University of California, Berkeley
U.C. Berkeley Division of Biostatistics Working Paper Series

Year 2004 *Paper* 148

Mean Response Models of Repeated Measurements in Presence of Varying Effectiveness Onset

Ying Qing Chen* Su-Chun Cheng†

*Division of Biostatistics, School of Public Health, University of California, Berkeley, yqchen@stat.berkeley.edu

†Department of Epidemiology & Biostatistics, University of California, San Francisco, scheng@biostat.ucsf.edu

http://biostats.bepress.com/ucbbiostat/paper148

Mean Response Models of Repeated Measurements in Presence of Varying Effectiveness Onset

Ying Qing Chen and Su-Chun Cheng

Abstract

Repeated measurements are often collected over time to evaluate treatment efficacy in clinical trials. Most of the statistical models of the repeated measurements have been focusing on their mean response as function of time. These models usually assume that the treatment has persistent effect of constant additivity or multiplicity on the mean response functions throughout the observation period of time. In reality, however, such assumption may be confounded by the potential existence of the so-called effectiveness action onset, although they are often unobserved or difficult to obtain. Instead of including nonparametric time-varying coefficients in the mean response models, we propose and study some semiparametric mean response models to accommodate such effectiveness times. Our methodologies will be demonstrated by a real randomised clinical trial data.

1 Introduction

In some randomised clinical trials, repeated measurements of same subject are collected over time and compared to evaluate the efficacy of a new treatment. For example, a randomised clinical trial was conducted to evaluate the treatment efficacy between Buprenorphine and Methadone in reducing opiate use among a total of 162 addicts (Johnson, Jaffe and Fudala, 1992). In this trial, the repeated measurements of whether or not a subject failed a urine test were collected at 3 visits per week over a 17 week period. Another example was a randomised clinical trial to evaluate the treatment efficacy of memantine in the management of painful peripheral neuropathy in diabetic patients. The weekly repeated measurements of visual analog score (VAS) nocturnal pain intensity were collected over a 16-week follow-up period. More examples can be found in Albert (1999) and the book by Diggle, et al. (2001).

Assume that $(Y_1, Y_2, \ldots, Y_m)$ is the vector of the repeated measurements collected for some subject. They can be considered as observations of an underlying random response curve over time, $\{Y(t); t \geq 0\}$, observed at the finite number of time points of $(t_1, t_2, \ldots, t_m)$, where $Y(t_j) = Y_j$, $j = 1, 2 \ldots, m$. In the statistical literature, the means of these response curves as function of time have been studied in regression settings, and can be used to evaluate the treatment efficacy in the randomised clinical trials. For example, one such model was first proposed in Zeger and Diggle (1994) and later generalised in Lin and Ying (2001),

$$E\{Y(t) \mid Z(s); 0 \leq s \leq t\} = \mu(t) + \boldsymbol{\beta}^T \boldsymbol{Z}(t), \tag{1}$$

where $\boldsymbol{Z}(\cdot)$ is the covariate vector, $\boldsymbol{\beta} \in \mathbf{R}^p$ is the associated regression coefficient, $\mu(\cdot)$ is some unspecified function and $\varepsilon(\cdot)$ is zero-mean stationary Gaussian process. Here, ^{T} denotes vector transpose. When $\boldsymbol{Z}(t)$ is the treatment indicator, the parameter $\boldsymbol{\beta}$ can be used to characterise the treatment efficacy. The unspecified $\mu(t)$ in (1) would allow more flexibility when the treatment efficacy parameter $\boldsymbol{\beta}$ is of major interest.

To explore the mean response curves in the memantine trial, for example, the lowess curves of the VAS nocturnal pain intensity were plotted in Figure 1 for both the memantine and the placebo groups. In appearance, the memantine shows treatment efficacy to lower the average pain intensity curve consistently during the trial period. Further examination, however, finds that the two curves are relatively close to each other at the beginning of drug application and toward the trial completion as well. This might suggest that the treatment does not have the constant effect of proportionality as needed in model (1). In fact, if the primary endpoint is chosen to be the change from baseline during the 16-week period, most

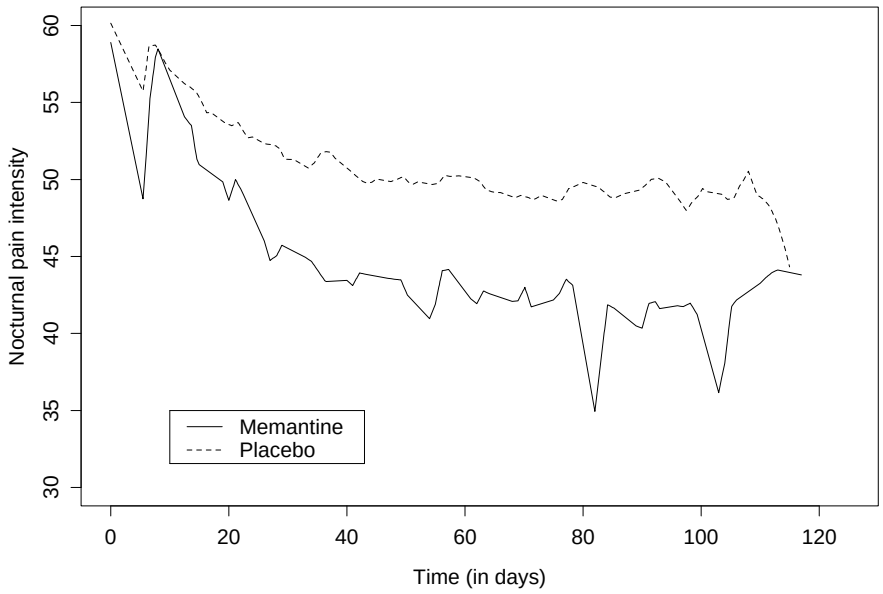

Fig. 1: Lowess curves of mean nocturnal pain intensity of memantine and control treatment groups

of the test statistics would fail to show significant results of treatment efficacy. Even if the robust estimation methods such as the Generalized Estimation Equations (GEE) are applied to the all the collected repeated measurements, the estimated $\widehat{\boldsymbol{\beta}}$ tends to have less power in detecting the treatment efficacy due to misspecified mean response.

In reality, there may be some practical reasons that lead to the nonconstant treatment effect as shown in Figure 1. Specifically in clinical drug development, it has long been understood that for some drugs, there are usually three distinct periods of drug action since the time of administration: pre-action onset, action onset, and post-action onset. Accordingly, there may be a pair of time points, $0 \leq U < V$, such that he treatment is usually only fully effective within the action onset time period of $[U, V]$, while not so for the pre-action onset period of $(0, U)$ or post-action period of (V, ∞). Due to the heterogeneity in human metabolism of drug compound, however, it is often difficult to determine or obtain the $[U, V]$'s without knowing the actual pharmaco-kinetic and pharmaco-dynamic profiles of the compound for all the individuals in the trials. The patterns of nonconstant treatment effect thus appears in the mean response curves by ignoring such $[U, V]$'s. In practice, the U and V are often called the treatment effectiveness lag time and saturation time, respectively.

As an alternative, it is mathematically convenient to extend mode1 (1) by including the time-varying coefficients, $\boldsymbol{\beta}(t)$,

$$E\{Y(t) \mid \boldsymbol{Z}(s), 0 \leq s \leq t\} = \mu(t) + \boldsymbol{\beta}(t)^{\tau}\boldsymbol{Z}(t), \tag{2}$$

similar to that in Hoover, et al. (1998). Various estimation methods, such as smoothing splines and locally weighted polynomials in Hoover, et al. (1998), component-wise smoothing spline in Chiang, Rice and Wu (2001) and basis function approximation in Huang, Wu and Zhou (2002), can be adapted to estimate $\boldsymbol{\beta}(\cdot)$. However, as an infinite-dimensional parameter, $\boldsymbol{\beta}(\cdot)$ itself usually lacks straightforward interpretation as treatment efficacy, and as a result, limits its practical application to establish guidelines in drug approval for the regulatory agencies or drug prescription for the clinical practitioners.

In this article, we will propose some new models based on model (1) to accommodate (U, V) and hence account for the unobserved treatment effectiveness lag and saturation times. The extended models will be proposed in §2. The semiparametric inference procedures and the associated statistical properties will be studied in §2.3. Further model extensions will be in §3. Numerical studies are presented in §4. Some concluding remarks and discussion will be in §5.

2 Methods

2.1 *Models for bivariate action onset times*

Suppose there are n subjects in a study. For the subject i, denote U_i and V_i the treatment effectiveness lag and saturation times, respectively, where $0 \leq U_i < V_i$, $i = 1, 2, \ldots, n$. Let $W_i = V_i - U_i$, which is the length of time interval of the action onset. Consider the bivariate vectors of (U_i, W_i) for the action onset times. Denote their joint density function $f_{U,W}(u, w; \theta)$. Since U_i and W_i reflect a human subject's individual reaction to the treatment, they are often neither identical nor independent among individuals.

