Efficient Panel Designs for Longitudinal Recurrent Event Studies Recording Panel Counts: Supplementary Material

ELIZABETH JUAREZ-COLUNGA

Department of Biostatistics and Informatics, University of Colorado Denver, Aurora, CO, USA, 80045.

C. B. DEAN*

Department of Statistical and Actuarial Sciences, Western University, London, ON, Canada, N6A 5B7.

dean@stats.uwo.ca

ROBERT BALSHAW

BC Centre for Disease Control, 655 West 12th Ave, Vancouver, BC, Canada, V5Z 4R4.

APPENDIX

*To whom correspondence should be addressed.
A. Notation

$i$: indexes individuals

t$: indexes time

$k$: number of treatments under test

$s$: indexes segments; $s = 1, \ldots, S$

$m_j$: number of individuals on treatment $j$, $j = 1, \ldots, k$

$M$: sample size of the study; $M = \sum_{j=1}^k m_j$

$Y_i(t)$: observation process for individual $i$

$N_i(t)$: counting process for individual $i$

$\bar{N}_i(t)$: observed counting process for individual $i$; $\bar{N}_i(t) = \int_0^t Y_i(u)dN_i(u)$

$x_i$: vector of treatment indicators for individual $i$

$z_i$: covariates for individual $i$

$\nu_i$: individual-specific random effect $i = 1, \ldots, M$

$\alpha$: parameters determining the baseline intensity function

$\beta$: treatment effects

$\gamma$: covariate effects

$\eta_1$: regression parameters; $\eta_1 = (\beta', \gamma')'$

$\eta$: regression parameters including parameters in the baseline intensity; $\eta = (\beta', \gamma', \alpha')'$

$\tau$: overdispersion parameter
\( \theta \): full set of parameters; \( \theta = (\beta', \gamma', \alpha', \tau)' \)

\((T_{s-1}, T_s]\): time period of segment \( s \); \( T^0 = 0 \)

\( I_{(T_{s-1}, T_s]}(t) \): indicator of whether individual \( i \) is observed at time \( t \)

\( \lambda_i(t) \): counting process intensity function

- 1-segment: \( \lambda_i(t) = \nu_i \rho(t; \alpha) e^{x_i' \beta + z_i' \gamma} \)
- \( S \) segments: \( \lambda_i(t) = \nu_i \rho(t; \alpha) \exp \left\{ x_i' \left[ \sum_{s=1}^{S} \beta_s I_{(T_{s-1}, T_s]}(t) \right] + z_i' \gamma \right\} \)

\( \rho(t; \alpha) \): baseline intensity function

\( d_\alpha \): dimension of \( \alpha \)

\( d_z \): dimension of \( \gamma \)

\( a \): indexes parameters in the baseline intensity; \( a = 1, \ldots, d_\alpha \)

\( n_{ip} \): number of events observed in panel \( p \) for individual \( i \), \( p = 1, \ldots, e_i \)

\( n_{i+} \): total number of events observed for individual \( i \); \( n_{i+} = \sum_{p=1}^{e_i} n_{ip} \)

\((T_{i,p-1}, T_{i,p}] \): \( p \)th panel period observed for individual \( i \) (for the 1-segment study)

\( T_{ei} \): termination for individual \( i \) (for the 1-segment study)

\( R_{ip} \): cumulative baseline intensity function over panel period \( p \) for individual \( i \) (for the 1-segment study);

\[ R_i = \int_{T_{i,p-1}}^{T_{i,p}} Y_i(t) \rho(t; \alpha) dt \]

\( R_i \): cumulative baseline intensity function over the entire observation period for individual \( i \); \( R_i = \int_0^\infty Y_i(t) \rho(t; \alpha) dt \)

\( \mu_{ip} \): expected mean number of events for individual \( i \) in panel \( p \) (for the 1-segment study); \( \mu_{ip} = R_{ip} \exp \{ x_i' \beta + z_i' \gamma \} \)

