Web Appendix A: Demonstration of Unbiasedness of Estimating Function (6)

We provide an argument that (6) of the main paper is an unbiased estimating function for \( \xi \) when the coarsening probabilities may not be correctly specified, so that the posited probabilities \( \pi \{ r, G_r(Z) \} \neq \pi_0 \{ r, G_r(Z) \} \) for some \( r = 1, \ldots, M, \infty \), but the models \( h_r \{ G_r(Z), \xi \} \) are correct, and that estimators for \( \xi \) based on (6) will converge in probability to \( \xi_0 \) for arbitrary choice of the functions \( q_r \{ G_r(Z), \xi \} \).

The argument proceeds by a series of iterated conditional expectations to show that (6) has mean zero under these conditions. When \( r = M \), the summand in (6) is

\[
I(C = \infty)q_M \{ G_M(Z), \xi \} [m(Z) - h_M \{ G_M(Z), \xi \}] . \tag{A.1}
\]

The conditional expectation of (A.1) given \( Z \) is

\[
K_M \{ G_M(Z) \} q_M \{ G_M(Z), \xi \} [m(Z) - h_M \{ G_M(Z), \xi \}] , \tag{A.2}
\]

and the conditional expectation of (A.2) given \( G_M(Z) \) is

\[
K_M \{ G_M(Z) \} q_M \{ G_M(Z), \xi \} [E \{ m(Z) | G_M(Z) \} - h_M \{ G_M(Z), \xi \}] . \tag{A.3}
\]

Under the correctly specified model \( h_M \{ G_M(Z), \xi_0 \} = E \{ m(Z) | G_M(Z) \} \), the expectation of (A.3) is zero when \( \xi = \xi_0 \), and hence (A.1) has expectation zero at \( \xi = \xi_0 \). We may similarly argue that an arbitrary summand in (6) has expectation zero. The conditional expectation of the \( r \)th summand, given \( Z \), is \( K_r \{ G_r(Z) \} q_r \{ G_r(Z), \xi \} [h_r \{ G_{r+1}(Z), \xi \} - h_r \{ G_r(Z), \xi \}] \),
which in turn has conditional expectation given \( G_r(Z) \) equal to \( K_{r0} \{ G_r(Z) \} q_r \{ G_r(Z), \xi \} \times (E[h_{r+1} \{ G_{r+1}(Z), \xi \} | G_r(Z)] - h_r \{ G_r(Z), \xi \}) \). When the \( h_r \{ G_r(Z), \xi \} \) are correctly specified, \( E[h_{r+1} \{ G_{r+1}(Z), \xi_0 \} | G_r(Z)] = E[E \{ m(Z) | G_{r+1}(Z) \} | G_r(Z)] = E \{ m(Z) | G_r(Z) \} = h_r \{ G_r(Z), \xi_0 \} \), where the second-to-last equality follows by CAR. Hence, at \( \xi = \xi_0 \), each summand in (6) has expectation zero, so that (6) is an unbiased estimating function for \( \xi \) even if the coarsening probabilities are misspecified, and estimators for \( \xi \) based on (6) will converge in probability to \( \xi_0 \) for arbitrary choice of the functions \( q_r \{ G_r(Z), \xi \} \).

Web Appendix B: Demonstration of Optimality of (7)

We show that substituting (7) of the main paper into (6) has expectation zero at \( \xi = \xi^{opt} \) when the coarsening probabilities are correctly specified but the functions \( h_r \{ G_r(Z), \xi \} \) may or may not be and hence is an unbiased estimating function under these conditions, so that the proposed estimator \( \hat{\xi}^{opt} \) using this choice converges in probability to \( \xi^{opt} \), ensuring condition (i) in the main paper is satisfied.