To choose appropriate distributions for (U, W), we consider the widely used shared frailty models for the bivariate times of (U, W). Assume γ_i is the frailty of the ith subject, $i = 1, 2, \ldots, n$, following the distribution with density function of $g(\gamma; \boldsymbol{\alpha})$, where $\boldsymbol{\alpha}$ is the parameter. Conditional on γ_i, the hazard functions for U_i and W_i are $\gamma_i \lambda_U(t)$ and $\gamma_i \lambda_W(t)$, respectively. Thus the bivariate survival function of (U, W) is,

$$S_{U,W}(u, w \mid \gamma_i) = \text{pr}\{U > u, W > w \mid \gamma_i\} = \exp[-\gamma_i\{\Lambda_U(u) + \Lambda_W(w)\}],$$

where $\Lambda_U(t) = \int_0^t \lambda_U(s)ds$ and $\Lambda_W(t) = \int_0^t \lambda_W(s)ds$, respectively. By integrating out the γ_i's, the marginal bivariate survival function for (U, W) is $S_{U,W}(u, w) = E \exp[-\gamma_i\{\Lambda_U(u) +$

$\Lambda_W(w)\}]$, which is also the Laplace transform of γ_i's distribution, $L(\cdot)$, at $\Lambda_U(u) + \Lambda_W(w)$. As a result,

$$f_{U,W}(u, w) = \lambda_U(u)\lambda_W(w)L^{(2)}\{\Lambda_U(u) + \Lambda_W(w)\}.$$

With different choices of $g(\cdot)$, $\lambda_U(\cdot)$ and $\lambda_W(\cdot)$, $f_{U,W}(\cdot, \cdot)$ embraces a variety of choices of bivariate distributions for (U, W).

One widely used family is the Clayton-Oakes model, or the Gamma frailty model (Clayton, 1978; Oakes, 1989). In this model, γ_i are assumed to follow the Gamma density function,

$$g(\gamma; \alpha) = \frac{(\gamma/\alpha_2)^{\alpha_1 - 1} \exp(-\gamma/\alpha_2)}{\alpha_2 \Gamma(\alpha_1)},$$

where $\Gamma(\gamma) = \int_0^\infty s^{\gamma-1} \exp(-s)ds$. Therefore, the bivariate density function for (U, W) is thus $\alpha_1(1+\alpha_1)\alpha_2^{-\alpha_1}\lambda_U(u)\lambda_W(w)\{\alpha_2^{-1} + \Lambda_U(u) + \Lambda_W(w)\}^{-\alpha_1-2}$. In practice, given the multiplicative form of the γ_i on $\lambda_U(\cdot)$ and $\lambda_W(\cdot)$, it is usually sensible to further assume that $\alpha_1 = 1/\alpha_2$, which leads to $E\gamma_i \equiv 1$. For other families of bivariate distributions, such as the positive stable frailty distributions, the book by Hougaard (2000, §7) offers a comprehensive account on these distributions.

2.2 *Mean response models with varying action onset*

Suppose there is a finite upper time limit for the study, $\tau > 0$, say. Let $Y_i(t)$ be the underlying random response curve of the ith subject, observed at the set of time points of $(T_{i1}, T_{i2}, \ldots, T_{i,m_i})$. The observed repeated measurements are denoted as $(Y_{i1}, Y_{i2}, \ldots, Y_{i,m_i})$ with $Y_{ij} = Y(T_{ij})$. Let $\boldsymbol{Z}_i(t) = (\boldsymbol{Q}_i(t)^T, \boldsymbol{R}_i(t)^T)^T$ be the associated covariates, which consists of $\boldsymbol{Q}_i(t)$ being the treatment assignment and $\boldsymbol{R}_i(t)$ the prognostic covariates to be adjusted. In a two-arm randomised clinical trial, for instance, $\boldsymbol{Q}_i(\cdot)$ can be 1 if the subject is in the treatment arm and 0 otherwise, while $\boldsymbol{R}_i(\cdot)$ contains the concomitant risk factors or confounding variables.

Based on the additive model of (1), we propose the following model:

$$E\{Y_i(t) \mid \boldsymbol{Z}_i(s), U_i, W_i; 0 \le s \le t\} = \mu(t) + \boldsymbol{\beta}_Q^T \boldsymbol{Q}_i(t) I(U_i \le t \le U_i + W_i) + \boldsymbol{\beta}_R^T \boldsymbol{R}_i(t), \quad (3)$$

where $I(\cdot)$ is the indicator function, and $\boldsymbol{\beta} = (\boldsymbol{\beta}_Q^T, \boldsymbol{\beta}_R^T)^T$ are the parameters of the same dimensions as $(\boldsymbol{Q}_i(t)^T, \boldsymbol{R}_i(t)^T)^T$, respectively. Conditional on (U_i, W_i), the mean response of the repeated measurements is $\mu(t)\exp\{\boldsymbol{\beta}_Q^T \boldsymbol{Q}_i(t) + \boldsymbol{\beta}_R^T \boldsymbol{R}_i(t)\}$ on $[U_i, U_i + W_i]$, and $\mu(t)\exp\{\boldsymbol{\beta}_R^T \boldsymbol{R}_i(t)\}$ otherwise. Thus the parameter $\boldsymbol{\beta}_Q$ describes the differences in the mean responses due to

the effect of $\boldsymbol{Q}_i(t)$ during the action onset period. When $\boldsymbol{Q}_i(t)$ is the treatment indicator, it characterises the actual treatment effect on the individuals.

This model is in fact a changepoint model with two subject-specific changepoints at U_i and $U_i + W_i$, respectively. Many changepoint models, however, focus on hypothesis testing of the fixed changepoints in the analysis of time series, such as Wu, Woodroofe and Mentz (2001). In the proposed model (3), not only does the different (U_i, W_i) reflect the actual biological mechanism, but also allow further modelling in regression settings to estimate the magnitude of the treatment effect. More straightforward algebra shows that the model becomes

$$E\{Y_i(t)|\boldsymbol{Z}_i(s); 0 \leq s \leq t\} = \mu(t) + \boldsymbol{\beta}_Q^T \boldsymbol{Q}_i(t) H(t; \boldsymbol{\theta}) + \boldsymbol{\beta}_R^T \boldsymbol{R}_i(t), \tag{4}$$

when marginalised over (U_i, W_i). Here $H(t; \boldsymbol{\theta}) = \int_0^t \int_{t-u}^{\infty} f_{U,W}(u, w; \boldsymbol{\theta}) dw du$.

As seen in the marginal model of (4), the inclusion of the varying (U_i, W_i)'s induces additional time-dependent structure upon $\boldsymbol{\beta}_Q^T \boldsymbol{Q}_i(t)$, which is modified as $\boldsymbol{\beta}_Q^T \boldsymbol{Q}_i(t) H(t)$. Two perspectives can be applied to view $H(t)$: (1) $\boldsymbol{\beta}_Q H(t)$, termed as "marginal treatment efficacy," is a special form of the time-varying $\boldsymbol{\beta}(t)$ in model (2); (2) the covariates $\boldsymbol{Q}_i(t)$ are "updated" by $H(t)$ and replaced by $\boldsymbol{Q}_i^*(t) = \boldsymbol{Q}_i(t) H(t)$ in model (1). Nevertheless, the constant parameter $\boldsymbol{\beta}_Q$ itself maintains the appealing interpretation in treatment effect of action onset. Furthermore, if the parameters in $H(t)$ can be appropriately estimated, it will enable us to estimate EU_i and EW_i, respectively, which may certainly yield valuable information in predicting the timing of an individual's action onset.

In fact, $H(t)$ itself carries some interesting properties, such as:

1. $0 \leq H(t) \leq 1$;
2. $\lim_{t \to 0} H(t) = \lim_{t \to \infty} H(t) = 0$;
3. $H'(t) = \int_0^t f(u, t-u) du - \int_0^{\infty} f(t, w) dw$; $H^{(2)}(t) = f(t, 0) + \int_0^t \partial f(u, w)/\partial w|_{w=t-u} du - \int_0^{\infty} \partial f(u, w)/\partial u|_{u=t} dw$.

The first property mandates that the marginal treatment efficacy will be no larger than the actual treatment efficacy $\boldsymbol{\beta}_Q$. The second property implies that the treatment efficacy diminishes to null at the beginning of randomised trial for the short-term and also toward the long-run, which echoes the observations of the possible treatment effectiveness lag and saturation. The third property will allow us to calculate the time point when $H(t)$ reaches its peak, and the turning points as well when the overall trend of $H(t)$ changes. In addition,

the lag time of pre-onset action cannot be too long. There should exist at least one subject, $i^* \in \{1, 2, \ldots, n\}$, such that $\text{pr}\{U_i^* \leq \tau\} > 0$; otherwise, $H(t) \equiv 0$ for any $0 \leq t \leq \tau$, which would cause $\boldsymbol{\beta}_{Q}$ to be nonidentifiable.