\( \mu_{i+} \): total expected mean number of events for individual \( i \); \( \mu_{i+} = R_i \exp \{ x_i' \beta + z_i' \gamma \} \)
$L_p$: likelihood based on panel data

$L_f$: likelihood based on full data, i.e. continuous followup; $L_f = L_{\alpha,f}$

$w_{ia} = \partial \log L_{\alpha,f}(\alpha)/\partial \alpha_a$: a function of the termination time for individual $i$

$\omega_{ipl}$: time of the $l$th event from start time ($t = 0$) for individual $i$ in panel period $p$

$g_p$: estimating equations for panel data analysis

$g_f$: estimating equations for full data analysis

$g_{\boldsymbol{\eta}_i}$: estimating equations for $(\boldsymbol{\beta}', \gamma')'$

$g_{\alpha,d}$: estimating equations for $\alpha$

$g_{\tau}$: estimating equation for $\tau$

$h_i$: correction term to reduce small sample bias;

$h_i = \text{diag}(U_1^{1/2}V_1'(V_1'U_1V_1)^{-1}V_1'U_1^{1/2})$

$V$: treatment and other covariates $V = (X \ Z)$

$W$: matrix with entries $w_{ia}$

$V_1$: covariates including $W$: $V_1 = (X \ Z \ W)$

$u_i = \mu_i + (1 + \tau\mu_i)$: function of the expected number of events and the overdispersion parameter for individual $i$

$T_{i,p}$: $p$th panel followup time for individual $i$, $i = 1, \ldots, M$, $p = 1, \ldots, e_i$

$T_{e_i}$: termination time for individual $i$; also denoted as $T_{i,e_i}$

$I_p, I_f$: information matrix based on likelihoods for the panel and full data, respectively

$\hat{\alpha}, \hat{\beta}, \hat{\gamma}, \hat{\tau}$: estimators obtained from an analysis of panel data
\( \hat{\alpha}, \hat{\beta}, \hat{\gamma}, \hat{\tau} \): estimators obtained from an analysis of the full data

\( G_j \): set of individuals who receive treatment \( j \)

\([n]_{j+} \): total number of events observed for all individuals receiving treatment \( j \), \([n]_{j+} = \sum_{i \in G_j} n_i^{+} \)

\([u]_{j+} \): function of the expected number of events for all individuals receiving treatment \( j \), \([u]_{j+} = \sum_{i \in G_j} u_i \)

\([u z]_{j+} \): function of the expected number of events and the covariates for all individuals receiving treatment \( j \), \([u z]_{j+} = \sum_{i \in G_j} u_i z_i \)

\([u w]_{j+} \): function of the expected number of events and the termination times for all individuals receiving treatment \( j \), \([u w]_{j+} = \sum_{i \in G_j} u_i w_i \)

\( \hat{\delta}_j \): estimator of time-weighted mean effect of treatment, specifically \( \hat{\delta}_j = \sum_{S=1}^{S} \Delta_s^{+} \hat{\beta}_s \), where \( \Delta_s^{+} = (T_s^{+} - T^s)/T^S \)

\( Q \): corrected variation of the \( w_i \)'s, \( Q = \phi_{w,w}(1 - \phi^2_{z,w}/\phi_{z,z} \phi_{w,w}) \)

\( \phi_{z,w} \): weighted covariation of \( z_i \)'s and \( w_i \)'s,
\[ \phi_{z,w} = \sum_{j=1}^{k} \sum_{i \in G_j} u_i (z_i - [u z]_{j+}/[u]_{j+}) (w_i - [u w]_{j+}/[u]_{j+}) \]

\( \phi_{z,z} \): weighted variation of \( z_i \)'s, \( \phi_{z,z} = \sum_{j=1}^{k} \sum_{i \in G_j} u_i (z_i - [u z]_{j+}/[u]_{j+})^2 \)

\( \phi_{w,w} \): weighted variation of \( w_i \)'s, \( \phi_{w,w} = \sum_{j=1}^{k} \sum_{i \in G_j} u_i (w_i - [u w]_{j+}/[u]_{j+})^2 \)

\( r_{z,w} = \phi_{z,w}/(\phi_{z,z} \phi_{w,w}) \): ratio of weighted covariation between \( z_i \) and \( w_i \); weights are \( u_i \).