Note that (6) may be written as \( S_1 + S_2 + S_3 \), where \( S_1 = I(C = \infty)q_M \{ G_M(Z), \xi \} m(Z), S_2 = -\sum_{r=2}^{M} I(C > r)q_r \{ G_r(Z), \xi \} - I(C > r - 1)q_{r-1} \{ G_{r-1}(Z), \xi \} h_r \{ G_r(Z), \xi \}, \) and \( S_3 = -I(C > 1)q_1 \{ G_1(Z), \xi \} h_1 \{ G_1(Z), \xi \} \). It is straightforward to show, by substituting (7), recalling that \( \pi \{ r, G_r(Z) \} = \pi_0 \{ r, G_r(Z) \} \), and first finding \( E(S_1|Z) \) using \( E \{ I(C = \infty) | Z \} = \pi_0(\infty, Z) \), that we have \( E(S_1) = E[-m(Z) \sum_{j=1}^{M} \lambda_j \{ G_j(Z) \} h_{j\xi} \{ G_j(Z) \} / K_{j0} \{ G_j(Z) \}] \), which matches the first term in (5) of the main paper. Considering an arbitrary summand \( S_{2r}, \) say, in \( S_2 \), and using \( E \{ I(C > r)|Z \} = K_{r0} \{ r, G_r(Z) \} \), we have

\[
E(S_{2r}|Z) = \left[ \sum_{j=1}^{r} \frac{\lambda_j \{ G_j(Z) \} h_{j\xi} \{ G_j(Z), \xi \}}{K_{j0} \{ G_j(Z) \}} - \sum_{j=1}^{r-1} \frac{\lambda_j \{ G_j(Z) \} h_{j\xi} \{ G_j(Z), \xi \}}{K_{j0} \{ G_j(Z) \}} \right] h_r \{ G_r(Z), \xi \},
\]

so that \( E(S_{2r}) = E[\lambda_{r0} \{ G_r(Z) \} h_{r\xi} \{ G_r(Z), \xi \} / K_{r0} \{ G_r(Z) \}] h_r \{ G_r(Z), \xi \} \), and thus \( E(S_2) \) is equal to the second term in (5) except for the summand at \( r = 1 \). An analogous argument
applied to $S_3$ shows that $E(S_3) = E[\lambda_{1,0} \{G_1(Z) \} h_{1\xi} \{G_1(Z), \xi \} / K_{1,0} \{G_1(Z) \} h_{1} \{G_1(Z), \xi \}]$. Combining these results, the estimating function found by substituting (7) into (6) has the same expectation as that of (7) when the coarsening probabilities are correctly specified, which is equal to zero when $\xi = \xi^{opt}$. Thus, $\hat{\xi}^{opt}$ converges in probability to $\xi^{opt}$, ensuring (i) as claimed.

**Web Appendix C: Derivation of Score Vector $S_{\psi} \{C, G_C(Z), \psi \}$**

The likelihood for $\psi$ may be shown to be

$$\prod_{r=1}^M \prod_{i:C_i \geq r} \lambda_r \{G_r(Z_i), \psi \}^{I(C_i=r)} [1 - \lambda_r \{G_r(Z_i), \psi \}]^{I(C_i>r)} ,$$

which may be written equivalently as $\prod_{r=1}^M \prod_{i:C_i \geq r} \lambda_r \{G_r(Z_i), \psi \}^{I(C_i=r)} [1 - \lambda_r \{G_r(Z_i), \psi \}]^{I(C_i>r)}$ (Tsiatis, 2006, Section 8.2). The score vector for $\psi$ may be shown to be

$$S_{\psi} \{C, G_C(Z), \psi \} = \sum_{r=1}^M \frac{\lambda_r \{G_r(Z), \psi \}}{\lambda_r \{G_r(Z), \psi \} [1 - \lambda_r \{G_r(Z), \psi \}]} dM_r \{G_r(Z), \psi \} ,$$

which, multiplying and dividing by $K_r \{G_r(Z), \psi \}$ and noting that $1 - \lambda_r \{G_r(Z), \psi \} = K_r \{G_r(Z), \psi \} / K_{r-1} \{G_r(Z), \psi \}$, may be written as in Section 4 of the main paper.

