Published in final edited form as:

J Appl Stat. 2012 ; 39(4): 883–899. doi:10.1080/02664763.2011.626848.

Mixed-Poisson Point Process with Partially-Observed Covariates: Ecological Momentary Assessment of Smoking

Benjamin Neustifter^{1,*}, Stephen L. Rathbun², and Saul Shiffman³

¹Food and Drug Administration, 10903 New Hampshire Ave., Silver Spring, MD 20903

²Department of Epidemiology and Biostatistics, University of Georgia, Athens, GA 30605 USA

³Department of Psychology, University of Pittsburgh, Pittsburgh, PA 15620 USA

Abstract

Ecological Momentary Assessment is an emerging method of data collection in behavioral research that may be used to capture the times of repeated behavioral events on electronic devices, and information on subjects' psychological states through the electronic administration of questionnaires at times selected from a probability-based design as well as the event times. A method for fitting a mixed Poisson point process model is proposed for the impact of partially-observed, time-varying covariates on the timing of repeated behavioral events. A random frailty is included in the point-process intensity to describe variation among subjects in baseline rates of event occurrence. Covariate coefficients are estimated using estimating equations constructed by replacing the integrated intensity in the Poisson score equations with a design-unbiased estimator. An estimator is also proposed for the variance of the random frailties. Our estimators are robust in the sense that no model assumptions are made regarding the distribution of the time-varying covariates or the distribution of the random effects. However, subject effects are estimated under gamma frailties using an approximate hierarchical likelihood. The proposed approach is illustrated using smoking data.

Keywords

Behavioral data; estimating equations; hierarchical likelihood; probability-sampling; increasing-domain asymptotics; point processes

1. Introduction

Investigation of the impact of time-varying covariates on recurrent events can be implemented using Ecological Momentary Assessment (EMA). EMA is an emerging method of data collection in behavioral research [34, 38] that often uses electronic devices to record information about each subject's behavioral states and repeated behavioral events as they occur in the subjects' natural environments. EMA reduces recall biases in results and avoids effects introduced by an artificial laboratory environment, thereby increasing the "ecological validity" of the data [34]. Previous uses of EMA include the study of eating disorders [37], stress coping [40], and pain [39,41], to list a few. By utilizing probability-based sampling of subjects' psychological states and other variables, EMA allows for the unbiased estimation of their mean states. Furthermore, by additionally assessing subjects at event times, this method makes it possible to estimate the effects of time-varying covariates upon the timing of the events of interest.

*bbarone@gmail.com.

The motivating data comes from, Shiffman et al.'s [36] nicotine patch trial on a sample of 412 smokers. Prior to treatment, subjects recorded the exact times that they smoked on electronic diaries and answered questions regarding their mood (e.g. restlessness) and environment (e.g. whether other smokers were present) at randomly selected smoking times and other randomly selected times during the day. By analyzing these data, we hope to gain insight regarding the impact of covariates on cigarette consumption rates in smokers; this will contribute to our understanding of what prompts smoking and may assist in designing effective therapies for smoking cessation.

Rathbun et al. [28] had previously fit a modulated Poisson point process [8] to the EMA smoking data, predicting the intensity of recurrent behavioral events as a function of partially-observed, time-varying covariates. The modulated Poisson process is particularly well-suited for modeling the impact of time-varying covariates on rates of behavioral events, and is a special case of an inhomogeneous Poisson process [9] whose log intensity is a linear function of the covariates. Writing the log intensity as a linear function of the covariates guarantees that the intensity is nonnegative, and the logarithm is the canonical link function for an inhomogeneous Poisson process [25]. Since the intercept term did not depend on subject, Rathbun et al.'s [28] implementation of the modulated Poisson process model implicitly assumes that all subjects have identical baseline smoking rates, yet it is well established that different smokers smoke at different rates. Failure to account for variation among subjects in baseline smoking rates may result in the underestimation of the standard errors of model parameters [1,14]. In this paper, we consider frailty models for point process intensity that allow for variation among subjects in baseline smoking rates. In the context of recurrent-events data, the frailty is an unobserved, subject specific factor that describes the impact of risk factors not included in the model [15,44].

To account for heterogeneity among subjects, Lawless [19] and Thall [42] introduced Poisson point process models with gamma frailties. The intensity of the point process was modeled parametrically, but time-varying covariates were not considered. Since the gamma distribution is conjugate for the Poisson point process, the likelihood can be written in closed form. Lawless and Zhan [21] suggested the use of Poisson score equations for subject-level covariate effects, and use a robust standard error estimator to account for over dispersion resulting from random variation among subjects. For time-varying covariates, Poisson score equations require that the covariates to be known functions of time, a requirement that is seldom satisfied in practice. To model the impact of partially-observed, time-varying covariates on the lifetimes to single events, Wulfsohn and Tsiatis [47] construct a joint model for the time-varying covariates and the event lifetimes. This joint modeling approach has been extended to recurrent event modeling by Zhang et al. [53] and Liu and Huang [24]. The estimators for covariate effects are consistent under correctly specified models, but may be biased if the model for the mean of the covariates as a function of time is misspecified [29].

Smoking rate may be a function of mood [16,35], which may vary in response to multiple stimuli occurring at arbitrary intervals and at frequencies greater than the frequency at which mood is sampled. No model is likely to adequately capture this variation in mood, so the joint modeling approach may yield biased estimates of point process model parameters. Rathbun et al. [28] propose a probability sampling framework for modeling the impact of partially-observed covariates on Poisson process intensity that does not involve a model for the time-varying covariates. A design-unbiased estimator for the integrated intensity is constructed from the probability sample of the covariates, and substituted into the likelihood equations, which are then solved to obtain an estimator for point process parameters. Waagepetersen [45] improves on this approach by using covariate information from the

events as well as the probability sample to construct an unbiased estimator for the integrated intensity.

After describing the point process model with random effects, Section 2 proposes the use of the estimating equations of Rathbun et al. [28] and Waagepetersen [45] for the estimation of point process parameters. A robust estimator is also proposed for the variance of the frailties that does not require any assumptions regarding the distribution of the frailties. Subjects' frailties may be estimated based on approximation to the $\log$ -likelihood of Lee and Nelder [22]. In Section 3, we state Theorems that demonstrate that the proposed estimating equations yield consistent estimators of point process parameters and that the variance estimator is consistent as well. Standard errors of point process parameters are adjusted to account for variation in frailties among study subjects. Section 4 will describe the motivating data set in detail and demonstrate the use of the derived estimators on these data, comparing results with those received from Rathbun et al. [28] and Waagepetersen's [45] estimators. The Appendices contain the proofs of consistency and asymptotic normality of the estimators.

2. Statistical Inference for the Random Effects Point Process Model

Suppose that n subjects are independently sampled, and that subject i is observed over a set of times belonging to the Borel set $A_i \subset \mathbb{R}$. Here, the sampling window A_i may be a single interval of time, but in the current application, A_i is the union of a finite number of disjoint time intervals corresponding to the set of times during which the electronic diary is actively collecting data. The electronic diary is suspended during sleep intervals, naps, and when the subject turned it off for meetings, movies, etc. For each subject i , N_i recurrent events are observed with event times $t_{i1} < \dots < t_{iN_i} \in A_i$. Let $z_i(t)$; $t \in A_i$ denote a $p \times 1$ vector of covariates; both time-varying and time-invariant covariates may be included.

To permit inter-subject variability in baseline event rates, we could consider a fixed effects model under which the event times for each subject i are realized from an inhomogeneous Poisson process with intensity

$$\lambda_i(t) = u_i \exp\{\beta^T z_i(t)\}; t \in A_i, i = 1, \dots, n, \quad (1)$$

where the u_i 's are fixed subject effects. Excluding the intercept term from β , this overparameterized fixed-effects model has $n + p$ parameters. Consistent estimators of subject effects can only be achieved under an asymptotic regime where for each subject, the Lebesgue measure $|A_i| \rightarrow \infty$ as $n \rightarrow \infty$ see Royden [30] for an explanation of Lebesgue measure. In the context of the current research, $|A_i|$ is the total length of time subject i is observed. Moreover, estimates of model parameters, especially the subject effects u_i may be unstable and imprecise in this over-parameterized model.