\( l_1 \): measure of balance relating to \( \beta_1 \), \( l_1 = [u w]_{1+}/[u]_{1+} - [u z]_{1+}/[u]_{1+} \times \phi_{z,w}/\phi_{z,z} \),

\( l_j \): measure of balance relating to \( \beta_j \), \( j \geq 2 \);
\[ l_j = \sqrt{\phi_{w,w} \{l_{w,j}/\sqrt{\phi_{w,w}} - l_{z,j}/\sqrt{\phi_{z,z} r_{z,w}} \}} \]

\( l_{z,j} \): measure of balance relating to \( \beta_j \), and with respect to covariates; \( l_{z,j} = [u z]_{j+}/[u]_{j+} - [u z]_{1+}/[u]_{1+} \)
$l_{w,j}$: measure of balance relating to $\beta_j$, and with respect to termination times; $l_{w,j} = [uw]_{j+}^s/[u]_{j+}^s - [uw]_{1+}^s/[u]_{1+}^s$

$n_{ip}^s$: number of observed events for individual $i$ in panel $p$ in segment $s$, $n_{ip}^s = \bar{N}_i(T_{i,p}^s) - \bar{N}_i(T_{i,p-1}^s)$

$\mu_{ip}^s$: expected number of events for individual $i$ in panel $p$ in segment $s$, $\mu_{ip}^s = E(n_{ip}^s) = R_{ip}^s e^{x_i^s \beta^s + z_i^s \gamma}$

$R_{ip}^s$: cumulative baseline function for individual $i$ in panel $p$ in segment $s$, $R_{ip}^s = \int_{T_{i,p-1}^s}^{T_{i,p}^s} \rho(u; \alpha)Y_i(u)du$

$\mu_{i+}^s$: expected number of events for individual $i$ in segment $s$, $\mu_{i+}^s = \sum_{p=1}^{c_i^s} \mu_{ip}^s = \sum_{p=1}^{c_i^s} R_{ip}^s e^{x_i^s \beta^s + z_i^s \gamma}$

$R_i^s$: cumulative baseline function for individual $i$ in segment $s$, $R_i^s = \sum_{p=1}^{c_i^s} R_{ip}^s$

$u_i^s$: $u_i^s = \mu_{i+}^s/(1 + \tau \mu_{i+}^s)$

$w_i^s$: function of the $T_e$’s for individual $i$ in segment $s$, $w_i^s = \partial \log R_i^s / \partial \alpha$

$\beta^s$: treatment effects in segment $s$; $\beta_j^s$: effect of the $j$th treatment in segment $s$

$l_{w,j}^s$, $l_{u,j}^s$, $l_{j}^s$: segment-specific measures of balance with respect to covariates, termination times, and overall, respectively.

$[u]_{j+}^s$: function of the expected number of events for all individuals receiving treatment $j$ in segment $s$, $[u]_{j+}^s = \sum_{i \in G_j} u_i^s$

$[uz]_{j+}^s$: function of the expected number of events and the covariates for all individuals receiving treatment $j$ in segment $s$, $[uz]_{j+}^s = \sum_{i \in G_j} u_i^s z_i$

$[uw]_{j+}^s$: function of the expected number of events and the termination times for all individuals receiving treatment $j$ in segment $s$, $[uw]_{j+}^s = \sum_{i \in G_j} u_i^s w_i^s$
B. ASYMPTOTIC RELATIVE EFFICIENCIES OF THE ESTIMATORS OBTAINED FROM AN ANALYSIS OF PANEL DATA