**Web Appendix D: Derivation of Approximate Standard Errors Via the Sandwich Method**

We provide expressions required to calculate the asymptotic variances of the three estimators for $\beta$ ($p \times 1$) considered in the main paper: $\hat{\beta}_{ipw}$, $\hat{\beta}_{br^*}$, and $\hat{\beta}_{opt^*}$. Let $\tau$ be the collection of unknown parameters involved in obtaining the estimators for $\beta$; in particular, $\tau = (\psi^T, \beta^T)^T$ for $\hat{\beta}_{ipw}$, $\tau = (\psi^T, \xi_1^T, ..., \xi_M^T, \beta^T)^T$ for $\hat{\beta}_{br^*}$, and $\tau = (\psi^T, \xi^T, \theta^T, \beta^T)^T$ for $\hat{\beta}_{opt^*}$. The estimator for $\tau$, $\hat{\tau}$, in each case can be obtained by solving a set of M-estimating equations given by
\( \sum_{i=1}^{n} \rho_i(\tau) = 0 \) (Stefanski and Boos, 2002), where the last \( p \) entries of \( \rho_i(\tau) \) correspond to the estimating equation for \( \beta \), and \( \rho_i(\tau) \) is defined for each estimator below. Let \( A_n = n^{-1} \sum_{i=1}^{n} A_i = n^{-1} \sum_{i=1}^{n} \partial / \partial \tau \{ \rho_i(\tau) \} \), and \( B_n = n^{-1} \sum_{i=1}^{n} \rho_i(\tau) \rho_i^T(\tau) \). Following standard theory, the asymptotic covariance matrix of \( \hat{\tau} \) can be approximated by the empirical sandwich matrix \( V_n = n^{-1} A_n^{-1} B_n (A_n^{-1})^T \). Therefore, the asymptotic variances of the three estimators can be approximated by the lower, rightmost diagonal \((p \times p)\) submatrix of the corresponding matrix \( V_n \). We present the form of \( \rho_i(\tau) \) and \( A_i \) for each of the estimators, from which the form of \( V_n \) may be calculated. The desired diagonal submatrix of \( V_n \) may then be obtained numerically, with the required matrix inversion carried out by standard routines.

Throughout, we assume that \( \lambda_r \{ G_r(Z), \psi_r \}, r = 1, \ldots, M, \) are logistic regression models, and \( \psi = (\psi_1^T, \ldots, \psi_M^T)^T \) are estimated via separate ML fits for each \( r = 1, \ldots, M, \) where \( \tilde{X}_{i,r} \) is a row vector consisting of the covariates used in the modeling of \( \lambda_r \{ G_r(Z_i), \psi_r \} \), including a “1” for the intercept term. For \( \hat{\beta}_{ipw} \), \( \rho_i(\tau) \) is given by

\[
\rho_i(\tau) = \begin{pmatrix}
\sum_{r=1}^{M} \frac{dM_r \{ G_r(Z_i), \psi \} K_{r-1} \{ G_r(Z_i), \psi \} \lambda_r \psi \{ G_r(Z_i), \psi \}}{K_r \{ G_r(Z_i), \psi \}} \\
\frac{I(C_i = \infty)m(Z_i, \beta)}{\pi(\infty, Z_i, \psi)} \\
\end{pmatrix}
\]

\[
= \begin{pmatrix}
dM_1 \{ G_1(Z_i), \psi_1 \} \tilde{X}_{i,1}^T \\
\vdots \\
dM_M \{ G_M(Z_i), \psi_M \} \tilde{X}_{i,M}^T \\
\frac{I(C_i = \infty)m(Z_i, \beta)}{\pi(\infty, Z_i, \beta)} \\
\end{pmatrix},
\]
and $A_i$ is given by