In the following, we consider a random frailty model, under which the frailties u_i in (1) are independently sampled from a distribution with mean 1, variance σ^2 , and finite third and fourth moments. Note that, if an intercept term is included in the parameter vector β , the first moment, if not equal to 1, can be absorbed into the intercept term. Note that this random intensity defines a simple Cox process [7], a Poisson point process with random intensity. A particular distribution which satisfies these assumptions is the gamma distribution with mean one and variance $\sigma^2 = \gamma^{-1}$; that is, a gamma distribution with density $f(u) = \gamma^\gamma u^{\gamma-1} e^{-\gamma u} / \Gamma(\gamma)$. The gamma distribution is particularly appropriate to this setting, as it is conjugate for the Poisson process and therefore allows for a closed-form marginal likelihood. However, as the consistency of our estimator depends only upon the first four moments of u_i , any distribution meeting the above assumptions will suffice.

2.1 Fully-Observed Covariates

To motivate the methods for statistical inference, we temporarily assume in this subsection that the covariates $z_i(t)$ are known functions of time t for all $t \in A_i$ and all $i = 1, \dots, n$. For the moment, we also assume that the baseline rates u_i follow a gamma distribution with mean one and variance $\sigma^2 = \gamma^{-1}$. Then we obtain the marginal log-likelihood for $\theta = (\beta, \sigma^2)^T$ to be

$$\ell_n(\theta) = n\gamma \ln \gamma + \sum_{i=1}^n \left\{ \sum_{t \in X_i} \beta^T z_i(t) - (\gamma + N_i) \ln(\gamma + \Lambda_i(A_i; \beta)) + \ln(\gamma + N_i) \right\}, \quad (2)$$

where

$$\Lambda_i(A_i; \beta) = \int_{A_i} \exp\{\beta^T z_i(t)\} dt$$

is the integrated intensity, N_i is the number of events, and $X_i \subset A_i$ is the random set of event times for the i th subject. Other distributions may be assumed for the u_i terms, but the gamma allows for a closed-form marginal log-likelihood and gives score function that inspire the suggested estimating function. The score equations for β are

$$\sum_{i=1}^n \left\{ \sum_{t \in X_i} z_i(t) - \frac{(\gamma + N_i) \Lambda_i^{(1)}(A_i; \beta)}{\gamma + \Lambda_i(A_i; \beta)} \right\} = 0, \quad (3)$$

where $\Lambda_i^{(1)}(A_i; \beta) = (\partial / \partial \beta^T) \Lambda_i(A_i; \beta)$. The score equations are unbiased in the sense that the expected value of the left-hand side is zero under the true value of β [11]. This is true even if the random effects are not gamma distributed, provided that its expected value is one. This suggests that the solution to (3) may be robust to violation of the assumption that the random effect is gamma distributed.

Let $v_i = \log u_i$ denote the canonical scale with respect to β [23]. Then the h-log likelihood for the Poisson-gamma hierarchical model is

$$h_n(\theta, v) = \sum_{i=1}^n \left[\left(\sum_{t \in X_i} \beta^T z_i(t) \right) + (N_i + \gamma) v_i - (\Lambda_i(A_i; \beta) + \gamma) e^{v_i} + \gamma \ln \gamma - \ln \Gamma(\gamma) \right].$$

Taking the derivatives with respect to β and v_i yields the h-score equations

$$\sum_{i=1}^n \left[\sum_{t \in X_i} z_i(t) - \Lambda_i^{(1)}(A_i; \beta) e^{v_i} \right] = 0 \quad (4)$$

and

$$N_i + \gamma - (\Lambda_i(A_i; \beta) + \gamma) e^{v_i} = 0 \quad (5)$$

Solving the latter for v_i yields the following estimator for the random effect

$$\widehat{u}_i(\beta, \gamma) = \frac{N_i + \gamma}{\Lambda_i(A_i; \beta) + \gamma}, \quad (6)$$

which when substituted into (4) yields the score equations (3) for β . Thus, the marginal maximum likelihood estimator for β can be obtained from the joint maximization of the h-likelihood. The h-score equations for β are unbiased in the sense that the left-hand side of (4) is zero under the true value of β [11]. Note, however, that for each θ ,

$$E_{\theta}\{\widehat{u}_i(\beta, \gamma) | u_i\} = u_i + \frac{\gamma(1 - u_i)}{\Lambda_i(A_i; \beta) + \gamma},$$

so the conditional expectations of $\widehat{u}_i(\beta, \gamma)$ given u_i are shrunk towards one. Assuming that the intensity is bounded below by a constant strictly greater than zero, $\Lambda_i(A_i; \beta) \rightarrow \infty$ the second term in the right-hand side of the expression above converges to zero, and so the conditional expectation of $\widehat{u}_i(\beta, \gamma)$ given u_i converges to u_i as $|A_i| \rightarrow \infty$. Both $\widehat{u}_i(\beta, \gamma)$ and u_i have unit mean under the true value of the model parameter, and the mean squared error of $\widehat{u}_i(\beta, \gamma)$ is

$$E_{\theta}\{\widehat{u}_i(\beta, \gamma) - u_i\}^2 = \{\Lambda_i(A_i; \beta) + \gamma\}^{-1},$$

which converges to zero as $|A_i| \rightarrow \infty$. So, by Chebychev's inequality, $\widehat{u}_i(\beta, \gamma)$ converges in probability to u_i as $|A_i| \rightarrow \infty$. Moreover, since $\widehat{u}_i(\beta, \gamma)$ is continuous with respect to β , $\widehat{u}_i(\hat{\beta}, \gamma)$ remains consistent for u_i provided that $\hat{\beta}$ is a consistent estimator for β .

To obtain a simpler estimating function for β that is still unbiased, we note that since $E_{\theta}(N_i) = \Lambda_i(A_i; \beta)$, the ratio term in (3), evaluated at the true parameter θ , has expected value

$$E_{\theta}\left\{\frac{(\gamma_0 + N_i)\Lambda_i^{(1)}(A_i; \beta)}{\gamma_0 + \Lambda_i(A_i; \beta)}\right\} = \Lambda_i^{(1)}(A_i; \beta).$$

Thus, an unbiased estimating equation for β can be obtained by substituting $\Lambda_i^{(1)}(A_i; \beta)$ for the ratio in (3), yielding the Poisson score function $\Psi_n(\beta) = n^{-1} \sum_{i=1}^n \psi_i(\beta)$, where

$$\psi_i(\beta) = \sum_{t \in X_i} z_i(t) - \Lambda_i^{(1)}(A_i; \beta) \quad (7)$$

This estimating function is unbiased in the sense that $E_{\beta}\{\Psi_n(\beta)\} = 0$ for all $\beta \in \mathbb{R}^p$ [11]. Note that this estimating equation does not involve the variance parameter $\gamma = \sigma^2$, and that it is unbiased regardless of the distribution of u_i provided that it has unit mean. The former is particularly helpful, as it allows us to treat σ^2 as a nuisance parameter to be estimated after a final estimate for β has been obtained, rather than simultaneously with β . While some loss of efficiency is expected relative to the maximum likelihood estimation under gamma frailties,

the solution to the Poisson estimating equations is robust to violations of the assumption that frailties are gamma distributed [20].