Under standard conditions for the application of asymptotic results to estimating equations, \( \sqrt{M}(\hat{\theta} - \theta) \) is asymptotically normal with asymptotic covariance

\[
E \left( \left. - \lim_{M \to \infty} \frac{\partial g_f}{\partial \theta} \; \right| \frac{\partial g_f}{\partial \theta} \right)^{-1} E \left\{ \left. \frac{\partial g_f}{\partial \theta} \right| \frac{\partial g_f}{\partial \theta} \right\} \left\{ \left. \frac{\partial g_f}{\partial \theta} \right| \frac{\partial g_f}{\partial \theta} \right\}^{-1}.
\]

(B.1)

The asymptotic variance of \( \sqrt{M}(\eta - \eta) \), \( \eta = (\beta', \gamma', \alpha')' \), from the full data analysis is \( (\lim_{M \to \infty} \frac{1}{M} I_f)^{-1} \), where \( I_f \) has the form (2.9), because of three sets of identities: (i) \( E(-\partial g_{\eta_1}/\partial \tau) = 0 \), (ii) \( E(-\partial g_{\alpha_f}/\partial \tau) = 0 \), and (iii) the \( \{k + d_z + d_\alpha\} \times \{k + d_z + d_\alpha\} \) upper left submatrix of \( E(g_f g_f') \) is the same as the corresponding submatrix of \( E(-\partial g_f/\partial \theta) \).

Finite sample variance estimates are obtained by substituting \( \hat{\theta} \) for \( \theta \) and omitting the expressions \( \lim_{M \to \infty} \). In this case there are two usual options for approximating the expectation of the terms in (B.1). The first is a model-based approach, which requires specification of 3rd and 4th moments and is used here to derive the results of Theorem 1. The second option, the one we have employed in the illustration, is an empirical approach, which substitutes \( E\{\sum_{i=1}^{M} g_{if} g_{if}'\} \) by \( \{\sum_{i=1}^{M} g_{if} g_{if}'\} \); where \( g_{if} \) denotes the contribution to the score equation from individual \( i \). Note

\[
E\left\{ \left. - \frac{\partial g_f}{\partial \theta} \right| \frac{\partial g_f}{\partial \theta} \right\} = \left( \begin{array}{c} I_f \\ b' \\ b_0 \end{array} \right),
\]

and

\[
E\{g_{if} g_{if}'\} = \left( \begin{array}{c} I_f \\ c' \\ c_0 \end{array} \right).
\]

Here \( I_f \) is given in (2.9), \( 0, b, \) and \( c \) are of dimension \((k + d_z + d_\alpha) \times 1\) vectors, \( 0 \) is a vector of zeros, \( b \) and \( c \) have elements

\[
b_r = \sum_{i=1}^{M} \frac{\mu_i (1 + 2 \tau \mu_i) y_{ir}}{(1 + \tau \mu_i)^2}, \quad r = 1, \ldots, k + d_z + d_\alpha,
\]

\[
c_r = \sum_{i=1}^{M} \frac{\gamma_{3i} y_{ir}}{(1 + \tau \mu_i)^3}, \quad \gamma_{3i} = E(Y_i - \mu_i)^3, \quad r = 1, \ldots, k + d_z + d_\alpha,
\]
The group-specific weighted covariation \( \phi \) are given by (B.2) and (B.3) below:

\[
b_0 = \sum_{i=1}^{M} \frac{\mu_i^2}{(1 + \tau \mu_i)^2}, \\
c_0 = \sum_{i=1}^{M} \frac{\gamma_{4i} - \mu_i^2(1 + \tau \mu_i)^2}{(1 + \tau \mu_i)^4}, \quad \gamma_{4i} = E(Y_i - \mu_i)^4.
\]

The asymptotic variance of \( \hat{\theta} \) is then estimated as

\[
\left( \frac{I_f^{-1}}{b_0} \mathbf{c} - \mathbf{b}'I_f^{-1} \right) \left( \frac{1}{b_0} (c_0 - 2\mathbf{c}'I_f^{-1} \mathbf{b} + \mathbf{b}'I_f^{-1} \mathbf{b}) \right),
\]

replacing \( \theta \) by \( \hat{\theta} \).