$$
A_i = \begin{pmatrix}
D_{i,1} & 0 & \cdots & \cdots & \cdots & 0 \\
0 & \ddots & 0 & \cdots & \cdots & 0 \\
0 & 0 & D_{i,r} & 0 & \cdots & 0 \\
0 & \cdots & 0 & \ddots & 0 & 0 \\
0 & \cdots & \cdots & 0 & D_{i,M} & 0 \\
E_{i,1} & \cdots & E_{i,r} & \cdots & E_{i,M} & D_{i,\beta}
\end{pmatrix},
$$

where

$$
D_{i,r} = -I(C_i \geq r) \lambda_r \{G_r(Z_i), \psi_r\} [1 - \lambda_r \{G_r(Z_i), \psi_r\}] \tilde{X}_i^T \tilde{X}_i, \quad r = 1, \ldots, M,
$$

$$
E_{i,r} = \frac{I(C_i = \infty) m(Z_i, \beta)}{\pi(\infty, Z_i, \psi)} \lambda_r \{G_r(Z_i), \psi_r\} \tilde{X}_i, \quad r = 1, \ldots, M,
$$

$$
D_{i,\beta} = \frac{I(C_i = \infty)}{\pi(\infty, Z_i, \psi)} m_\beta(Z_i, \beta),
$$

and $m_\beta(Z_i, \beta)$ is a column vector of partial derivatives of $m(Z_i, \beta)$ with respect to $\beta$.

We implemented $\hat{\beta}_{hr}$ as described in Bang and Robins (2005); i.e., we added as a covariate $\hat{K}_r^{-1} \{G_r(Z), \hat{\psi}_1, \ldots, \hat{\psi}_r\}$ in the conditional mean functions $h_r^* \{G_r(Z), \xi_r\}$ corresponding to a generalized linear model with canonical link, where $\hat{K}_r^{-1} \{G_r(Z), \hat{\psi}_1, \ldots, \hat{\psi}_r\}$ is an estimate for the true cumulative hazard $K_r^{-1} \{G_r(Z), \psi_1, \ldots, \psi_r\}, r = 1, \ldots, M$. We write the new conditional mean function including additional covariate $\hat{K}_r^{-1} \{G_r(Z), \hat{\psi}_1, \ldots, \hat{\psi}_r\}$ as
\( h_r^* \{ G_r(Z), \psi_1, \ldots, \psi_r, \xi_r, \beta \} \). For this estimator, \( \rho_i(\tau) \) is given by

\[
\rho_i(\tau) = \begin{pmatrix}
  dM_1 \{ G_1(Z_i), \psi_1 \} \tilde{X}_{i,1}^T \\
  \vdots \\
  dM_M \{ G_M(Z_i), \psi_M \} \tilde{X}_{i,M}^T \\
  I(C_i > 1) \left[ h_2^* \{ G_2(Z_i), \psi_1, \psi_2, \xi_2, \beta \} - h_1^* \{ G_1(Z_i), \psi_1, \xi_1, \beta \} \right] \\
  \times h_{1,\psi_1,\xi_1}^* \{ G_1(Z_i), \psi_1, \xi_1, \beta \} \\
  \vdots \\
  I(C_i > M) \left[ m(Z_i, \beta) - h_M^* \{ G_M(Z_i), \psi, \xi_M, \beta \} \right] \\
  \times h_{M,\psi,\xi_M}^* \{ G_M(Z_i), \psi, \xi_M, \beta \} \\
  h_i^* \{ G_1(Z_i), \psi_1, \xi_1, \beta \}
\end{pmatrix},
\]

where \( h_{r,\psi_1,\ldots,\psi_r,\xi_r}^* \{ G_r(Z_i), \psi_1, \ldots, \psi_r, \xi_r, \beta \} \) is the column vector of partial derivatives of \( h_r^* \{ G_r(Z_i), \psi_1, \ldots, \psi_r, \xi_r, \beta \} \) with respect to \( \psi_1, \ldots, \psi_r, \xi_r, r = 1, \ldots, M \).