To obtain an unbiased estimate for σ^2 , define the inverse of the intensity

$$\varphi_i(\beta, t) = \exp\{-\beta^T z_i(t)\}. \quad (8)$$

Then

$$\hat{\sigma}_n^2(\beta_0) = n^{-1} \sum_{i=1}^n \left\{ |A_i|^{-2} \sum_{s \neq t \in X_i} \varphi_i(\beta_0, s) \varphi_i(\beta_0, t) - 1 \right\}. \quad (9)$$

is an unbiased estimator for σ^2 , where β_0 is the true value of β . Since β_0 is unknown, we employ the plug-in estimator $\hat{\sigma}_n^2(\hat{\beta})$, where $\hat{\beta}$ is a consistent estimator for β . While the estimating equation $\hat{\Psi}_n(\beta) = 0$ for β remains unbiased even if the data are not realized from a Poisson process [32], $\hat{\sigma}_n^2(\beta_0)$ is unbiased for σ^2 if and only if the second-order intensity [12] satisfies $\lambda_2(s, t) = \lambda(s)\lambda(t)$, a condition satisfied by the Poisson process.

2.2 Partially-Observed Covariates

To evaluate the function $\Lambda_i(A_i; \beta)$ and derivatives necessary for estimating θ , the covariates $z_i(t)$ must be a known function of time $t \in A_i$ for all $i = 1, \dots, n$. However, in many applied settings it is only possible to sample the covariates. While the results established in the previous subsection assume that the covariates are a known function of time, we will now address the problem of obtaining estimates of the parameters with desirable asymptotic properties when this is not the case.

Suppose that for each subject i , covariates are assessed at times t belonging to a random set $D_i \subset A_i$ of covariate sample times generated by a known probability-based sampling design. Probability-based sampling of covariates is a traditional feature of EMA, so such data are readily available. For each β , $\Lambda_i^{(1)}(A_i; \beta)$ may be regarded as a population total for $\lambda_i^{(1)}(t) = x_i(t) \exp\{\beta^T x_i(t)\}$, where the population is taken to be the collection of all points $t \in A_i$. Cordy [6] introduced design-based inference for spatial sampling designs. For a sample of fixed size m_i , define the inclusion density $\pi_i(t) = \sum_{j=1}^{m_i} f_{ij}(t)$, where $f_{ij}(t)$ is the marginal density for the j th covariate sample site. Alternatively, the covariates may be sampled according to Poisson point processes with known intensities $\pi_i(t)$. In either case, the Horvitz-Thompson estimator

$$\hat{\Lambda}_i^{(1)}(A_i; \beta) = \sum_{t \in D_i} \frac{z_i(t) \exp\{\beta^T z_i(t)\}}{\pi_i(t)} \quad (10)$$

is design-unbiased for $\Lambda_i^{(1)}(A_i; \beta)$ provided $\pi_i(t) > 0$ for all $t \in A_i$ [6]. Waagepetersen [45] suggests the unbiased estimator

$$\tilde{\Lambda}_i^{(1)}(A_i; \beta) = \sum_{t \in X_i \cup D_i} \frac{z_i(t) \exp\{\beta^T z_i(t)\}}{\pi_i(t) + \exp\{\beta^T z_i(t)\}}, \quad (11)$$

Waagepetersen's [45] estimator should be more efficient than (10) since it uses covariate information from not only from the probability sample D_i but also from the times of the events X_i . Substituting either (10) or (11) into (7) and rearranging terms gives weighted estimating functions of the form $\tilde{\Psi}_n(\beta) = n^{-1} \sum_{i=1}^n \hat{\psi}_i(\beta)$, where

$$\hat{\psi}_i(\beta) = \sum_{t \in X_i} z_i(t) w(t; \beta) - \sum_{t \in D_i} \frac{z_i(t) w(t; \beta) \exp\{\beta^T z_i(t)\}}{\pi_i(t)}, \quad (12)$$

$w(t, \beta) = 1$ for (10) and Waagepetersen's weights $w_i(t, \beta) = \pi_i(t) [\pi_i(t) + \exp\{\beta^T z_i(t)\}]^{-1}$ for (11). The proposed estimator is obtained by solving $\tilde{\Psi}_n(\beta) = 0$ for $\tilde{\beta}_n$. Given $\tilde{\beta}_n$, the variance σ^2 in the baseline event rate is estimated using the plug-in estimator $\hat{\sigma}_n^2(\tilde{\beta}_n)$; see expression (9). In the following, let $\tilde{\theta}_n = (\tilde{\beta}_n, \hat{\sigma}_n^2(\tilde{\beta}_n))$.

Estimators of the form (10) or (11) may be substituted into the h-score equations (4) and (5), the solution of which might be used to obtain an approximate marginal maximum likelihood estimator. The resulting approximate h-score equations remain unbiased. It is possible to demonstrate that solution is consistent, provided that the sizes of the sample windows $|A_i| \rightarrow \infty$ as $n \rightarrow \infty$. In most practical applications, however, the sample window sizes are bounded from above, and some subjects may drop out early resulting in small sample windows. The proposed estimating functions (12) are unbiased. Moreover, the next Section will demonstrate that their solutions are consistent under bounded sample window sizes.

3. Asymptotic Properties

We demonstrate the large-sample properties of the proposed estimator $\tilde{\theta}_n$ as the number of subjects $n \rightarrow \infty$. Kutoyants [18], Brillinger [3], Krickeberg [17], Rathbun and Cressie [27], Rathbun et al. [28], and Waagepetersen [45] have demonstrated the consistency and asymptotic normality of parameter estimators for the Poisson process under increasing domain asymptotics. Here, only a single realization of the Poisson is observed on a sample region A , and large-sample properties are demonstrated as the volume $|A| \rightarrow \infty$. Owing to the independent increments property of the Poisson process, these results readily extend to sampling n independent realizations of a Poisson process, and letting $n \rightarrow \infty$. The following considers the mixed effects model (1) with frailties having mean 1, variance σ^2 , and finite fourth moments. We prove that the proposed estimator $\tilde{\theta}_n$ is consistent and asymptotic normal as $n \rightarrow \infty$ under fixed-sample covariate sample designs. Similar results can also be obtained for Poisson point process sampling of covariates; remarks will be provided indicating modifications of results required to accommodate Poisson point process sampling.

Consider the following assumptions:

Assumption 1

The subjects $i = 1, \dots, n$ are independently sampled in uniformly bounded study intervals A_i , and the sets of covariate sample times D_i are independent of the recurrent event times X_i , both within and between subjects.

Assumption 2

There exists a constant $M < \infty$ such that $|z_i(t)| < M$ for almost all $t > 0$ and all $i = 1, \dots, n$.

Assumption 3

The weight function $w_i(t; \beta)$ is bounded from above and below by a positive constant, and has continuous first- and second-order derivatives with respect to β . Moreover, the family of first derivatives $\{w_i^{(1)}(t; \beta) : t \in A_i, i \geq 1\}$ is equicontinuous in β .

Assumption 4

$$n^{-1} \sum_{i=1}^n \int_{A_i} x_i(t) x_i(t)^T dt \rightarrow B$$

as $n \rightarrow \infty$, where the minimum eigenvalue of B is strictly positive.

Assumption 5

The random covariate assessments $\{u \in D_i; i = 1, \dots, n\}$ are obtained according to a known probability-based sampling scheme with single and pairwise inclusion densities $\pi_i(u)$ and $\pi_i(u, v)$ [6], where the sampling intensity is such that

$$\frac{\sum_{i=1}^n m_i}{\sum_{i=1}^n |A_i|} \rightarrow \rho$$

as $n \rightarrow \infty$, where $0 < \rho < \infty$ and m_i is the number of times the covariates were sampled for subject i .

Assumption 6

The inclusion densities $\pi_i(u)$ and pairwise inclusion densities $\pi_i(u, v)$ are bounded below by constants greater than zero for all $t \in A_i; i = 1, \dots, n$.