2.3 *Inference procedures*

Denote $N_i(t) = \sum_{j=1}^{m_i} I(T_{ij} \leq t)$ and assume that $E\{N_i(t)\} = \Omega(t)$ is unspecified. Let C_i be the follow-up time and $\Delta_i(t) = I(C_i \geq t)$. Conditional on $\boldsymbol{Z}_i(\cdot)$, $(Y_i(\cdot), C_i)$ are assumed to be independent. The true parameters hereinafter are denoted as their respective counterparts with the subscript "$_*$." For instance, the true parameters for $\boldsymbol{\beta}_{Q}$ and $\boldsymbol{\beta}_{R}$ in (3) are $\boldsymbol{\beta}_{Q*}$ and $\boldsymbol{\beta}_{R*}$, respectively. Consider the cumulative sum of the repeated measurements on residuals for the ith subject, $X_i(t) = \int_0^t \{Y_i(s) - \nu_i(s)\} dN_i(s)$, where $\nu_i(t; \boldsymbol{\beta}, \boldsymbol{\theta}) = \boldsymbol{\beta}_Q^T \boldsymbol{Q}_i(t) H(t) + \boldsymbol{\beta}_R^T \boldsymbol{R}_i(t)$. Then

$$E\{dX_i(t) \mid \boldsymbol{Z}_i(s), C_i; 0 \leq s \leq t, \boldsymbol{\beta}_*, \boldsymbol{\theta}_*\} = \Delta_i(t) d\Omega_\mu(t), \tag{5}$$

where $d\Omega_\mu(t) = \mu(t) d\Omega(t)$.

Let $M_i(t) = X_i(t) - \int_0^t \Delta_i(s) d\Omega_\mu(s)$. Then $M_i(\cdot; \boldsymbol{\beta}_*, \boldsymbol{\theta}_*)$ are the zero-mean stochastic processes. Similar to those in Lin and Ying (2001), the following estimating equations generalise the normal equations of the least-squares in the linear regression models to estimate the parameters in the proposed model (4),

$$\sum_{i=1}^{n} \int_0^\tau \Delta_i(t) \Psi(t) \boldsymbol{\varphi}_i(t) dM_i(t) = 0, \tag{6}$$

where $\Psi(\cdot)$ is the positive weight function which converges uniformly to a deterministic function $\psi(t) \in [0, \tau]$, and $\boldsymbol{\varphi}_i(t)$ are the smooth functions of the same dimensions as $(\boldsymbol{\beta}^T, \boldsymbol{\theta}^T)^T$ such that $\boldsymbol{\varphi}_i(t)$ are measurable with respect to $\{\boldsymbol{Z}_i(s), C_i; 0 \leq s \leq t, i = 1, 2, \ldots, n\}$. For instance, $\boldsymbol{\varphi}_i(\cdot)$ can be chosen as $\boldsymbol{Z}_i(\cdot)$ and some of its functionals.

In addition to the unknown parameters of $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$ in (6), the infinite-dimensional function of $\Omega_\mu(\cdot)$ is also unknown. An estimator of the Breslow-type, however, can be obtained for $\Omega_\mu(\cdot)$,

$$\widehat{\Omega}_\mu(t) = \int_0^t \frac{\sum_{i=1}^n dX_i(s)}{\sum_{i=1}^n \Delta_i(s)},$$

which is unbiased to $\Omega_\mu(t)$. Let $\widehat{M}_i(t) = X_i(t) - \int_0^t \Delta_i(s) d\widehat{\Omega}_\mu(s)$. Replace the $M_i(\cdot)$'s in (6) and thus result in $\sum_{i=1}^n \int_0^\tau \Delta_i(t) \Psi(t) \boldsymbol{\varphi}_i(t) d\widehat{M}_i(t) = 0$. Straightforward algebra further leads

to

$$\boldsymbol{\mathcal{E}}(\boldsymbol{\beta},\boldsymbol{\theta}) = \sum_{i=1}^{n} \int_0^{\tau} \Delta_i(t)\Psi(t)\left\{\boldsymbol{\varphi}_i(t) - \bar{\boldsymbol{\varphi}}(t)\right\} dX_i(t) = 0, \tag{7}$$

where $\bar{\boldsymbol{\varphi}}(t) = \sum_{i=1}^{n} \Delta_i(t)\boldsymbol{\varphi}_i(t) / \sum_{i=1}^{n} \Delta_i(t)$. Assume that $\widehat{\boldsymbol{\beta}}$ and $\widehat{\boldsymbol{\theta}}$ are the solutions in (7), respectively.

Let $\boldsymbol{\nu}_i'(t)$ be the derivative of $\nu_i(t)$, $i = 1, 2, \ldots, n$. Then $-n^{-1}\boldsymbol{\mathcal{E}}\ (\boldsymbol{\beta}_*, \boldsymbol{\theta}_*)$ goes to

$$D = E\left[\int_0^{\tau} \Delta_1(t)\psi(t)\{\boldsymbol{\varphi}_1(t) - \bar{\boldsymbol{\varphi}}_*(t\}\boldsymbol{\nu}_1'(t)^{\tau} d\Omega(t)\right],$$

where $\bar{\boldsymbol{\varphi}}_*(t)$ is the limit of $\bar{\boldsymbol{\varphi}}(t)$ almost surely, as $n \to \infty$. When $f_{U,W}(u,w)$ degenerates to 1 at $(u,w) = (0,\infty)$ and 0 otherwise, and $\boldsymbol{\varphi}_i(\cdot)$ are chosen to be $\boldsymbol{Z}_i(\cdot)$, the proposed model (4) becomes the model (1) and D reduces to the nonsingular matrix of $\widetilde{D}$ in Lin and Ying (2000). In general, when the elements in $\boldsymbol{\varphi}_i(\cdot)$ are not linearly related, D is nonsingular. Thus under mild conditions, the solutions to $\boldsymbol{\mathcal{E}}(\boldsymbol{\beta},\boldsymbol{\theta}) = 0$ are strongly consistent as $n \to \infty$ as shown in the Appendix. If the total variation of $\boldsymbol{\varphi}_i(\cdot)$, $i = 1, 2, \ldots, n$, are bounded, it is true that

$$n^{-1/2}\boldsymbol{\mathcal{E}}(\boldsymbol{\beta}_*, \boldsymbol{\theta}_*) \doteq n^{-1/2} \sum_{i=1}^{n} \int_0^{\tau} \Delta_i(t)\psi(t)\{\boldsymbol{\varphi}_i(t) - \bar{\boldsymbol{\varphi}}_*(t)\} dM_i(t; \boldsymbol{\beta}_*, \boldsymbol{\theta}_*).$$

By the Central Limit Theorem, it is shown in the Appendix that $n^{-1/2}\boldsymbol{\mathcal{E}}(\boldsymbol{\beta}_*, \boldsymbol{\theta}_*)$ is asymptotically normal with mean zero and the variance-covariance matrix,

$$\Sigma = E\left[\int_0^{\tau} \Delta_1(t)\psi(t)\{\boldsymbol{\varphi}_1(t) - \bar{\boldsymbol{\varphi}}_*(t)\} dM_1(t)\right]^{\otimes 2},$$

where $a^{\otimes 2}$ denotes aa^{τ}. In addition, a Taylor's expansion of $\boldsymbol{\mathcal{E}}(\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\theta}})$ at $(\boldsymbol{\beta}_*, \boldsymbol{\theta}_*)$ yields that $n^{1/2}(\widehat{\boldsymbol{\beta}}^{\tau} - \boldsymbol{\beta}_*^{\tau}, \widehat{\boldsymbol{\theta}}^{\tau} - \boldsymbol{\theta}_*^{\tau})^{\tau}$ is asymptotically equivalent to $\{-\boldsymbol{\mathcal{E}}'(\boldsymbol{\beta}_*, \boldsymbol{\theta}_*)/n\}^{-1} \cdot n^{-1/2}\boldsymbol{\mathcal{E}}(\boldsymbol{\beta}_*, \boldsymbol{\theta}_*)$. As shown in the Appendix, $\widehat{\boldsymbol{\beta}}$ and $\widehat{\boldsymbol{\theta}}$ are consistent, and

$$n^{1/2}\begin{pmatrix} \widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_* \\ \widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_* \end{pmatrix} \to N(0, D^{-1}\Sigma D^{-1})$$

in distribution in a neighbourhood of $(\boldsymbol{\beta}_*, \boldsymbol{\theta}_*)$, where D and Σ can be approximated by their empirical counterparts,

$$\begin{aligned} \widehat{D} &= n^{-1} \sum_{i=1}^{n} \int_0^{\tau} \Delta_i(t)\Psi(t)\{\boldsymbol{\varphi}_i(t) - \bar{\boldsymbol{\varphi}}(t)\}\boldsymbol{\nu}_i'(t)^{\tau} dN_i(t), \text{ and} \\ \widehat{\Sigma} &= n^{-1} \sum_{i=1}^{n} \left[\int_0^{\tau} \Delta_i(t)\Psi(t)\{\boldsymbol{\varphi}_i(t) - \bar{\boldsymbol{\varphi}}(t; \widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\theta}})\} d\widehat{M}_i(t; \widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\theta}})\right]^{\otimes 2}, \end{aligned}$$

respectively.