The matrix \( A_i \) is given by

\[
A_i = \begin{pmatrix}
  A_{1i} \\
  A_{2i} \\
  A_{3i}
\end{pmatrix},
\]

\[
A_{1i} = \begin{pmatrix}
  D_{i,1} & 0 & \cdots & \cdots & \cdots & 0 \\
  0 & \ddots & 0 & \cdots & \cdots & 0 \\
  0 & 0 & D_{i,r} & 0 & \cdots & 0 \\
  0 & \cdots & 0 & \ddots & 0 & 0 \\
  0 & \cdots & \cdots & 0 & D_{i,M} & 0
\end{pmatrix},
\]

and \( A_{3i} = \begin{pmatrix}
  F_{1,i} & 0 & \cdots & 0 & F_{2,i} & 0 & \cdots & 0 & F_{3,i}
\end{pmatrix}, \)
where

\[ D_{i,r} = -I(C_i \geq r) \lambda_r \{ G_r(Z_i), \psi_r \} \left[ 1 - \lambda_r \{ G_r(Z_i), \psi_r \} \right] \tilde{X}_{i,r}^T \tilde{X}_{i,r}, \quad r = 1, \ldots, M, \]

\[ F_{1,i} = h^*_1 \{ G_1(Z_i), \psi_1, \xi_1, \beta \}, \quad F_{2,i} = h^*_1 \{ G_1(Z_i), \psi_1, \xi_1, \beta \}, \]

\[ F_{3,i} = h^*_1 \{ G_1(Z_i), \psi_1, \xi_1, \beta \}; \]

i.e., \( F_{1,i}, F_{2,i}, F_{3,i} \) are partial derivatives of \( h^*_1 \{ G_1(Z_i), \psi_1, \xi_1, \beta \} \) with respect to \( \psi_1, \xi_1, \) and \( \beta, \) respectively. The \( A_{2i} \) term involves the partial derivatives of the column vector

\[
\rho_{2,i}(\tau) = \begin{cases} 
I(C_i > 1) [h^*_2 \{ G_2(Z_i), \psi_1, \psi_2, \xi_2, \beta \} - h^*_1 \{ G_1(Z_i), \psi_1, \xi_1, \beta \}] \\
\times h^*_1 \{ G_1(Z_i), \psi_1, \xi_1, \beta \} \\
\vdots \\
I(C_i > M) [m(Z_i, \beta) - h^*_M \{ G_M(Z_i), \psi, \xi_M, \beta \}] \\
\times h^*_M \{ G_M(Z_i), \psi, \xi_M, \beta \}
\end{cases}
\]

with respect to \( \tau. \) Often in practice, it is cumbersome to obtain the analytical derivatives of \( \rho_{2,i}(\tau) \) with respect to \( \tau. \) In our implementation, we used numerical derivatives as an approximation to the analytical derivatives. For example, to calculate the derivative of \( \rho_{2,i}(\tau) \) with respect to the \( k \)th element of \( \tau, \) we used a one-sided numerical approximation of the form \( \{ \rho_{2,i}(\tau + \epsilon 1_k) - \rho_{2,i}(\tau) \} / \epsilon \) for small enough \( \epsilon > 0, \) where \( 1_k \) is a column vector with 1 on the \( k \)th entry and all other entries 0.
For $\tilde{\beta}_{opt}$, $\rho_i(\tau)$ is given by

$$
\rho_i(\tau) = \begin{pmatrix}
\sum_{r=1}^{M} I(C_i > r) \bar{q}_r \{ G_r(Z_i), \tilde{\xi}, \psi \} \
I(C_i = \infty) m(Z_i, \beta) / \pi(\infty, Z_i, \psi) + \sum_{r=1}^{M} \frac{dM_r \{ G_r(Z_i), \psi \}}{K_r \{ G_r(Z_i), \psi \}} h_r \{ G_r(Z_i), \xi \}
\end{pmatrix},
$$