The asymptotic model-based variance of \( \tilde{\eta} \) is estimated as above, replacing \( I_f \) with \( I_p \) and \( \theta \) with \( \tilde{\theta} \).

If we assume 3rd and 4th moments as for the negative binomial distribution, \( \gamma_{3i} = \mu_i(1 + \tau \mu_i)(1 + 2\tau \mu_i) \), \( \gamma_{4i} = 6\mu_i^2(1 + \tau \mu_i)^2 + \mu_i(1 + \tau \mu_i)(1 + 3\mu_i + 3\tau \mu_i^2) \), and \( \mathbf{c} = \mathbf{b} \). In this case, the estimators of \( \eta = (\beta', \gamma', \alpha')' \) and \( \tau \) from either the full or panel data analysis are asymptotically independent. Note that only mean and variance assumptions are required for consistency of the asymptotic variance of \( \tilde{\eta} \) or \( \tilde{\eta} \).

To obtain the asymptotic variances of \( \tilde{\beta} \) and \( \tilde{\beta} \), \( \text{Asvar}(\tilde{\beta}) \) and \( \text{Asvar}(\tilde{\beta}) \), respectively, we consider the partition of the information matrix (2.9) into blocks related to \( X, Z, \) and \( W \), and obtain the inverse; these expressions are given by (B.2) and (B.3) below:

\[
\text{Asvar}(\tilde{\beta}) = (X'UX)^{-1} + (X'UX)^{-1}X'UZ\phi_{z, z}^{-1}Z'UX(X'UX)^{-1} + L_J(Q + H_p)^{-1}L_J; \quad (B.2)
\]

\[
\text{Asvar}(\tilde{\beta}) = (X'UX)^{-1} + (X'UX)^{-1}X'UZ\phi_{z, z}^{-1}Z'UX(X'UX)^{-1} + L_J(Q + H_f)^{-1}L_J; \quad (B.3)
\]

where \( \phi_{z, z} = Z'UZ - Z'UX(X'UX)^{-1}X'UZ, \phi_{w, w} = W'UW - W'UX(X'UX)^{-1}X'UW \), \( \phi_{z, w} = Z'UW - Z'UX(X'UX)^{-1}X'UW, Q = \phi_{w, w} - \phi_{z, w}^{-1}\phi_{z, z} \), and

\[
L_J = (X'UX)^{-1}X'UW - (X'UX)^{-1}X'UZ\phi_{z, z}^{-1}\phi_{z, w} \cdot (B.4)
\]

The group-specific weighted covariation \( \phi_{z, w} \) is given explicitly in (B.5):

\[
\phi_{z, w(p, s)} = \sum_{j=1}^{k} \sum_{t \in G_j} u_i \left( z_{ip} - \left[ u_{j+} \right]_{j+} \right) \left( w_{is} - \left[ w_{j+} \right]_{j+} \right), \quad (B.5)
\]
where \( p = 1, 2, \ldots, d_z \) and \( s = 1, 2, \ldots, d_\alpha. \)

The elements of \((X'UX)^{-1}X'UZ\) are:

\[
\begin{pmatrix}
\left[u \times_{1+}\right]_1 / \left[u \times_{1+}\right]_1 \\
\left[u \times_{2+}\right]_2 / \left[u \times_{1+}\right]_1 - \left[u \times_{2+}\right]_1 / \left[u \times_{1+}\right]_1 \\
\vdots \\
\left[u \times_{k+}\right]_k / \left[u \times_{1+}\right]_1 - \left[u \times_{1+}\right]_1 / \left[u \times_{1+}\right]_1
\end{pmatrix},
\]

and do not depend on the \( \beta_j \)'s; they are function of the \( T_z \)'s and \( \alpha \) and \( \gamma. \) Similar expressions may be obtained for \((X'UX)^{-1}X'UW.\)