The estimating equations used in the weighted estimating equations of (7) are somewhat *ad hoc*, although the estimators defined in the equations carry the appealing statistical properties such as consistency and asymptotic normality. It is desirable to choose an optimal weight function to minimize the variance among the estimators. When there is no varying action onset and the variance-covariance structure are identical among subjects, it is straightforward with an application of Cauchy-Schwarz inequality to see that such choice is $1/\text{var}\{Y(t)-\nu(t)\}$, which is the essentially the diagonal elements in the variance-covariance matrix of $Y(\cdot)$, as indicated in Lin and Ying (2001). Since the inclusion of the varying action onset only modifies the mean structure marginally, it does not introduce additional variability on $Y(\cdot)$, so the optimal choice of $\psi(\cdot)$ would improve the efficiency. However, as pointed out in the comments following Lin and Ying (2001) by Wang and Wang (2001), the efficiency should be further improved if the weight function can be selected among the bivariate functions of $\Phi(s,t)$ to account for the covariance of $(Y(s),Y(t))$ for different $s>0$ and $t>0$.

To estimate the baseline $\mu(\cdot)$, it is natural to consider the estimator of

$$\widetilde{\mu}(t)=\bar{Y}(t)-\bar{\nu}(t;\widehat{\boldsymbol{\beta}},\widehat{\boldsymbol{\theta}}),$$

where $\bar{Y}(t)=\sum_{i=1}^{n}\Delta_i(t)Y_i(t)/\sum_{i=1}^{n}\Delta_i(t)$ and $\bar{\nu}(t;\boldsymbol{\beta},\boldsymbol{\theta})=\sum_{i=1}^{n}\Delta_i(t)\nu_i(t;\boldsymbol{\beta},\boldsymbol{\theta})/\sum_{i=1}^{n}\Delta_i(t)$, respectively. This is the pointwise average of $Y_i(t)-\nu_i(t)$ when $\Delta_i(t)=1$, i.e., the subjects are still "at risk." When the observation times are observed in a continuous time scale, some smoothing technique has to be implemented to obtain a reasonable estimate. In Lin and Ying (2001), a simple singleton nearest neighbour smoother was used. This approach may not be the most efficient. But it has advantage "in non-linear, non-Gaussian situations" without constructing explicit smoothers (Rice, 2003). To improve efficiency, however, more sophisticated smoothing techniques such as the one by Capra and Müller (1997) can be can be adapted to estimate $\mu(\cdot)$. Specifically, consider the time interval $[0,\tau]$ is partitioned into L consecutive equidistant intervals: (t_{l-1},t_l), with $l=1,2\ldots,L\to\infty$ and $t_0=0$. Assume the smoothing parameter h such that $h\to 0$ and $n_*h\to 0$, as $n_*\to 0$, where n_* is the total number of observation time points. Then a smoothed estimate of $\widetilde{\mu}(\cdot)$ is

$$\widehat{\mu}(t)=\arg\min_{a_0,a_1}\left[\sum_{l=1}^{L}K\left(\frac{t-t_l}{h}\right)\{\widetilde{\mu}(t_l)-a_0-a_1(t_l-t)\}^2\right].$$

Here $K(s)=1-s^2$, if $|s|\le 1$, and 0 otherwise. Other smoothers including higher-order kernel smoothers or local fitting with high-order polynomials can be also used under the

necessary conditions of linearity, consistency and consistency with needed rate in Capra and Müller (1997).

3 Extensions

3.1 *Multiplicative mean response models*

In addition to the additive model in (1), there is also a parallel multiplicative model proposed in the literature (Cheng and Wei, 2000),

$$E\{Y_i(t) \mid \boldsymbol{Z}_i(s); 0 \le s \le t\} = \mu(t) \exp\{\boldsymbol{\beta}^T \boldsymbol{Z}_i(t)\}, \tag{8}$$

to analyze the repeated measurements. This model is equivalent to the additive model when the response curves are properly transformed, for instance, if the $Y(t)$ in model (8) is log-transformed. However, the Cheng-Wei model also assumes constant treatment effect and may not be appropriate in presence of the action onset times. To include the varying action onset, we propose the following model:

$$E\{Y_i(t) \mid \boldsymbol{Z}_i(s), U_i, W_i; 0 \le s \le t\} = \mu(t) \exp\{\boldsymbol{\beta}_Q^T \boldsymbol{Q}_i(t) I(U_i \le t \le U_i + W_i) + \boldsymbol{\beta}_R^T \boldsymbol{R}_i(t)\}. \tag{9}$$

The marginalised version of this model is thus

$$E\{Y_i(t) \mid \boldsymbol{Z}_i(s); 0 \le s \le t\} = \mu(t) \exp\{\boldsymbol{\beta}_R^T \boldsymbol{R}_i(t)\} \left[\exp\{\boldsymbol{\beta}_Q^T \boldsymbol{Q}_i(t)\} H(t; \boldsymbol{\theta}) + \{1 - H(t; \boldsymbol{\theta})\}\right]. \tag{10}$$

Apparently, $\exp\{\boldsymbol{\beta}_Q^T \boldsymbol{Q}_i(t)\} H(t) + \{1 - H(t)\}$, which is a weight average of $\exp\{\boldsymbol{\beta}_Q^T \boldsymbol{Q}_i(t)\}$ and 1, would approach to 1 as t goes to 0 or ∞. When $\boldsymbol{Q}_i(\cdot)$ is the treatment indicator, this property should better characterise the observed response curves in presence of the potential action onset. Unlike the additive model (4), however, the marginalised multiplicative model does not maintain the linear structure on $\boldsymbol{\beta}_Q$, which may add complexity in estimation.

To estimate the parameters $(\boldsymbol{\beta}, \boldsymbol{\theta})$ in model (9), consider $X_i(t) = \int_0^t Y_i(s) dN_i(s)$. Since

$$E\{dX_i(t) | Z_i(s), C_i; 0 \le s \le t, \boldsymbol{\beta}_*, \boldsymbol{\theta}_*\} = \Delta_i(t) \exp\{\rho_i(t; \boldsymbol{\beta}_*, \boldsymbol{\theta}_*)\} d\Omega_\mu(t),$$

where $\rho_i(t; \boldsymbol{\beta}, \boldsymbol{\theta}) = \boldsymbol{\beta}_R^T \boldsymbol{R}_i(t) + \log[\exp\{\boldsymbol{\beta}_Q^T \boldsymbol{Q}_i(t)\} H(t; \boldsymbol{\theta}) + \{1 - H(t; \boldsymbol{\theta})\}]$, then $M_{\rho,i}(t) = X_i(t) - \int_0^t \Delta_i(s) \exp\{\rho_i(s)\} d\Omega_\mu(s)$ are the zero-mean stochastic processes. The following estimating equations can thus be used

$$\mathcal{E}_2(\boldsymbol{\beta}, \boldsymbol{\theta}) = \sum_{i=1}^n \int_0^\tau \Delta_i(t) \Psi(t) \left\{\boldsymbol{\varphi}_i(t) - \bar{\boldsymbol{\varphi}}_\rho(t; \boldsymbol{\beta}, \boldsymbol{\theta})\right\} dX_i(t) = 0,$$

where $\bar{\boldsymbol{\varphi}}_\rho(t;\boldsymbol{\beta},\boldsymbol{\theta}) = \sum_{i=1}^n \Delta_i(t)\exp\{\rho_i(t;\boldsymbol{\beta},\boldsymbol{\theta})\}\boldsymbol{\varphi}_i(t)/\sum_{i=1}^n \Delta_i(t)\exp\{\rho_i(t;\boldsymbol{\beta},\boldsymbol{\theta})\}$. Again, denote $(\widehat{\boldsymbol{\beta}},\widehat{\boldsymbol{\theta}})$ the solutions to $\mathcal{E}_2(\boldsymbol{\beta},\boldsymbol{\theta}) = 0$. Then the similar techniques applied in the additive model lead to the large-sample properties of consistency as well as asymptotic normality,

$$n^{1/2}\begin{pmatrix}\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}_*\\ \widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}_*\end{pmatrix} \to N(0, D_\rho^{-1}\Sigma_\rho D_\rho^{-1}) \tag{11}$$

in distribution, where

$$\begin{aligned} D_\rho &= -E\left[\int_0^\tau \Delta_1(t)\psi(t)\exp\{\rho_1(t)\}\bar{\boldsymbol{\varphi}}'_\rho(t)d\Omega_\mu(t)\right]\\ \Sigma_\rho &= E\left[\int_0^\tau \Delta_1(t)\psi(t)\left\{\boldsymbol{\varphi}_i(t)-\bar{\boldsymbol{\varphi}}_\rho(t;\boldsymbol{\beta},\boldsymbol{\theta})\right\}dM_{\rho,1}(t)\right]^{\otimes 2}. \end{aligned}$$

Here, D_ρ and Σ_ρ can be estimated by their empirical counterparts respectively.