where

$$
\bar{h}_r \{ G_r(Z_i), \tilde{\xi} \} = h_r \{ G_r(Z_i), \xi \} - \theta^r \frac{K_{r-1} \{ G_r(Z_i), \psi \} \lambda_r \{ G_r(Z_i), \psi \}}{\lambda_r \{ G_r(Z_i), \psi \}}
$$

and

$$
\bar{q}_r \{ G_r(Z_i), \tilde{\xi}, \psi \} = -[K_r \{ G_r(Z_i), \psi \}]^{-1} \sum_{j=1}^{r} \lambda_j \{ G_j(Z_i), \psi \} \left( \begin{array}{c}
\bar{h}_j \xi \{ G_j(Z_i), \tilde{\xi}, \psi \} \\
\bar{h}_j \theta \{ G_j(Z_i), \tilde{\xi}, \psi \}
\end{array} \right),
$$

$$
\bar{h}_j \theta \{ G_j(Z_i), \tilde{\xi}, \psi \} = -K_{j-1} \{ G_j(Z_i), \psi \} \lambda_j \frac{\lambda_j \{ G_j(Z_i), \psi \}}{\lambda_j \{ G_j(Z_i), \psi \}},
$$

$$
\bar{h}_j \xi \{ G_j(Z_i), \tilde{\xi}, \psi \} = h_j \xi \{ G_j(Z_i), \xi, \psi \}.
$$

The matrix $A_i$ is given by

$$
A_i = \begin{pmatrix}
A_{1i} \\
A_{2i}
\end{pmatrix}, \quad A_{1i} = \begin{pmatrix}
D_{i,1} & 0 & \cdots & \cdots & \cdots & 0 \\
0 & \ddots & 0 & \cdots & \cdots & 0 \\
0 & 0 & D_{i,r} & 0 & \cdots & 0 \\
0 & \cdots & 0 & \ddots & 0 & 0 \\
0 & \cdots & \cdots & 0 & D_{i,M} & 0
\end{pmatrix},
$$

and

$$
A_{2i} = \left( \partial / \partial \tau \{ \rho_{3,i}(\tau) \} \right),
$$

\[8\]
where

\[
D_{i,r} = -\mathbb{I}(C_i \geq r) \lambda_r \{G_r(Z_i), \psi_r\} [1 - \lambda_r \{G_r(Z_i), \psi_r\}] \tilde{X}_i^T \tilde{X}_{i,r}, \quad r = 1, \ldots, M,
\]

\[
\rho_{3,i}(\tau) = \left\{ \begin{array}{l}
\sum_{r=1}^{M} I(C_i > r) \tilde{q}_r \left\{ G_r(Z_i), \tilde{\xi}, \psi \right\} \left[ \tilde{h}_{r+1} \left\{ G_{r+1}(Z_i), \tilde{\xi}, \psi \right\} - \tilde{h}_r \left\{ G_r(Z_i), \tilde{\xi}, \psi \right\} \right] \\
\frac{I(C_i = \infty) m(Z_i, \beta)}{\pi(\infty, Z_i, \psi)} + \sum_{r=1}^{M} \frac{dM_r \{G_r(Z_i), \psi\}}{K_r \{G_r(Z_i), \psi\}} h_r \{G_r(Z_i), \xi\}
\end{array} \right. 
\]

Analogous to the strategy for \( \hat{\beta}_{br}^* \), in our implementation, we used numerical derivatives as an approximation to the analytical derivatives of \( \rho_{3,i}(\tau) \) with respect to \( \tau \).