When \( \alpha \) is a scalar, \( L_J \) simplifies to a \( k \times 1 \) vector (B.7):

\[
L_J = (l_1, l_2, \ldots, l_k)' = (l_{w,1}, l_{w,2}, \ldots, l_{w,k})' - 
\begin{pmatrix}
l_{z,1} \\
l_{z,2} \\
\vdots \\
l_{z,k}
\end{pmatrix}
\frac{\phi_{z,w}}{\phi_{z,z}}
\]

where \( (l_{w,1}, l_{w,2}, \ldots, l_{w,k})' = (X'UX)^{-1}X'UW, \) and \( (l_{z,1}, l_{z,2}, \ldots, l_{z,k})' = (X'UX)^{-1}X'UZ.\)

Hence the asymptotic variances of the estimators \( \tilde{\beta} \) and \( \tilde{\alpha} \) are

\[
\text{Asvar}(\tilde{\beta}_1) = \left[u_{1+}\right]_1^{-1} + l_{z,1}^2 \phi_{z,z}^{-1} + l_1^2 (H_p + Q)^{-1},
\]

\[
\text{Asvar}(\tilde{\beta}_j) = \left[u_{j+}\right]_j^{-1} + \left[u_{j+}\right]_1^{-1} + l_{z,j}^2 \phi_{z,z}^{-1} + l_j^2 (H_p + Q)^{-1}, \quad j = 2, 3, \ldots, k,
\]

and

\[
\text{Asvar}(\tilde{\alpha}) = (H_p + Q)^{-1}.
\]

The inverse of \( I_f, \) the information matrix based on the full data, can be similarly computed to obtain \( \text{Asvar}(\hat{\beta}_j), \) \( j = 1, \ldots, k, \) and \( \text{Asvar}(\hat{\alpha}). \) These are calculated using identical formulae as above, except \( H_p \) is replaced with \( H_f. \) Thus, Theorem 1 follows from computing the ratio of these asymptotic variances, and the precise formulation stated offers emphasis on the important elements regarding efficiency.
C. Figures
Fig. W1. AREs of the estimator of $\alpha$, the shape parameter of the baseline intensity function, for different designs with 2, 4, and 8 panels, varying values of the covariate effect, $\gamma$ ($x$-axis), and of the overdispersion parameter, $\tau$ ($y$-axis).
Fig. W2. AREs of the estimator of $\beta_1$, the overall mean baseline treatment effect, for different designs with 2, 4, and 8 panels, varying values of the covariate effect ($x$-axis), $\gamma$, and of the overdispersion parameter, $\tau$ ($y$-axis).
Fig. W3. AREs of $\hat{\beta}_2^2$ for the 2-segment model for different designs with 1 and 2 panels per segment. Parameter values are $\alpha = 1.1$, $\beta_1 = -3.5$ for the baseline intensity function; $\beta_1 = 0.2, 0.075, -0.050$ when the segment cut point is at 16, 24, and 32 months, respectively; $\beta_2 = -0.8, -0.925, -1.050$ when the segment cut point is at 16, 24, and 32 months, respectively.
Fig. W4. AREs of the estimate of the time-weighted mean treatment effect, $\hat{\delta}_2$, based on the 2-segment model for different designs with 1 and 2 panels per segment. Parameter values are $\alpha = 1.1$, $\beta_1 = -3.5$ for the baseline intensity function; $\beta^2_1 = 0.2, 0.075, -0.050$ when the segment cut point is at 16, 24, and 32 months, respectively; $\beta^2_2 = -0.8, -0.925, -1.050$ when the segment cut point is at 16, 24, and 32 months, respectively.