These assumptions are similar to those posed by Rathbun et al. [28]. Assumption 1 is necessary to ensure that there is no correlation either among subjects, or between events and covariates sample times within subjects. The study intervals are uniformly bounded if their volumes $|A_i| < a$ for some constant $0 < a < \infty$. This condition ensures that data from a finite number of subjects do not dominate the estimator for the baseline variance σ^2 . Assumption 2 ensures that the estimating equations are bounded, and also that $E\{\widehat{\Lambda}_i^{(1)}(A_i; \beta)\} = \Lambda_i^{(1)}(A_i; \beta)$ [6]. The lower bound on the weight function in Assumption 3 ensures that all events and dummy points receive some weight in the computation of estimating function, and the upper bound ensures that the estimating function does not become dominated by a finite number of events or covariate sample times as $n \rightarrow \infty$. Existence of the derivatives of the weight function with respect to β ensures existence of the corresponding derivatives of the estimating function. Under Assumption 4, $E_{\beta}\{(\partial / \partial \beta^T) \widehat{\Psi}_n(\beta)\}$ will be positive definite for sufficiently large n ; this assumption also ensures that there is no redundancy among the covariates. Assumptions 5 and 6 place some weak restrictions on the density for the sampling of the covariate sample points. It is sufficient for Assumption 5 that the random assessments are realized from a probability sampling design where the expected number of assessments per unit time is equal across subjects or subject/stratum combinations. Assumption 6 ensures

that the random assessments adequately cover the study region, and allows for unbiased estimation of $\Lambda_i^{(1)}(A_i; \beta)$. For Poisson point process sampling, Assumptions 5 and 6 are replaced by the assumption that the intensity $\pi_A(t)$ is bounded from above, and below by a constant exceeding zero.

The following lemmas are required to obtain the existence of a consistent solution $\tilde{\beta}_n$ to $\hat{\Psi}_n(\beta) = 0$:

Lemma 1—Under Assumptions 1–3, 5 and 6 and for each neighborhood $S(\beta_0, \epsilon)$ of β_0 the estimating functions $\hat{\Psi}_n(\beta)$ converge uniformly in β to a nonstochastic limit $\Psi_\infty(\beta)$ that is equal to zero at β_0 as $n \rightarrow \infty$.

Lemma 2—Under Assumptions 1–6 and for each neighborhood $S(\beta_0, \epsilon)$ of β_0 the Jacobians $\tilde{\Psi}_n^{(1)}(\beta)$ converge uniformly in β to a nonstochastic limit that is negative definite at β_0 as $n \rightarrow \infty$.

Proof of both lemmas may be found in Appendix A. The result that $\hat{\Psi}_n(\beta_0)$ converges to zero almost surely is necessary to achieve a consistent solution. The constancy of $\tilde{\beta}_n$ is an immediate consequence of Theorem 3 of Yuan and Jennrich [49]:

Theorem 1

Suppose that Assumptions 1–3, 5 and 6 are satisfied, then $\tilde{\beta}_n \rightarrow \beta_0$ almost surely as $n \rightarrow \infty$.

To obtain the large-sample variance of $\tilde{\beta}_n$ define

$$J_n(\beta) = E_\theta \left\{ \frac{\partial}{\partial \beta} \tilde{\Psi}_n(\beta) \right\} \\ = -n^{-1} \sum_{i=1}^n \left\{ \int_{A_i} z_i(t) z_i(t)^T w(t; \beta) \exp\{\beta^T z_i(t)\} dt \right\}, \quad (13)$$

$$\Sigma_n^Q(\beta) = \text{var}_\theta \left\{ n^{-\frac{1}{2}} \sum_{i=1}^n \sum_{t \in X_i} z_i(t) w(t; \beta) \right\} \\ = n^{-1} \sum_{i=1}^n \sigma^2 \left\{ \int_{A_i} \int_{A_i} z_i(s) z_i(t)^T w(s; \beta) w(t; \beta) \exp\{\beta^T (z_i(s) + z_i(t))\} ds dt + \int_{A_i} z_i(t) z_i(t)^T w(t; \beta)^2 \exp\{\beta^T z_i(t)\} dt \right\}, \quad (14)$$

and

$$\Sigma_n^R(\beta) = \text{var}_\theta \left\{ n^{-\frac{1}{2}} \sum_{i=1}^n \sum_{t \in D_i} \frac{z_i(t) w_i(t; \beta) \exp\{\beta^T z_i(t)\}}{\pi_i(t)} \right\} \\ = n^{-1} \sum_{i=1}^n \left\{ \int_{A_i} \int_{A_i} Y_i(s) Y_i(t)^T \left(\frac{\pi_i(s, t) - \pi_i(s) \pi_i(t)}{\pi_i(s) \pi_i(t)} \right) ds dt + \int_{A_{ni}} Y_i(t) Y_i(t)^T \pi_i(t)^{-1} dt \right\}. \quad (15)$$

where $Y_i(t, \beta) = z_i(t) w_i(t; \beta) \exp\{\beta^T z_i(t)\}$. Here, $J_n(\beta)$ is the information matrix, and

$\Sigma_n^Q(\beta_0)$ and $\Sigma_n^R(\beta_0)$ are the variance components of the estimating equation attributed to the event assessments and the random assessments, respectively. If covariates are sampled from a Poisson point process with intensity $\pi_A(t)$, then the pair inclusion density $\pi_A(s, t) = \pi_A(s) \pi_A(t)$, so the first term on the right-hand side of (15) vanishes. Let $J_\infty(\beta_0)$, $\Sigma_\infty^Q(\beta_0)$, and $\Sigma_\infty^R(\beta_0)$ be the limits of $J_n(\beta_0)$, $\Sigma_n^Q(\beta_0)$, and $\Sigma_n^R(\beta_0)$, respectively as $n \rightarrow \infty$. The

existence of these limits is guaranteed under Assumptions 1–4, and 6 and the bounded convergence theorem. For Poisson point process sampling of covariates, the first term on the right-hand side of (15) is equal to zero.

Before obtaining the asymptotic normality of $\tilde{\beta}_n$, we must first demonstrate the asymptotic normality of the estimating function $\tilde{\Psi}_n(\beta_0)$:

Lemma 3—Under Assumptions 1–3, 5, and 6, $n^{\frac{1}{2}}\tilde{\Psi}_n(\beta_0)$ converges in distribution to a Gaussian random vector with mean 0 and variance

$$\sum_{\infty}^Q(\beta_0) + \sum_{\infty}^R(\beta_0),$$

where $\sum_{\infty}^Q(\beta_0)$ and $\sum_{\infty}^R(\beta_0)$ are the respective limits of $\sum_n^Q(\beta)$ and $\sum_n^R(\beta)$ defined in expressions (14) and (15) as $n \rightarrow \infty$

This Lemma is also proved in Appendix A. The asymptotic normality of the estimator $\tilde{\beta}_n$ then follows from Theorem 4 of Yuan and Jennrich [49]:

Theorem 2

Under Assumptions 1–6,

$$n^{-\frac{1}{2}}(\tilde{\beta}_n - \beta_0) \xrightarrow{\mathcal{L}} N(0, V(\beta_0))$$

where

$$V(\beta_0) = \{J_{\infty}(\beta_0)\}^{-1} \sum_{\infty}^Q(\beta_0) \{J_{\infty}(\beta_0)\}^{-1} + \{J_{\infty}(\beta_0)\}^{-1} \sum_{\infty}^R(\beta_0) \{J_{\infty}(\beta_0)\}^{-1},$$

as $n \rightarrow \infty$.

Here, the variance of $\tilde{\beta}_n$ is partitioned into two sources of uncertainty, that due to the Poisson point process model generating the events, and that due to the random sampling of the covariates. The large-sample variance of $\tilde{\beta}_n$ involves integrals of functions evaluated at all sites, or all pairs of sites in the sample domain. As shown in Appendix B, unbiased estimates may be constructed as functions of covariates at the locations of the events, the locations of the covariate samples, or both. Since the latter uses more information, however, it should also yield the most precise estimates of these integrals.