3.2 *Covariate-dependent observation times*

Usually in a well-designed randomised clinical trial, the repeated measurements are supposed to collect at a pre-determined or fixed set of time points to avoid potential bias or missing in the data set. In reality, however, they may be actually observed at varying sets of time points for different individuals, which may be further affected by the subjects' covariates (Sun and Wei, 2000; Lin and Ying, 2000). In the statistical literature, when the mean functions of the counting processes are different, the following model are usually used,

$$E\{N_i(t) \mid \boldsymbol{Z}_i(s); 0 \le s \le t\} = \eta(t)\exp\{\boldsymbol{\kappa}^T\boldsymbol{Z}_i(t)\}, \tag{12}$$

where $\boldsymbol{\kappa}$ is parameter and $\eta(\cdot)$ is unspecified baseline function, as in Pepe and Cai (1993) and Lawless and Nadeau (1995). Hence, the following estimating equations can be used to estimate $\boldsymbol{\kappa}$,

$$\mathcal{E}_N(\boldsymbol{\kappa}) = \sum_{i=1}^n \int_0^\tau \{\boldsymbol{Z}_i(t) - \bar{\boldsymbol{Z}}(t;\boldsymbol{\kappa})\}dN_i(t) = 0,$$

where $\boldsymbol{Z}_i(t) = \sum_{i=1}^n \Delta_i(t)\exp\{\boldsymbol{\kappa}^T\boldsymbol{Z}_i(t)\}\boldsymbol{Z}_i(t)/\sum_{i=1}^n \Delta_i(t)\exp\{\boldsymbol{\kappa}^T\boldsymbol{Z}_i(t)\}$, by differentiating the log of partial likelihood function of

$$\sum_{i=1}^n \int_0^\tau \left\{\boldsymbol{\kappa}^T\boldsymbol{Z}_i(t) - \log\left[\sum_{k=1}^n \Delta_k(t)\exp\{\boldsymbol{\kappa}^T\boldsymbol{Z}_i(t)\}\right]\right\}dN_i(t)$$

with respect to $\boldsymbol{\kappa}$. Moreover, since now

$$E\{dX_i(t) \mid \boldsymbol{Z}_i(s), C_i; 0 \le s \le t, \boldsymbol{\beta}_*, \boldsymbol{\theta}_*, \boldsymbol{\kappa}_*\} = \Delta_i(t)\exp\{\boldsymbol{\kappa}_*^T\boldsymbol{Z}_i(t)\}d\Omega_{\mu,\eta}(t),$$

where $d\Omega_{\mu,\eta}(t) = \eta(t)d\Omega_{\mu}(t)$, the following estimating equations can be similarly established for $(\boldsymbol{\beta}^T, \boldsymbol{\theta}^T, \boldsymbol{\kappa}^T)^T$ as in (7),

$$\boldsymbol{\mathcal{E}}_1(\boldsymbol{\beta}, \boldsymbol{\theta}, \boldsymbol{\kappa}) = \sum_{i=1}^{n} \int_0^{\tau} \Delta_i(t)\Psi(t) \left\{\boldsymbol{\varphi}_i(t) - \bar{\boldsymbol{\varphi}}_1(t)\right\} dX_i(t) = 0,$$

where $\bar{\boldsymbol{\varphi}}_1(t) = \sum_{i=1}^{n} \Delta_i(t) \exp\{\boldsymbol{\kappa}^T \boldsymbol{Z}_i(t)\}\boldsymbol{\varphi}_i(t) / \sum_{i=1}^{n} \Delta_i(t) \exp\{\boldsymbol{\kappa}^T \boldsymbol{Z}_i(t)\}$. Denote $(\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\theta}}, \widehat{\boldsymbol{\kappa}})$ the solutions such that $\mathcal{E}_N(\widehat{\boldsymbol{\kappa}}) = \mathcal{E}_1(\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\theta}}, \widehat{\boldsymbol{\kappa}}) = 0$. Then following the arguments in Sun and Wei (2000), it is true that they are consist and have the asymptotic normality as,

$$n^{1/2} \begin{pmatrix} \widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_* \\ \widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_* \end{pmatrix} \doteq n^{1/2} \left(D_1^{-1}(\boldsymbol{\beta}_*, \boldsymbol{\theta}_*, \boldsymbol{\kappa}_*), -D_N^{-1}(\boldsymbol{\kappa}_*)\right) \begin{pmatrix} \mathcal{E}_1(\boldsymbol{\beta}_*, \boldsymbol{\theta}_*, \boldsymbol{\kappa}_*) \\ \mathcal{E}_N(\boldsymbol{\kappa}_*) \end{pmatrix}, \tag{13}$$

where $D_1 = -\lim_{n\to\infty} n^{-1}\mathcal{E}_1'$ and $D_N = -\lim_{n\to\infty} n^{-1}\mathcal{E}_N'$. Hence, by the normal approximation of $(\mathcal{E}_1, \mathcal{E}_N)$ as shown the Appendix, the asymptotic variance of (13) can be estimated by $(\widehat{D}_1^{-1}, -\widehat{D}_0^{-1})\widehat{\Sigma}_1(\widehat{D}_1^{-1}, -\widehat{D}_N^{-1})^T$, where $\widehat{D}_N$, $\widehat{D}_1$ and $\widehat{\Sigma}_1$ are their respective empirical estimates.

3.3 *Isotonic regression of mean response models*

In either the additive model (1) or the multiplicative model (8), the mean of the baseline response curves are assumed to be arbitrarily unspecified. In the randomised trials, for instance, this means that the mean response curves of the subjects in the control group is completely unspecified. The overall trend on the lowess curve of the pain intensity score over time for the control group, however, may suggest that there is a pattern of the curve as a monotonically decreasing function of time. This is usually not surprising especially in the drug trials, when most of the pre-clinical stability studies showing the effectiveness of the drug compound with decreasing drug potency over time (Chen, et al, 2003). Therefore it is reasonable to extend the model (4) to the following one,

$$E\{Y_i(t) \mid Z_i(s); 0 \le s \le t\} = \mu(t) + \boldsymbol{\beta}_Q^T \boldsymbol{Q}_i(t) H(t; \boldsymbol{\theta}) + \boldsymbol{\beta}_R^T \boldsymbol{R}_i(t), \tag{14}$$

with $\mu(\cdot) \in \mathcal{M}$, where $\mathcal{M}$ is the set of all the monotonic functions. In the memantine trial, $\mathcal{M}$ should include all the monotonically non-decreasing functions. Denote the set by $\mathcal{M}^-$.

When there is no covariate information included in (14), the regression model reduces to a simple isotonic estimation problem. That is, we need to find $\mu(\cdot) \in \mathcal{M}^-$ such that

$$\mu^-(\cdot) = \arg \min_{\mu \in \mathcal{M}^-} \sum_{i=1}^{n} \|Y_i - \mu\|^2,$$

with the norm $\|\cdot\|$ defined as in Rice and Silverman (1991). Thus the computational algorithms, such as the most widely used Pooled Adjacent Violators Algorithm or the Minimum Lower Set Algorithm, can be used (Robertson, Wright and Dykstra, 1988). When the covariate information is included as proposed in the model, we can adapt the back-fitting algorithm as in Zeger and Diggle (1994) to obtain the final estimates of the baseline function $\mu\cdot$ and the parameters.

Algorithm.

1. Consider $(\widehat{\boldsymbol{\beta}}_{[k]}, \widehat{\boldsymbol{\theta}}_{[k]})$ are obtained in the kth iterative step, $k = 1, 2, \ldots$, where $\widehat{\boldsymbol{\beta}}_{[0]} = \widehat{\boldsymbol{\theta}}_{[0]} = 0$. Use one of the aforementioned algorithm to compute $\mu^{-}_{[k+1]}(\cdot) \in \mathcal{M}$ such that

$$\mu^{-}_{[k+1]}(\cdot) = \arg \min_{\mu \in \mathcal{M}^{-}} \sum_{i=1}^{n} \left\| Y_{[k],i} - \mu \right\|^2,$$

   where $Y_{[k],i}(t) = Y_i(t) - \{\boldsymbol{\beta}^T_{Q,[k]} \boldsymbol{Q}_i(t) H(t; \boldsymbol{\theta}_{[k]}) + \boldsymbol{\beta}^T_{R,[k]} \boldsymbol{R}_i(t)\}$;

2. Given $\mu^{-}_{[k+1]}(\cdot)$, obtain $(\widehat{\boldsymbol{\beta}}_{[k+1]}, \widehat{\boldsymbol{\theta}}_{[k+1]})$ by minimizing

$$\sum_{i=1}^{n} \int_0^{\tau} \Phi(t) [Y_i(t) - \{\mu^{-}_{[k+1]}(t) + \boldsymbol{\beta}^T_Q \boldsymbol{Q}_i(t) H(t; \boldsymbol{\theta}) + \boldsymbol{\beta}^T_R \boldsymbol{R}_i(t)\}]^2 dN_i(t).$$

In fact, the proposed isotonic regression model belongs to a more general additive isotonic model (Bacchetti, 1989),

$$E\{Y_i(t) \mid Z_i(s); 0 \le s \le t\} = \sum_{l=1}^{P} \mu_l(t) + \boldsymbol{\beta}^T_Q \boldsymbol{Q}_i(t) H(t; \boldsymbol{\theta}) + \boldsymbol{\beta}^T_R \boldsymbol{R}_i(t),$$

where $(\mu_1, \mu_2, \ldots, \mu_P)$ are the P-dimensional isotonic function vector. When there is no covariate information involved, the backfitting algorithm by Hastie and Tibshirani (1990) can be used with the Pooled Adjacent Violators Algorithm to individual μ_i iteratively. When the covariate information is included, it is straightforward to further extend the above algorithm for the estimation in this model. To avoid complicated variance calculation of the estimators, the computer-intensive methods such as bootstrapping (Efron and Tibshirani, 1994) can be used.

