R_i(\gamma, \beta)]^2}\right\}, \\ I_{\beta\phi} &= \frac{-\partial^2 \ell(\gamma, \beta, \phi)}{\partial\beta\,\partial\phi} = \sum_{i=1}^{m}\left\{\frac{(\partial R_i/\partial\beta)[n_i - R_i(\gamma, \beta)]}{[1 + \phi R_i(\gamma, \beta)]^2}\right\}. \end{aligned}$$

The Andersen–Gill model of Appendix A with added frailty can be handled by the R/S-Plus Cox model function `coxph`. This implementation returns an estimate $\hat\beta$ and standard error, as well as a maximum likelihood value, so that a likelihood ratio or Wald test of $\beta = 0$ can be used. A score statistic analogous to (B.1) for the Cox model is

$$U_\beta(0) = \sum_{i=1}^{m}\left\{\sum_{j=1}^{n_i} Z_i(t_{ij}) - \frac{(1+\tilde\phi n_i)\sum_{t_r^* \le \tau_i} Z_i(t_r^*)(dN_\cdot(t_r^*)/Y_\cdot(t_r^*))}{1+\tilde\phi\sum_{t_r^* \le \tau_i}(dN_\cdot(t_r^*)/Y_\cdot(t_r^*))}\right\},$$

where the t_r^*, $dN_\cdot(t_r^*)$ and $Y_\cdot(t_r^*)$ are as defined in Appendix A. This statistic has the form "*Observed–Expected*;" the function `coxph` does not give it as output so some additional coding is required.

SUPPLEMENTARY MATERIAL

Additional simulation results (DOI: 10.1214/12-AOAS560SUPP; .pdf). The supplementary file contains detailed simulation results to support the discussion in Sections 2 and 4. Each simulation study in the supplementary file has its own description and title.

REFERENCES

ANDERSEN, P. K., BORGAN, Ø., GILL, R. D. and KEIDING, N. (1993). *Statistical Models Based on Counting Processes*. Springer, New York. MR1198884

ASCHER, H. and FEINGOLD, H. (1984). *Repairable Systems Reliability*: *Modeling*, *Inference*, *Misconceptions and Their Causes*. *Lecture Notes in Statistics* **7**. Dekker, New York. MR0762088

BAKER, R. D. (1996). Some new tests of the power law process. *Technometrics* **38** 256–265. MR1411882

BAKER, R. D. (2001). Data-based modeling of the failure rate of repairable equipment. *Lifetime Data Anal*. **7** 65–83. MR1836202

ÇIĞŞAR, C. (2010). Some models and tests for carryover effects and trends in recurrent event processes. Ph.D. thesis, Univ. Waterloo, Waterloo, Ont., Canada. Available at http://uwspace.uwaterloo.ca/handle/10012/5554.

ÇIĞŞAR, C. and LAWLESS, J. F. (2012). Supplement to "Assessing transient carryover effects in recurrent event processes, with application to chronic health conditions." DOI:10.1214/12-AOAS560SUPP.

COOK, R. J. and LAWLESS, J. F. (2007). *The Statistical Anlaysis of Recurrent Events*. Springer, New York.

COX, D. R. and ISHAM, V. (1980). *Point Processes*. Chapman & Hall, London. MR0598033

COX, D. R. and LEWIS, P. A. W. (1966). *The Statistical Analysis of Series of Events*. Methuen, London. MR0199942

DALEY, D. J. and VERE-JONES, D. (2003). *An Introduction to the Theory of Point Processes*. *Vol. I*: *Elementary Theory and Methods*, 2nd ed. Springer, New York. MR1950431

DUCHATEAU, L., JANSSEN, P., KEZIC, I. and FORTPIED, C. (2003). Evolution of recurrent asthma event rate over time in frailty models. *J. R. Stat. Soc. Ser. C. Appl. Stat*. **52** 355–363. MR1997294

FARRINGTON, C. P. and HOCINE, M. N. (2010). Within-individual dependence in self-controlled case series models for recurrent events. *J. R. Stat. Soc. Ser. C. Appl. Stat*. **59** 457–475. MR2756545

FARRINGTON, C. P. and WHITAKER, H. J. (2006). Semiparametric analysis of case series data. *J. R. Stat. Soc. Ser. C. Appl. Stat*. **55** 553–594. MR2291406

FARRINGTON, C. P., WHITAKER, H. J. and HOCINE, M. N. (2009). Case series analysis for censored, perturbed, or curtailed post-event exposures. *Biostatistics* **10** 3–16.

KARR, A. F. (1991). *Point Processes and Their Statistical Inference*, 2nd ed. *Probability*: *Pure and Applied* **7**. Dekker, New York. MR1113698

LAWLESS, J. F. (1987). Regression methods for Poisson process data. *J. Amer. Statist. Assoc*. **82** 808–815. MR0909986

LAWLESS, J. F., WIGG, M. B., TULI, S., DRAKE, J. and LAMBERTI-PASCULLI, M. (2001). Analysis of repeated failures or durations, with application to shunt failures for patients with paediatric hydrocephalus. *J. R. Stat. Soc. Ser. C. Appl. Stat*. **50** 449–465. MR1871799

LINDQVIST, B. H. (2006). On the statistical modeling and analysis of repairable systems. *Statist. Sci.* **21** 532–551. MR2369984

OGATA, Y. (1983). Likelihood analysis of point processes and its application to seismological data. *Bull. Int. Statist. Inst.* **50** 943–961.

PEÑA, E. A. (1998). Smooth goodness-of-fit tests for composite hypothesis in hazard based models. *Ann. Statist.* **26** 1935–1971. MR1673285

PEÑA, E. A. (2006). Dynamic modeling and statistical analysis of event times. *Statist. Sci.* **21** 487–500. MR2369983

TULI, S., DRAKE, J., LAWLESS, J. F., WIGG, M. and LAMBERTI-PASCULLI, M. (2000). Risk factors for repeated cerebrospinal shunt failures in patients with pediatric hydrocephalus. *J. Neurosurgery* **92** 31–38.

VERONA, E., PETROV, D., CSERHATI, E., HOFMAN, J., GEPPE, N., MEDLEY, H. and HUGHES, S. (2003). Fluticasone propionate in asthma: A long term dose comparison study. *Arch. Dis. Child.* **88** 503–509.

XIE, M., SUN, Q. and NAUS, J. (2009). A latent model to detect multiple clusters of varying sizes. *Biometrics* **65** 1011–1020. MR2756488

XU, S., ZHANG, L., NELSON, J. C., ZENG, C., MULLOOLY, J., MCCLURE, D. and GLANZ, J. (2011). Identifying optimal risk windows for self-controlled case series studies of vaccine safety. *Stat. Med.* **30** 742–752. MR2767769

WOMEN'S COLLEGE RESEARCH INSTITUTE
WOMEN'S COLLEGE HOSPITAL
UNIVERSITY OF TORONTO
790 BAY STREET, 7TH FLOOR
TORONTO, ONTARIO M5G 1N8
CANADA
E-MAIL: candemir.cigsar@wchospital.ca

DEPARTMENT OF STATISTICS
AND ACTUARIAL SCIENCE
UNIVERSITY OF WATERLOO
200 UNIVERSITY AVENUE WEST
WATERLOO, ONTARIO N2L 3G1
CANADA
E-MAIL: jlawless@math.uwaterloo.ca