A consistent estimator $\tilde{\beta}_n$ of β_0 may be substituted into (9) to obtain the plug-in estimator $\tilde{\sigma}_n^2(\tilde{\beta}_n)$ for the variance of the random effect. The consistency of the plug-in estimator is proved in Theorem 3:

Theorem 3

Suppose that Assumptions 1–3, 5 and 6 are satisfied, then $\tilde{\sigma}_n^2(\tilde{\beta}_n) \rightarrow \sigma_0^2$ almost surely as $n \rightarrow \infty$.

This theorem is proved in Appendix A.

4. Ecological Momentary Assessment of Smoking

The methods developed here are illustrated using baseline data from a nicotine-patch trial designed to investigate the psychological processes that mediate the effects of treatment on smoking lapse. A total of 412 smokers were given electronic diaries, upon which they were instructed to record each cigarette smoked (hereafter referred to as an “event”) over a 16-day period prior to a designated quit date. The first three days of monitoring were designated to allow participants to acclimate to the use of the diary; data from these days were not used in the analyses. The diary prompted subjects to complete assessments of mood and environment at randomly selected event times and random times are chosen according to a stratified sampling design, stratified by day. Assessments of mood were on an 11-point scale, and environmental variables include whether or not the smoker was alone, whether or not other smokers were present, and smoking prohibitions with three levels: prohibited, discouraged, and allowed. Positive affect, negative affect and attention disturbance factors were constructed from mood variables as described in Ferguson et al. [10], while restlessness is retained as a single item measured on an 11-point scale. Details regarding the study design may be found in Shiffman et al. [36].

To reduce the burden on study subjects, not all cigarettes were assessed for the covariates. To target 5 cigarette assessments per day, smoking events for subject i on day j were independently selected to be assessed with known probability

$$p_{ij} = \min\{5N_{i,j-1}^{-1}, 1\}, \quad (16)$$

where $N_{i,j-1}$ is the number of events for subject i on day $j-1$. Conditional on $N_{i,j-1}$ and u_i , the thinned point process of assessed events is realized from a Poisson process with intensity $p_{ij}u_i \exp\{\beta^T z_i(t)\}$. Then β may be estimated by solving the estimating equation $\tilde{\Psi}_n(\beta) = 0$, where

$$\tilde{\Psi}_n(\theta) = n^{-1} \sum_{i=1}^n \left[\sum_{j=1}^{d_i} \left\{ \sum_{t \in X_{ij}} z_i(t) w_{ij}(t; \beta) - p_{ij} \sum_{t \in D_{ij}} \frac{z_i(t) w_{ij}(t; \beta) \exp\{\beta^T z_i(t)\}}{\pi_{ij}(t)} \right\} \right]$$

[28]. Here, the sums are over the n subjects, the d_i days that subject i was sampled, and the respective sets X_{ij} and D_{ij} of event times and covariate sample times for subject i on day j . For the stratified random sampling design used to sample time-varying covariates, the inclusion densities are $\pi_{ij}(t) = m_{ij}/|A_{ij}|$, where m_{ij} is the number of elements in D_{ij} and $|A_{ij}|$ is the length of time (in hours) that the diary was active for subject i on day j . Two weighting schemes were considered, the unweighted scheme where $w_{ij}(t; \beta) = 1$, and weights set to $w_{ij}(t; \beta) = \pi_{ij}(t) / \{\pi_{ij}(t) + \pi_{ij} \exp\{\beta^T z_i(t)\}\}$ as in Waagepetersen [45].

Table 1 compares the results for fixed and random effects models under both weighting schemes. Effects of the three levels of smoking prohibition in Table 1 are computed under the mean levels of the mood and remaining environmental variables. Both fixed and random effects models obtain the same point estimates for these parameters as they use the same estimating equations under their respective weighting schemes. For both weighting schemes, standard errors were substantially increased under the random effects model with appropriate reductions in significance. This result indicates that failure to account for variation in smoking rates among smokers can result in false positive significance tests.

Whereas negative affect and attention were significant in the fixed effects model, these variables were not significant under the random effects model. Significantly positive coefficients indicated that smoking rate increased with increasing restlessness, when smokers were alone, and when other smokers were present under the random effects model. Conversely, negative coefficients indicated smoking rate decrease with increasing arousal and smoking prohibitions under the random effects model.

For both fixed and random effects models, standard errors were marginally smaller under the weighted estimating equations than under the unweighted estimating equations. This result is consistent with Waagepetersen's [45] findings for the spatial Poisson point process model, and is a consequence of the fact that the weighted estimating equations use covariate information from the realization of the point process as well as the random sample to estimate $\Lambda_i^{(1)}(A_i; \beta)$, whereas the unweighted equations only use information from the random sample to estimate this quantity. Moreover, the estimate of the variance of the random effects was smaller under the weighted estimating equations, also suggesting a better fit.

Table 2 shows the estimated hazard ratios under the random effects model, computed by taking the exponential function of the parameter estimates computed under the weighted estimating equations. The results suggest that the environmental variables had the strongest effects on smoking rates. After adjusting for the effects of the remaining variables, smokers smoked an estimated 0.535 cigarettes per hour (cph) when smoking was prohibited, 0.892 cph when smoking was discouraged, and 1.354 cph when smoking was allowed. Smoking rates were an estimated 1.642 times higher when other smokers were present, and 1.325 times higher when the smoker was alone. Others' smoking could influence the smoker by providing visual, olfactory or social cues encouraging a smoker to light up [35], and laboratory studies have found that the sight or smell of cigarettes [31] and the sight of someone smoking [26,46] can elicit craving for cigarettes. Two mood variables also showed significant effects; smoking rate increased by 7.8% for every unit increase in restlessness, and decreased by 1.8% for every unit increase in arousal. The latter suggests that the proposed approach can detect even very subtle effects of time-varying covariates on smoking rate. Subjectively-experienced restlessness may reflect the undirected activation of motivational systems that drive towards smoking [33]. Given the strength of the effects of the environmental variables, smoking cessation may be more successful if smokers are advised to spend more time in locations where smoking is prohibited, and with other people who are not smoking.

5. Discussion

The proposed approach for statistical inference for frailty models with partially observed time-varying covariates is offered as an alternative to the joint modeling approach. One limitation is that it requires that time-varying covariates be observed at times generated from a probability-based sampling design, and so, it is most suitable for EMA studies where the implementation of probability-based designs is routine. Since no covariate model is required, the proposed approach also does not require any modeling assumptions regarding the covariates so it is easier to implement than the joint modeling approach. While the joint modeling approach can be sensitive to violations of modeling assumptions regarding the time-varying covariates [29], a careful statistician may employ appropriate diagnostic tools to arrive at a correctly specified model provided that the data are adequate to capture the temporal variability in the covariates. A variety of models are available for time-varying covariates under the joint modeling approach including polynomial, spline, and stochastic process models [43] adding considerable flexibility to the approach. However, the joint modeling approach assumes that the covariates are smooth continuous functions of time, an

assumption that may not be satisfied by the behavioral data considered in EMA studies, where covariates can change abruptly as a result of events occurring during a subject's day.

The frailty model considered in this paper implicitly assumes that the effect of time-varying covariates on smoking rate does not vary among smokers. If we can identify what specific situational factors impact individual smoking behaviors, those factors may be targeted for more effective individual-based smoking cessation treatment. Although a number of authors have mentioned the possibility in passing (e.g., Cook and Lawless [5]; p. 36), very little attention has been given to random covariate effects. Ha and Lee [13] briefly consider hierarchical likelihood estimation of random covariate effects, but do not consider time-varying covariates. In a series of papers, Zeng and Lin [50–52] consider transformation models with random covariate effects that are time dependent. In our future work, we will extend our approach to recurrent events models with random covariate effects that are time dependent.