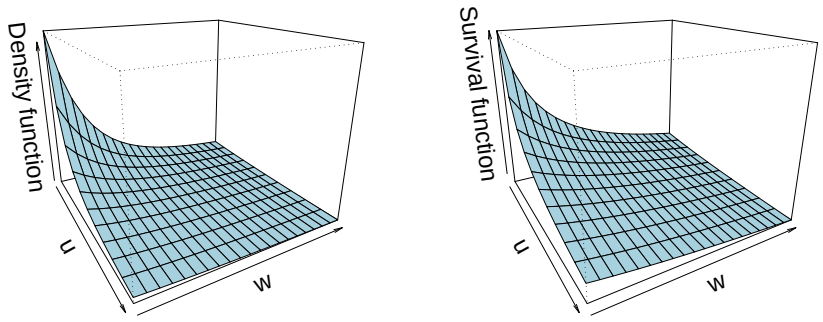

Fig. 2: Density function and survival function of the bivariate action onset

4 Examples

4.1 *Distributions of bivariate action onset times*

As proposed in §2.1, the bivariate action onset times can be modelled by the Gamma frailty model. To gain some concrete sense about this family of the distributions and their ultimate impact on the mean response curve, we choose some examples from this family of distributions. One special choice is to use the Weibull forms for $\lambda_U(\cdot)$ and $\lambda_W(\cdot)$,i.e., $\lambda_U\omega t^{\omega-1}$ and $\lambda_W\omega t^{\omega-1}$, respectively, where ω is parameter. Thus, the bivariate density function and survival function of (U, W) becomes

$$f_{U,W}(u, w; \theta) = \alpha(1+\alpha)\alpha^{-\alpha}\lambda_U\lambda_W u^{\omega} w^{\omega}(\alpha^{-1} + \lambda_U u^{\omega} + \lambda_W w^{\omega})^{-\alpha-2},$$

and

$$S_{U,W}(u, w; \theta) = \{1 + (\lambda_U u^{\omega} + \lambda_W w^{\omega})/\alpha\}^{-\alpha},$$

where $\boldsymbol{\theta} = (\alpha^T, \omega, \lambda_U, \lambda_W)^T$. This is the generalised Pareto power distribution, also called the bivariate Burr distribution. When $\omega = 1$, the marginal distributions of (U, W) become exponential. The bivariate density and survival functions are demonstrated in Figure 2 when $\lambda_U = \lambda_W = \omega = 1$ and $\alpha = 0.5$.

The impact of inclusion of varying action onset on the mean response model is in fact reflected by the shape of the function of $H(t)$, as demonstrated in model (4). The function of $H(t)$ under the mentioned distributions of (U, W) are plotted in Figure 3 under three situations of $\lambda_W = 0.5$, 1.0 and 1.5, represent relatively shorter/longer period of time of the action onset. It is not surprising to see that all the curves appear to be tied down toward 0 at both ends of 0 and ∞ with a peak in the middle. This means that the treatment may be observed to take effect gradually from the beginning, reach the peak efficacy and

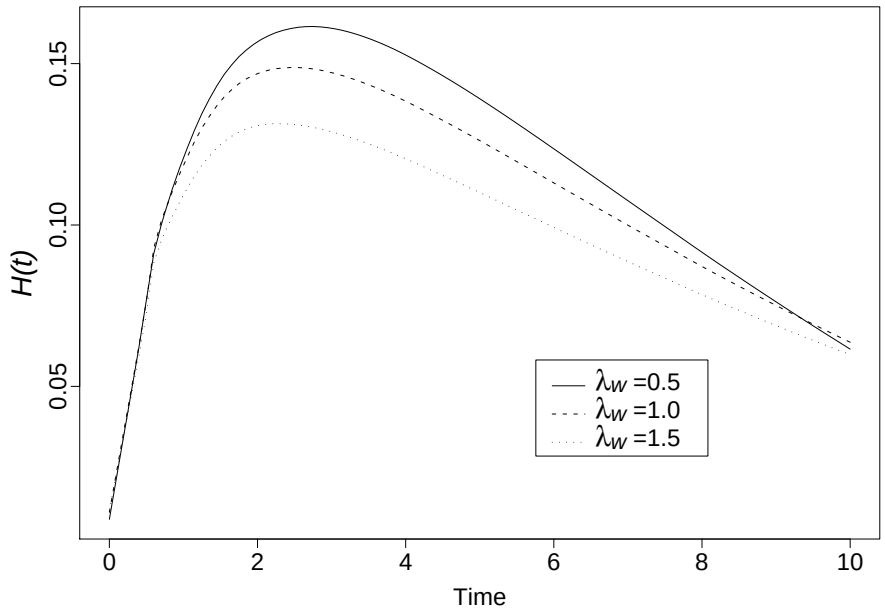

Fig. 3: The functions of $H(t)$ in model (4)

then dampen as time goes on. More interestingly, as λ_W increases, the time period of action onset becomes shorter, and the curves appears to have uniformly lower efficacy, i.e., $H(t;\lambda_W = 1.5) \leq H(t;\lambda_W = 1.0) \leq H(t;\lambda_W = 0.5)$.

4.2 *Simulations*

Moderate simulations are conducted mainly to demonstrate the validity of the estimation procedures. According to our models, there are three steps to simulate the data sets: (1) varying effectiveness times (u_i, w_i). These bivariate times are simulated following the Gamma frailty model. The ultimate density function used for the bivariate times is $0.75\sqrt{2}uw(u+w+2)^{-2.5}$; (2) observation times $(t_{i1}, t_{i2}, \ldots, t_{i,m_i})$. The observations time are simulated according to a random effect Poisson process with intensity rate following Gamma (1,0.5). The total time period of observation following uniform distribution with mean of 20, which yields about 11 observation times per subject; (3) repeated responses $(y(t_{i1}), y(t_{i2}), \ldots, y(t_{i,m_i}))$. The repeated responses are simulated according to the following model:

$$y_{ij}(t_{ij}) = \mu(t_{ij}) + \beta_Q QI(u_i < t_{ij} < u_i + w_i) + \beta_R R(t_{ij}) + \varepsilon(t_{ij}).$$

Here Q is the treatment indicator of Bernoulli random variable with the success probability of 50%, $R(t)$ are standard normal, $\varepsilon(t)$ is Gaussian process with $\text{cov}\{\varepsilon(s), \varepsilon(t)\} = \exp(-|s-t|)$

Table 1: Summary of simulation results. Each entry is the estimated bias with 95% empirical coverage probabilities in brackets

		$(\boldsymbol{\beta}_{Q^*}, \boldsymbol{\beta}_{R^*}) = (1,0)$		$(\boldsymbol{\beta}_{Q^*}, \boldsymbol{\beta}_{R^*}) = (0,0)$		$(\boldsymbol{\beta}_{Q^*}, \boldsymbol{\beta}_{R^*}) = (0,1)$	
n	$\mu(t)$	$\boldsymbol{\beta}_{Q}$	$\boldsymbol{\beta}_{R}$	$\boldsymbol{\beta}_{Q}$	$\boldsymbol{\beta}_{R}$	$\boldsymbol{\beta}_{Q}$	$\boldsymbol{\beta}_{R}$
50	$\sqrt{t}$	-0.027	-0.007	0.004	-0.007	0.005	-0.008
		(0.960)	(0.945)	(0.960)	(0.947)	(0.916)	(0.969)
50	$\sin(2\pi t)$	-0.004	-0.004	-0.002	0.002	-0.003	-0.015
		(0.940)	(0.944)	(0.952)	(0.926)	(0.943)	(0.972)
100	$\sqrt{t}$	0.002	-0.006	-0.014	0.014	-0.010	0.002
		(0.943)	(0.956)	(0.941)	(0.946)	(0.967)	(0.941)
100	$\sin(2\pi t)$	0.004	-0.014	0.008	0.013	-0.002	0.010
		(0.938)	(0.952)	(0.941)	(0.932)	(0.946)	(0.950)
200	$\sqrt{t}$	0.004	-0.004	-0.006	0.008	0.001	0.010
		(0.939)	(0.943)	(0.946)	(0.949)	(0.953)	(0.941)
200	$\sin(2\pi t)$	-0.014	0.011	-0.001	0.003	-0.003	0.011
		(0.940)	(0.958)	(0.964)	(0.971)	(0.927)	(0.943)

and $\mu(t) = t^{1/2}$ and $\sin(2\pi t)$, respectively. The true values of (β_R, β_Q) are (0,0), (0,1) and (1,0), respectively. The simulation results are summarised in Table 1. For each entry in the table, 1,000 replicates are simulated to estimate the bias and empirical coverage probability. The bias is defined as the difference between the sample mean of the estimates over the 1,000 replicated data sets and its true value. The empirical coverage probability is the percentage of Wald-type 95% confidence intervals that include the true parameters. It is evident that the estimators are virtually unbiased and the nominal confidence intervals carry reasonable coverages.