Acknowledgments

This work was supported by NSF grant SES-0720195 and NIH grant IRO1DA024687. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation, the National Institutes of Health, the Food and Drug Administration or the U.S. Government.

References

- Andersen PK, Klein JP, Knudsen K, Palacios RT. Estimation of variance in Cox's regression model with shared gamma frailties. *Biometrics*. 1997; 53:1475–1484. [PubMed: 9423262]
- Billingsley, P. Probability and Measure. Wiley; New York: 1986.
- Brillinger DR. Statistical inference for stationary point processes. *Stochastic Processes and Related Topics*. 1975; 1:55–99.
- Chung, KL. A Course in Probability Theory. 2. Academic Press; New York: 1974.
- Cook, RJ.; Lawless, JF. The Statistical Analysis of Recurrent Events. Springer; New York: 2007.
- Cordy CB. An extension of the Horvitz-Thompson theorem to point sampling from a continuous universe. *Statistics and Probability Letters*. 1993; 18:353–362.
- Cox DR. Some statistical methods connected with series of events. *Journal of the Royal Statistical Society B*. 1955; 17:129–164.
- Cox, DR. The statistical analysis of dependencies in point processes. In: Lewis, PAW., editor. *Stochastic Point Processes*. Wiley; New York: 1972.
- Cressie, N. Statistics for spatial data. Wiley; New York: 1991.
- Ferguson SG, Shiffman S, Gwaltney CJ. Does reducing withdrawal severity mediate nicotine patch efficacy? A randomized clinical trial. *Journal of Consulting and Clinical Psychology*. 2006; 74:1153–1161. [PubMed: 17154744]
- Godambe VP, Thompson ME. Estimating equations in the presence of a nuisance parameter. *Annals of Statistics*. 1974; 2:568–571.
- Guan Y, Loh JM. A thinned block bootstrap variance estimation procedure for inhomogenous spatial point patterns. *Journal of the American Statistical Association*. 2007; 102:1377–1386.
- Ha ID, Lee Y. Estimating frailty models via Poisson hierarchical generalized linear models. *Journal of Computational and Graphical Statistics*. 2003; 12:663–681.
- Henderson R, Oman P. Effect of frailty on marginal regression estimates in survival analysis. *Journal of the Royal Statistical Society B*. 1999; 61:367–379.
- Hougaard P. Life table methods for heterogeneous populations: Distributions describing the heterogeneity. *Biometrika*. 1984; 71:75–83.
- Kassel JD, Stroud LR, Paronis CA. Smoking, stress, and negative affect: Correlation, causation, and context across stages of smoking. *Psychological Bulletin*. 2003; 129:270–304. [PubMed: 12696841]

17. Krickeberg, K. Processus ponctuels en statistique. In: Hennequin, PL., editor. Ecole d'Eté de Probabilités de Sainte-Flour X. International Cooperative Publishing House; 1982.
18. Kutoyants, YA. Parameter Estimation for Stochastic Processes. Heldermann Verlag; Berlin: 1984.
19. Lawless JF. Regression methods for Poisson process data. *Journal of the American Statistical Association*. 1987; 82:808–815.
20. Lawless JF. Negative binomial and mixed Poisson regression. *Canadian Journal of Statistics*. 1987; 15:209–225.
21. Lawless JF, Zhan M. Analysis of interval-grouped recurrent-event data using piecewise constant rate functions. *Canadian Journal of Statistics*. 1998; 26:549–565.
22. Lee Y, Nelder JA. Hierarchical generalized linear models. *Journal of the Royal Statistical Society B*. 1996; 58:619–678.
23. Lee Y, Nelder JA. Hierarchical generalized linear models: A synthesis of generalized linear models, random- effect models and structured dispersions. *Biometrika*. 2001; 88:987–1006.
24. Liu L, Huang XL. Joint analysis of correlated repeated measures and recurrent events processes in the presence of death with application to a study on acquired immune deficiency syndrome. *Journal of the Royal Statistical Society C*. 2009; 58:65–81.
25. McCullagh, P.; Nelder, JA. Generalized Linear Models. 2. New York: Chapman and Hall; 1989.
26. McDermut W, Haaga DAF. Effect of stage of change on cue relativity in continuing smokers. *Experimental and Clinical Psychopharmacology*. 1998; 6:316–324. [PubMed: 9725115]
27. Rathbun SL, Cressie N. Asymptotic properties of estimators for the parameters of spatial inhomogeneous Poisson point processes. *Advances in Applied Probability*. 1994; 26:122–154.
28. Rathbun SL, Shiffman S, Gwaltney CJ. Modelling the effects of partially observed covariates on Poisson process intensity. *Biometrika*. 2007; 94:153–165.
29. Rathbun, SL.; Song, X.; Neustifter, B.; Shiffman, S. Probability sampling framework for modeling the impact of time-varying covariates on lifetimes to events. in review
30. Royden, HL. Real Analysis. McMillan Publishing Co; New York: 2006.
31. Sayette M, Hufford M. Effects of cue exposure and deprivation on cognitive resources in smokers. *Journal of Abnormal Psychology*. 1994; 103:812–818. [PubMed: 7822584]
32. Schoenberg FP. Consistent parametric estimation of the intensity of a spatio-temporal point process. *Journal of Statistical Planning and Inference*. 2005; 128:79–93.
33. Shiffman S, Waters AJ. Negative affect and smoking lapses: A prospective analysis. *Journal of Consulting and Clinical Psychology*. 2004; 72:192–201. [PubMed: 15065954]
34. Shiffman, S.; Stone, AA. Ecological Momentary Assessment: A new tool for behavioral medicine research. In: Krantz, DS.; Baum, A., editors. Technology and Methods in Behavioral Medicine. Lawrence Erlbaum Associates; Philadelphia: 1998.
35. Shiffman S, Paty JA, Gnys M, Kassel JD, Hickcox M. Immediate antecedents of cigarette smoking: An analysis from ecological momentary assessment. *Journal of Abnormal Psychology*. 2002; 111:531–545. [PubMed: 12428767]
36. Shiffman S, Scharf DM, Shadel WG, Gwaltney CJ, Dang Q, Paton SM. Analyzing milestones in smoking cessation: Illustration in a nicotine patch trial in adult smokers. *Journal of Consulting and Clinical Psychology*. 2006; 74:276–285. [PubMed: 16649872]
37. Smyth J, Wonderlich S, Miltenberger R, Mitchell J, Rorty M. The use of ecological momentary assessment approaches in eating disorder research. *International Journal of Eating Disorders*. 2001; 30:83–95. [PubMed: 11439412]
38. Stone AA, Shiffman S. Capturing momentary, self-report data: A proposal for reporting guidelines. *Annals of Behavioral Medicine*. 2002; 24:236–243. [PubMed: 12173681]
39. Stone AA, Broderick JB, Porter L, Kaell AT. The experience of rheumatoid arthritis pain and fatigue: Examining momentary reports and correlates over one week. *Arthritis Care and Research*. 1997; 10:185–193. [PubMed: 9335630]
40. Stone AA, Schwartz JE, Neale JM, Shiffman S, Marco CA, Hickcox M, Paty J, Porter LS, Cruise LJ. A comparison of coping assessed by ecological momentary assessment and retrospective recall. *Journal of Personality and Social Psychology*. 1998; 74:1670–1680. [PubMed: 9654765]