4.3 *A real randomised clinical trial*

The chronic pain due to damaged peripheral nerves is one of the leading complications of diabetic patients. Among the 10.3 million patients diagnosed with diabetes in the United States, more than 60% of them suffer some form of damaged nerves, which may lead to more than 1 million neuropathic pain cases. The basic function of Memantine is to restore of the function of damaged nerve cells and block the excitation of N-methyl-D-aspartate (NMDA) receptors. It was shown to be effective in reducing pain responses in rodent and

Table 2: Parameter estimates in model (4) with/without effectiveness onset: Cov., covariates; Est., parameter estimates; s.e., standard errors; CI, confidence interval. The reference groups for Gender, Analgesic usage and Treatment are male, no use and placebo, respectively.

	Without varying onset			With varying onset		
Cov.	Est.	s.e.	95% CI	Estimate	s.e.	95% CI
Gender	1.851	1.306	(-0.708,4.410)	-1.756	2.038	(-5.750,2.238)
Age	0.045	0.066	(-0.084,0.174)	0.116	0.103	(-0.086,0.318)
Analgesic usage	1.043	1.311	(-1.527,3.613)	-1.578	2.045	(-5.586,2.430)
Days	-0.070	0.009	(-0.088,-0.052)	-0.039	0.010	(-0.235,-0.019)
Treatment	-1.118	1.273	(-1.377,1.377)	-4.433	1.986	(-8.326,-0.540)

primate chronic pain model (Seltzer, et al., 1991). Clinical trials on human subjects have been conducted to evaluate the efficacy of memantine and its dose-response with relatively sample sizes, for example, in Sang, et al. (2002). A randomised clinical trial of larger scale was conducted to evaluate its efficacy in the treatment of diabetic patients with painful peripheral neuropathy. This is a 16-week, randomised, double-blinded placebo-controlled trial with a total of 420 dibetic patients. The primary efficacy outcomes are the repeated measurements of VAS nocturnal pain intensity measured weekly. When the primary endpoint is the change in the intensity scores from the baseline in 16 weeks, it is found that neither the usual approach of ANOVA nor the ANCOVA would yield significant reduction in the recorded pain intensities.

A closer examination of the graph of pain intensity scores shows that the memantine reduces the pain intensity gradually till around 20 days. Then the reduction stays stable through the most of the rest of trial period. The control group, however, seems to have a sudden declining pain intensity around the end of the trial, and thus the difference between the two groups diminishes. This might be the cause for the aforementioned ANOVA or ANCOVA approaches with less power to detect the overall differences. To actually implement our models (4) for the dependent variable of repeated measurements on the pain intensity scores, five covariates are selected: treatment indicator for $\boldsymbol{Q}(t)$, and gender, age, concomitant analgesic usage and days since randomization for $\boldsymbol{R}(t)$. The estimates are listed in Table 2 for the varying effectiveness onset being included and not.

As show in the table, the covariates of gender, age and analgesic usage are both not significant in the two models, while the time trend for the days since randomisation is significant in both models. However, the treatment appears not significant when the varying

effectiveness onset is not included, but significant otherwise. When the magnitude of estimates are examined, it is found that the treatment would have more impact in reducing the intensity scores with varying effectiveness onset. It is on average reduced about 4.5 considering the varying effectiveness onset in contrast to 1.1 not considering. But interestingly, the time effect appears less impact with varying effectiveness onset, it changes from 0.07 to 0.04 reduction per day. Another notable observation is that, the gender effect has different direction by comparing two models, although they are not significant. This lead to the conjecture that the varying effectiveness onset may be gender-specific, which still needs to be confirmed with larger sample size and further modelling of effectiveness onset on gender.

5 Discussion

The phenomenon of treatment efficacy gradually improving as time progresses has been studied in the statistical literature. For instance, Zucker and Lakatos (1990) coined the term of "treatment effectiveness lag" to characterise the slow onset of a treatment efficacy, and Chen, et al. (2002) developed regression tools to account for such treatment effectiveness lags in time-to-event data analysis. The phenomenon of saturation of treatment efficacy is, however, less explored in statistical literature, although it has been long recognised in physics. For example, in Beiser (1984), a decaying process was described by an exponential curve within a small time interval for the number of atoms emitted by Uranium, based on the quantum mechnical laws. But this process of decaying is eventually moderated by other factors to terminate the decaying trend but reach the stable saturation, which would otherwise violate the laws of energy or space constraints. These same laws apply to the saturation phenomenon in human metabolism and drug compound mechanism. The proposed model in this article to include varying action onset does not intend to specify any individual action onset time on the repeated measurements, but is able to describe the average effect of such action onset marginally.

Mathematically, the marginalised model (4) is in fact a time-varying coefficient model. Because it is based on the possible biological mechanisms, it has more direct interpretation on the parameters in the model, if compared with an arbitrary descriptive time-varying coefficient. The introduction of varying treatment effectiveness action onset is also equivalent to most of the smoothing techniques applied in the nonparametric estimation approaches: the distribution assumption on the unobserved varying action onset and its marginalisation essentially smoothes the differences in mean response between the action onset and otherwise.

The estimation approaches in this article following the counting processes formulation in analysis of repeated measurements by Cheng and Wei (2000), Sun and Wei (2000), Lin and Ying (2001) and others. This formulation is simple and does not require smoothing, with room for significant improvement in efficiency. In Lin and Ying (2001), an estimate of the baseline function was introduced to improve the efficiency by minimizing the variance of the proposed estimating equations. It is still ad hoc and unknown whether or not the efficiency reaches the semiparametric efficiency bound. Although there are other approaches that do not need smoothing yet may have better efficient estimation, for instance, the difference-based method by Yatchew (1997) for the partial linear models with less loss of efficiency, more future work in the semiparametric model efficiency framework of Bickel, et al. (1993) and van der Laan and Robins (2002) are needed. Along with the efficiency calculation, the technical development of asymptotic theory for the smoothing baseline estimators in §2 and the isotonic regression algorithms in §3 will be addressed in separate manuscripts, given the interest of these theory development beyond the scope of the current manuscript.

Appendix A: Asymptotics

A.1. Weak of Convergence of $n^{-1/2}\mathcal{E}(\cdot;\boldsymbol{\beta}_*,\boldsymbol{\theta}_*)$

Our proof follows an extension of the Appendix 2 in Cheng and Wei (2000). Denote $\mathcal{B}(t) = \sum_{i=1}^n \int_0^t \Delta_i(s)\Phi(s)dM_i(s)$ and $\mathcal{B}_\varphi(t) = \sum_{i=1}^n \int_0^t \Delta_i(s)\Phi(s)\varphi(s)dM_i(s)$. Then $\mathcal{E}(\boldsymbol{\beta}_*,\boldsymbol{\theta}_*) = \mathcal{B}_\varphi(\tau) - \int_0^\tau \bar{\varphi}(t)d\mathcal{B}(t)$. For any $t > 0$, $\mathcal{B}(t)$ and $\mathcal{B}_\varphi(t)$ are the sums of independently and identically distributed zero-mean terms. By the Central Limit Theorem, $n^{-1/2}(\mathcal{B}(t),\mathcal{B}_\varphi(t))$ converges in distribution to a zero-mean Gaussian process, $(\mathcal{W}(t),\mathcal{W}_\varphi(t))$, say.

Assume that $\varphi_i(\cdot)$, $i = 1, 2, \ldots, n$, are of bounded variation. Moreover, without loss of generality, $\varphi_i(\cdot)$ are assumed to be non-negative. Then the individual terms of $\mathcal{B}(\cdot)$ and $\mathcal{B}_\varphi(\cdot)$ can be written as sums of monotone functions in t and hence "manageable." Thus $n^{-1/2}(\mathcal{B}(t),\mathcal{B}_\varphi(t))$ converges weakly to $(\mathcal{W},\mathcal{W}_\varphi)$, as $n \to \infty$ (Pollard, 1990, p. 38 and p.53). By the strong embedding theorem in Shorack and Wellner (1986, p. 47), there exists an induced probability space such that $(n^{-1/2}\mathcal{B}(t), n^{-1/2}\mathcal{B}_\varphi(t), n^{-1}\sum_{i=1}^n \Delta_i(t), n^{-1}\sum_{i=1}^n \Delta_i(t)\varphi_i(t))$ converges almost surely. By the Lemma 8.2.3 in Chow and Teicher (1988, p.265) coupled with the Helly's theorem in Serfling (1980, p.352), it is true that

$$n^{-1/2}\int_0^t \frac{n}{\sum_{i=1}^n \Delta_i(s)}d\mathcal{B}(s) \to \int_0^t \frac{1}{E\Delta_1(s)}d\mathcal{W}(s) \text{ and } n^{-1/2}\int_0^t \bar{\varphi}(s)d\mathcal{B}(s) \to \int_0^t \bar{\varphi}_*(s)d\mathcal{W}(s)$$

almost surely and uniformly in t. The weak convergence of $n^{-1/2}\mathcal{E}(\boldsymbol{\beta}_*, \boldsymbol{\theta}_*)$ thus follows in the original probability space, due to their convergence almost surely to $\mathcal{W}_\varphi(\tau) - \int_0^\tau \bar{\varphi}_*(s)d\mathcal{W}(s)$ in the induced probability. The calculation of the variance-covariance matrix of Σ is straightforward.