41. Stone AA, Broderick JB, Shiffman S, Schwartz JE. Understanding recall of weekly pain from a momentary assessment perspective: Absolute agreement, between- and within-person consistency, and judged change in weekly pain. *Pain*. 2004; 107:61–69. [PubMed: 14715390]
42. Thall PF. Mixed Poisson likelihood regression models for longitudinal interval count data. *Biometrics*. 1988; 44:197–209. [PubMed: 3358988]
43. Tsiatis AA, Davidian M. Joint modeling of longitudinal and time-to-event data: An overview. *Statistica Sinica*. 2004; 14:809–834.
44. Vaupel JW, Manton KG, Stallard E. The impact of heterogeneity in individual frailty on the dynamics of mortality. *Demography*. 1979; 16:439–454. [PubMed: 510638]
45. Waagepetersen R. Estimating functions for inhomogeneous spatial point processes with incomplete covariate data. *Biometrika*. 2008; 95:351–363.
46. Warren CA, McDonough BE. Event-related brain potentials as indicators of smoking cue-reactivity. *Clinical Neurophysiology*. 1999; 110:1570–1584. [PubMed: 10479024]
47. Wulfsohn MS, Tsiatis AA. A joint model of survival and longitudinal data measured with error. *Biometrics*. 1997; 53:330–339. [PubMed: 9147598]
48. Yuan KH. A theorem on uniform convergence of stochastic functions with applications. *Journal of Multivariate Analysis*. 1997; 62:100–109.
49. Yuan KH, Jennrich RI. Asymptotics of estimating equations under natural conditions. *Journal of Multivariate Analysis*. 1998; 65:245–260.
50. Zeng D, Lin DY. Maximum likelihood estimation in semiparametric regression models with censored data. *Journal of the Royal Statistical Society B*. 2007; 69:507–564.
51. Zeng D, Lin DY. Semiparametric transformation models with random effects for recurrent events. *Journal of the American Statistical Association*. 2007; 102:167–180.
52. Zeng D, Lin DY. Semiparametric transformation models with random effects for joint analysis of recurrent and terminal events. *Biometrics*. 2009; 65:746–752. [PubMed: 18945267]
53. Zhang HP, Ye YQ, Diggle P, Shi J. Joint modeling of time series measures and recurrent events and analysis of the effects of air quality on respiratory systems. *Journal of the American Statistical Association*. 2008; 103:48–60.

Appendix A

This appendix contains proofs of the lemmas required to demonstrate the consistency and asymptotic normality of $\tilde{\beta}_n$. In addition, we will prove the consistency of $\tilde{\sigma}_n^2$.

Proof of Lemma 1

For each β , $\tilde{\Psi}_n(\beta)$ is the sample mean of independent zero-mean random variables whose variances share a common upper bound under Assumptions 1–3, 5 and 6. So by Theorem 5.1.2 of Chung [4], $\tilde{\Psi}_n(\beta)$ converges pointwise to a nonstochastic limit for each $\beta \in \Theta$ with probability 1. Moreover, since $\tilde{\Psi}_n(\beta)$ is an unbiased estimating function, $\tilde{\Psi}_n(\beta_0)$ converges to zero with probability one.

To obtain uniform convergence, partition $\tilde{\Psi}_n(\beta) - E_{\beta_0}\{\tilde{\Psi}_n(\beta)\} = h_n(\beta) - g_n(\beta)$, where

$$h_n(\beta) = n^{-1} \sum_{i=1}^n \left\{ \sum_{t \in X_i} z_i(t) w(t; \beta) - \int_{A_i} z_i(t) w(t; \beta) \exp\{\beta_0^T z_i(t)\} dt \right\}$$

and

$$g_n(\beta) = n^{-1} \sum_{i=1}^n \left\{ \sum_{t \in D_i} \frac{z_i(t) w(t; \beta) \exp\{\beta^T z_i(t)\}}{\pi_i(t)} - \int_{A_i} z_i(t) w(t; \beta) \exp\{\beta^T z_i(t)\} dt \right\}.$$

First we will demonstrate that $h_n(\beta)$ is equicontinuous in β . Take any $\varepsilon > 0$. Since the weight function is bounded under Assumption 3, there exists $k < \infty$ such that $|w(t; \beta_0) - w(t; \beta)| < k/2$.

For each $\beta \in \mathcal{S}(\beta_0, \varepsilon)$,

$$\begin{aligned} h_n(\beta_0) - h_n(\beta) &= n^{-1} \sum_{i=1}^n \left\{ \sum_{t \in X_i} z_i(t) (w(t; \beta_0) - w(t; \beta)) - \int_{A_i} z_i(t) \{w(t; \beta_0) - w(t; \beta)\} \exp\{\beta_0^T z_i(t)\} dt \right\} \\ &\leq \frac{k}{2n} \sum_{i=1}^n \left\{ \sum_{t \in X_i} |z_i(t)| - \int_{A_i} |z_i(t)| \exp\{\beta_0^T z_i(t)\} dt \right\} + \frac{k}{n} \sum_{i=1}^n \int_{A_i} |z_i(t)| \exp\{\beta_0^T z_i(t)\} dt. \end{aligned}$$

In the expression above, $|z_i(t)|$ denotes the absolute value of $z_i(t)$, element by element. The first term in the expression above converges to zero with probability 1, while the second term converges to

$$\delta = \lim_{n \rightarrow \infty} \frac{k}{n} \sum_{i=1}^n \int_{A_i} |z_i(t)| \exp\{\beta_0^T z_i(t)\} dt.$$

The existence of this limit is guaranteed under Assumptions 1 and 2. Therefore, $h_n(\beta)$ is equicontinuous in β . So by Yuan's [48] Lemma, $h_n(\beta)$ converges uniformly to zero as $n \rightarrow \infty$. A similar approach can be used to demonstrate the uniform convergence of $g_n(\beta)$.

Proof of Lemma 2

The derivative of $\tilde{\Psi}_n^{(1)}(\beta)$ with respect to β is

$$\tilde{\Psi}_n^{(1)}(\beta) = n^{-1} \sum_{i=1}^n \left\{ \sum_{t \in X_i} z_i(t) w^{(1)}(t; \beta) - \sum_{t \in D_i} \frac{z_i(t) w^{(1)}(t; \beta) \exp\{\beta^T z_i(t)\}}{\pi_i(t)} \right\} \quad (\text{A.1})$$

$$-n^{-1} \sum_{i=1}^n \sum_{t \in D_i} \frac{z_i(t) z_i(t)^T w(t; \beta) \exp\{\beta^T z_i(t)\}}{\pi_i(t)}, \quad (\text{A.2})$$

which has expectation

$$E_{\beta_0} \{\tilde{\Psi}_n^{(1)}(\beta)\} = n^{-1} \sum_{i=1}^n \int_{A_i} z_i(t) w^{(1)}(t; \beta) \{\exp\{\beta_0^T z_i(t)\} - \exp\{\beta^T z_i(t)\}\} dt - n^{-1} \sum_{i=1}^n \int_{A_i} z_i(t) z_i(t)^T w(t; \beta) \exp\{\beta^T z_i(t)\} dt$$

For each β , $\tilde{\Psi}_n^{(1)}(\beta)$, is the sample mean of independent random variables whose variances share a common upper bound under Assumptions 1–3, 5 and 6. Therefore, $\tilde{\Psi}_n^{(1)}(\beta_0)$ converges pointwise to a nonstochastic limit for each $\beta \in \Theta$ with probability 1. Since the first term of (A.3) is equal to zero when $\beta = \beta_0$, $\tilde{\Psi}_n^{(1)}(\beta_0)$ converges to a negative definite matrix under Assumption 4. Uniform convergence is proved in a manner similar to that used in Lemma 1.