A.2. Asymptotic variance of $n^{-1/2}(\mathcal{E}_1(\boldsymbol{\beta}_*, \boldsymbol{\theta}_*, \boldsymbol{\kappa}_*)^T, \mathcal{E}_N(\boldsymbol{\kappa}_*)^T)^T$

The asymptotic normality of the joint distribution of $n^{-1/2}(\mathcal{E}_1(\boldsymbol{\beta}_*, \boldsymbol{\theta}_*, \boldsymbol{\kappa}_*)^T, \mathcal{E}_N(\boldsymbol{\kappa}_*)^T)^T$ can be similarly established following the arguments in Lin and Wei (1989) and Sun and Wei (2000). To calculate its associated asymptotic variance, it is noted that

$$n^{-1/2}\mathcal{E}_N(\boldsymbol{\kappa}_*) = n^{-1/2}\sum_{i=1}^n \int_0^\tau \left\{\boldsymbol{Z}_i(t) - \bar{\boldsymbol{Z}}(t;\boldsymbol{\kappa})\right\} dM_{N,i}(t),$$

where $M_{N,i}(t) = N_i(t) - \int_0^t \Delta_i(s)\exp\{\boldsymbol{\kappa}^T\boldsymbol{Z}_i(s)\}d\Omega(s)$. Let $\boldsymbol{e}_i = E\int_0^\tau \left\{\boldsymbol{Z}_i(t) - \bar{\boldsymbol{Z}}(t;\boldsymbol{\kappa})\right\} dM_{N,i}(t)$ and its empirical estimates as $\widehat{\boldsymbol{e}}_i$, respectively. Thus the variance-covariance matrix of $n^{-1/2}(\mathcal{E}_1(\boldsymbol{\beta}_*, \boldsymbol{\theta}_*, \boldsymbol{\kappa}_*)^T, \mathcal{E}_N(\boldsymbol{\kappa}_*)^T)^T$ can be approximated by

$$\widehat{\Sigma}_1 = \begin{pmatrix} n^{-1}\sum_{i=1}^n \widehat{\boldsymbol{\epsilon}}_i\widehat{\boldsymbol{\epsilon}}_i^T & n^{-1}\sum_{i=1}^n \widehat{\boldsymbol{\epsilon}}_i\widehat{\boldsymbol{e}}_i^T \\ n^{-1}\sum_{i=1}^n \widehat{\boldsymbol{e}}_i\widehat{\boldsymbol{\epsilon}}_i^T & n^{-1}\sum_{i=1}^n \widehat{\boldsymbol{e}}_i\widehat{\boldsymbol{e}}_i^T \end{pmatrix},$$

where $\widehat{\boldsymbol{\epsilon}}_i = \int_0^\tau \Delta_i(t)\Phi(t)\{\boldsymbol{\varphi}_i(t) - \bar{\boldsymbol{\varphi}}_1(t)\}\exp\{\widehat{\boldsymbol{\kappa}}^T\boldsymbol{Z}_i(t)\}d\widehat{\Omega}_{\mu,\eta}(t)$ and

$$\widehat{\Omega}_{\mu,\eta}(t) = \int_0^t \frac{\sum_{i=1}^n dX_i(s)}{\sum_{i=1}^n \Delta_i(t)\exp\{\widehat{\boldsymbol{\kappa}}^T\boldsymbol{Z}_i(s)\}}.$$

References

Albert, P. S. (1999). Tutorial in Biostatistics: longitudinal data analysis (repeated measures) in clinical trials. *Statistics in Medicine* **18**, 1707-1732.

Bacchetti, P. (1989). Additive isotonic models. *Journal of American Statistical Association* **84**, 289-294.

Beiser, A. (1984). *Concepts of Modern Physics*. New York: McGraw-Hill.

Bickel, P. J., Klasssen, C. A. J., Ritov, Y. and Wellner (1993). *Efficiency and Adaptive Estimation for Semiparametric Models*. Baltimore: Johns Hopkins University Press.

Chen, Y. Q., Pong, A. P. and Xing, B. (2003). Rank regression in stability analysis. *Journal of Biopharmaceutical Statistics* **13**, 463-479.

Chen, Y. Q., Rohde, C. A. and Wang, M.-C. (2002). Additive hazards models with latent treatment effectiveness lag time. *Biometrika* **89**, 917-931.

Cheng, S. C. and L. J. Wei (2000). Inferences for a semiparametric model with panel data. *Biometrika* **87**, 89-97.

Chiang, C. T., Rice, J. A. and Wu, C. O. (2001). Smoothing spline estimation for varying coefficients models with repeatedly measured dependent variables. *Journal of the American Statistical Association* **96**, 605-619.

Chow, Y. S. and Teicher, H. (1988). *Probability Theory: Independence, Interchangeability, Martingales, 2nd Ed.* New York: Springer.

Clayton, D. G. (1978). A model for association in bivariate life tables and its application in epidemiologic studies of familial tendency in chronic disease incidence. *Biometrika* **65**, 141-151.

Diggle, P. J., Heagerty, P., Liang, K.-Y. and Zeger, S. L. (2001). *Analysis of Longitudinal Data*, 2nd Ed., New York: Oxford University Press.

Efron, B. and Tibshirani, R. J. (1994). *An Introduction to Bootstrapping*. New York: Chapman & Hall.

Hastie, T. and Tibshirani, R. (1990). *Generalized Additive Models*. London: Chapman & Hall.

Hoover, D. R., Rice, J. A., Wu, C. O. and Yang, L. P. (1998). Nonparametric smoothing estimates of time-varying coefficient models with longitudinal data. *Biometrika* **85**, 809-822.

Hougaard, P. (2000). *Analysis of Multivariate Survival Data*. New York: Springer.

Huang, J. H., Wu, C. O. and Zhou, L. (2002). Varying-coefficient models and basis function approximations for the analysis of repeated measurements. *Biometrika* **89**, 111-128.

Johnson, R. E., Jaffe, J. H. and Fudala, P. J. (1992). A controlled trial of buprenorphine treatment for opiate dependence. *Journal of the American Medical Association* **267**, 2750-2755.

Lawless, J. F. and Nadeau, C. (1995). Some simple robust methods for the analysis of recurrent event data. *Technometrics* **37**, 158-168.

Lin, D. Y. and Ying, Z. (2001). Semiparametric and nonparametric regression analysis of longitudinal data (with discussion). *Journal of American Statistical Association* **96**, 103-126.

Lin, D. Y. and Wei, L. J. (1989). The robust inference for the Cox proportional hazards model. *Journal of American Statistical Association* **84**, 1074-1078.

Oakes, D. R. (1989). Bivariate survival models induced by frailties. *Journal of American Statistical Association* **84**, 487-493.

Pollard, D. (1990). *Empirical Processes: Theory and Applications.* Hayward, California:Institute of Mathematical Statistics.

Pepe, M. S. and Cai, J. (1993). Some graphical displays and marginal regression analyses for recurrent failure times and time dependent covariates. *Journal of American Statistical Association* **88**, 811-820.

Rice, J. A. (2003). Functional and longitudinal data analysis: perspectives on smoothing. *UC Berkeley Department of Statistics Technical Report.*

Rice, J. A. and Silverman, B. W. (1991). Estimating the mean and covariance structure nonparametrically when the data are curves. *Journal of Royal Statistical Society, Series B.* **53**, 233-243.

Robertson, T., Wright, F. T. and Dykstra, R. L. (1988). *Order Restricted Statistical Inference.* New York: Wiley.

Sang, C. N., Booher, S., Gilron, I., Parada, S. and Max, M. B. (2002). Dextromethorphan and memantine in painful dibetic neuropathy and postherpetic neuralgia. *Anesthesiology* **96**, 1053-1061.

Seltzer, Z., Cohn, S., Ginzburg, R., Berilin, B. (1991). Modulation of neuropathic pain behavior in rats by spinal disinhibition and NMDA receptor blockade of injury discharge. *Pain* **49**, 121-128.

Serfling, R. J. (1980). *Approximation Theorems of Mathematical Statistics.* New York: Wiley.

Shorack, G. R. and Wellner, J. A. (1986). *Empirical Process with Applications to Statistics*. New York: Wiley.

Sun, J. and Wei, L. J. (2000). Regression analysis of panel count data with covariate-dependent observation and censoring times. *Journal of Royal Statistical Society, Series B* **62**, 293-302.

van der Laan, M. J. and Robins, J. M. (2003). *Unified Methods for Censored Longitudinal Data and Causality*. New York: Springer.

Wang, J.-L. and Wang, W. (2001). Comment on "Semiparametric and nonparametric regression analysis of longitudinal data" by Lin and Ying. *Journal of American Statistical Association* **96**, 119-123.

Wu, W. B., Woodroofe, M. and Mentz, G. (2001). Isotonic regression: another look at the changepoint problem. *Biometrika* **88**, 793-804.

Yatchew, A. (1997). An elementary estimator of the partial linear model. *Economics Letters* **57**, 135-143.

Zeger, S. L. and Diggle, P. J. (1994). Semiparametric models for longitudinal data with application to CD4 cell numbers in HIV seroconverters. *Biometrics* **50**, 689-699.

Zucker, D. M. and Lakatos, E. (1990). Weighted log rank type statistics for comparing survival curves when there is a time lag in effectiveness of treatment. *Biometrika* **77**, 853-864.