Proof of Lemma 3

Under Assumption 1, we may write $\tilde{\Psi}_n(\beta) = \frac{1}{n} \{Q_n(\beta) - R_n(\beta)\}$, where $Q_n(\beta) = \sum_{i=1}^n q_i(\beta)$ and $R_n(\beta) = \sum_{i=1}^n r_i(\beta)$ are independently distributed,

$$q_i(\beta) = \sum_{t \in X_{ni}} z_i(t) w(t; \beta)$$

and

$$r_i(\beta) = \sum_{t \in X_{ni}} \frac{z_i(t) w(t; \beta) \exp\{\beta^T z_i(t)\}}{\pi_{ni}(t)}.$$

Let a denote any unit vector. Under Assumptions 2 and 3 variance of $a^T Q_n(\beta_0)$ is equal to $a^T \sum_n^Q(\beta_0) a$, which is bounded below by $n\sigma^2 k_1 \sum_{i=1}^n |A_i|^2$ for some constant $k_1 > 0$. Moreover, $E|q_i(\beta)|^4 \leq M_4 k_2 |A_i|^4$, for some constant $k_2 > 0$ under assumptions 2 and 3, where M_4 is the fourth moment of the random effect. Lyapounov's condition (see Billingsley [2], p. 371) then follows since the sizes of the study regions are bounded from below and above, proving that $Q_n(\beta_0)$ converges to a Gaussian random vector with variance $\sum_n^Q(\beta_0)$. Similar arguments can be used to demonstrate that $R_n(\beta)$ also satisfies Lyapounov's condition under Assumptions 1, 2, 3, 5, and 6.

Proof of Theorem 3

The variance of $\tilde{\sigma}_n^2$ is

$$\text{var}_{\theta_0}(\tilde{\sigma}_n^2) = n^{-2} \sum_{i=1}^n |A_i|^{-4} V_i(\tilde{\beta}_n)$$

where

$$\begin{aligned}
V_i(\beta) = & (1 + \sigma_0^2) \int_{A_{ni}} \int_{A_{ni}} \varphi_i(\beta, 2s+2t) \varphi_i(\beta_0, s+t)^{-1} ds dt \\
& + 4M_3 \int_{A_{ni}} \int_{A_{ni}} \int_{A_{ni}} \varphi_i(\beta, 2s+t+u) \varphi_i(\beta_0, s+t+u)^{-1} ds dt du \\
& + (M_4 - (1 + \sigma_0^2)^2) \int_{A_{ni}} \int_{A_{ni}} \int_{A_{ni}} \int_{A_{ni}} \varphi_i(\beta, s+t+u+v) \varphi_i(\beta_0, s+t+u+v)^{-1} ds dt du dv
\end{aligned}$$

Since $\varphi_i(\beta, t)$ is continuous with respect to β , and $\tilde{\beta}_n$ is consistent for β_0 , $V_i(\tilde{\beta}_n) = V_i(\beta_0) + o_p(|A_i|^4)$. By Assumptions 1 and 2, $V_i(\beta_0)$ is bounded above, so $\text{var}_{\theta_0}(\tilde{\sigma}_n^2) \rightarrow 0$, as $n \rightarrow \infty$. The theorem then follows from Chebyshev's inequality.

Appendix B

The large sample variance of $\tilde{\beta}_n$ involves integrals of the forms

$$f(A; \beta) = \int_A f(z(t); \beta) dt \text{ and } f_2(A \times A; \beta) = \int_A \int_A f_2(z(s), z(t); \beta) ds dt;$$

see expressions (13)–(15). Unbiased estimators for these integrals may be constructed as functions of the events of the point process, the probability sample of covariates, or both. Using the former,

$$\widehat{f}(A; \beta) = \sum_{t \in X} \frac{f(z(t); \beta)}{\exp\{\beta^T z(t)\}} \text{ and } \widehat{f}_2(A \times A; \beta) = \sum_{(s, u) \in X^\#} \frac{f(z(s), z(t); \beta)}{(\sigma^2 + 1) \exp\{\beta^T (z(s) + z(t))\}},$$

are unbiased estimators for $f(A; \beta)$ and $f_2(A \times A; \beta)$, respectively, where the sum in the second expression is over the set $X^\# = \{s, u \in X: s \neq u\}$ of all pairs of events of the point process. Alternatively, the data from the probability sample of the covariates may be used to obtain the estimators

$$\widehat{f}(A; \beta) = \sum_{t \in D} \frac{f(z(t); \beta)}{\pi(t)} \text{ and } \widehat{f}_2(A \times A; \beta) = \sum_{(s, u) \in D^\#} \frac{f(z(s), z(t); \beta)}{\pi(s, t)},$$

where the second term is obtained by summing over the set $D^\# = \{s, u \in X: s \neq u\}$ of all pairs of sample points in the probability sample. More precise estimators may be obtained using both the events of the point process and the sample samples yielding the estimators

$$\widehat{f}(A; \beta) = \sum_{t \in X \cup D} \frac{f(z(t); \beta)}{\exp\{\beta^T z(t)\} + \pi(t)} \text{ and } \widehat{f}_2(A \times A; \beta) = \sum_{(s, u) \in X^\# \cup D^\#} \frac{f_2(z(s), z(t); \beta)}{(\sigma^2 + 1) \exp\{\beta^T (z(s) + z(t))\} + \pi(s, t)}.$$

Since the latter uses more information, it should yield more precise estimates of the large-sample variance of $\tilde{\beta}_n$. If the covariates are sampled according to a point process with intensity $\pi(t)$, then $\pi(s, t) = \pi(s)\pi(t)$.

Table 1

Weighted and unweighted estimates of point process model parameters for fixed and random effects models

Weights	Covariate	Fixed Effects			Random Effects		
		Estimate	SE	p-value	SE	p-value	
Unweighted	Restlessness	0.0762	0.0045	0.0001	0.0099	0.0001	
	Negative Affect	-0.0346	0.0073	0.0001	0.0194	0.0745	
	Positive Affect	-0.0042	0.0063	0.5030	0.0187	0.8213	
	Arousal	-0.0177	0.0039	0.0001	0.0085	0.0376	
	Attention	-0.0214	0.0066	0.0011	0.0175	0.2194	
	Alone	0.2773	0.0166	0.0001	0.0327	0.0001	
	Other Smokers	0.5135	0.0251	0.0001	0.0493	0.0001	
	Prohibited *	-0.6428	0.0230	0.0001	0.0604	0.0001	
	Discouraged *	-0.1501	0.0252		0.0779		
	Allowed *	0.2275	0.0099		0.0373		
	σ^2				0.6906		
Weighted	Restlessness	0.0752	0.0043	0.0001	0.0094	0.0001	
	Negative Affect	-0.0251	0.0071	0.0004	0.0189	0.1840	
	Positive Affect	-0.0019	0.0059	0.7449	0.0173	0.9113	
	Arousal	-0.0180	0.0037	0.0001	0.0080	0.0235	
	Attention	-0.0153	0.0064	0.0175	0.0172	0.3737	
	Alone	0.2813	0.0168	0.0001	0.0335	0.0001	
	Other Smokers	0.4958	0.0249	0.0001	0.0530	0.0001	
	Prohibited *	-0.6254	0.0229	0.0001	0.0563	0.0001	
	Discouraged *	-0.1139	0.0247		0.0725		
	Allowed *	0.3032	0.0101		0.0363		
	σ^2				0.5045		

* Estimates are for log smoking rates under the levels of smoking prohibition and mean values of mood variables from the probability sample.

Table 2

Risk ratios for the effects of mood variables on smoking rate. Smoking rates in cigarettes per hour are given for the three levels of smoking prohibition

Covariate	Risk Ratio	95% Confidence Interval
Restlessness	1.078	1.059, 1.098
Negative Affect	0.975	0.940, 1.012
Positive Affect	0.998	0.965, 1.032
Arousal	0.982	0.967, 0.998
Attention	0.985	0.952, 1.019
Alone	1.325	1.241, 1.415
Other Smokers	1.642	1.480, 1.822
Prohibited	0.535	0.479, 0.598
Discouraged	0.892	0.774, 1.029
Allowed	1.354	1.261, 1.454