Web Appendix E: Derivation of Conditional Expectations Implied by Assumed Mixed Models in Section 5

We derive the required conditional expectations \( E(Y|Y_1, \ldots, Y_j, \tilde{X}) \) for \( j = 1, \ldots, 4 \) implied by the linear mixed model (14) in Section 5 of the main paper. The random vector \( \Psi = (\alpha_0, \alpha_1, e_1, e_2, e_3, e_4)^T \) has multivariate normal distribution with mean \( \mu \) and variance \( \Sigma \), where

\[
\mu = \left( \mu_{a0}, \mu_{a1}, 0_{1 \times 4} \right)^T, \quad \Sigma = \begin{pmatrix}
\Sigma_a & 0_{2 \times 4} \\
0_{2 \times 4} & \sigma^2_e I_4
\end{pmatrix},
\]

0\(_{a \times b}\) is a zero matrix with dimension \((a \times b)\), and \( I_a \) is an \((a \times a)\) identity matrix. Therefore, the distribution of \((\alpha_0, \alpha_1, Y_1, Y_2, Y_3, Y_4)^T\), conditional on \( \tilde{X} \), follows multivariate normal distribution with mean \( \tilde{\mu} = A\mu + c \) and variance \( \tilde{\Sigma} = A\Sigma A^T \), where

\[
A = I_6 + \begin{pmatrix}
0_{2 \times 2} & 0_{2 \times 4} \\
A_{21} & 0_{4 \times 4}
\end{pmatrix}, \quad A_{21} = \begin{pmatrix}
1 & 1 & 1 & 1 \\
t_1 & t_2 & t_3 & t_4
\end{pmatrix}^T, \quad \text{and} \quad c = \gamma^T \tilde{X} \begin{pmatrix}
0_{2 \times 1} \\
1_{4 \times 1}
\end{pmatrix}.
\]
Hence, the conditional mean is given by

\[
E(Y|Y_1, Y_2, Y_3, Y_4, \bar{X}) = E(Y_5|Y_1, Y_2, Y_3, Y_4, \bar{X}) = \gamma^T X + E(\alpha_0|Y_1, Y_2, Y_3, Y_4, \bar{X}) + t_5 E(\alpha_1|Y_1, Y_2, Y_3, Y_4, \bar{X}).
\]

To calculate the conditional mean \( E(\alpha_k|Y_1, Y_2, Y_3, Y_4, \bar{X}), \ k = 0, 1, \) we use the following property of multivariate normal distribution. Suppose \((X_1^T, X_2^T)^T\) follows a \(N(\mu, \Omega)\) distribution. If \(\mu\) and \(\Omega\) are partitioned correspondingly as follows:

\[
v = \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} \quad \text{and} \quad \Omega = \begin{pmatrix} \Omega_{11} & \Omega_{12} \\ \Omega_{21} & \Omega_{22} \end{pmatrix},
\]

then \((X_1|X_2 = a) \sim N(\bar{\mu}, \overline{\Omega})\), where \(\bar{\mu} = v_1 + \Omega_{12} \Omega_{22}^{-1} (a - v_2)\). Straightforward application of the above property yields

\[
E \{(\alpha_0, \alpha_1)^T|Y_1, \ldots, Y_4, \bar{X}\} = \bar{\mu}_{1:2} + \bar{\Sigma}_{1:2,3:6} \bar{\Sigma}_{3:6,3:6}^{-1} \{(Y_1, Y_2, Y_3, Y_4)^T - \bar{\mu}_{3:6}\},
\]

where \(\bar{\mu}_{a:b}\) is a column vector consisting of \(a\)th to \(b\)th entries of \(\bar{\mu}\), and \(\bar{\Sigma}_{a:b,m:n}\) is the submatrix of \(\bar{\Sigma}\) with rows \(a\) to \(b\) and columns \(m\) to \(n\). Therefore the conditional expectation is

\[
E(Y|Y_1, Y_2, Y_3, Y_4, \bar{X}) = \gamma^T \bar{X} + (1, t_5) \left[ \bar{\mu}_{1:2} + \bar{\Sigma}_{1:2,3:6} \bar{\Sigma}_{3:6,3:6}^{-1} \{(Y_1, Y_2, Y_3, Y_4)^T - \bar{\mu}_{3:6}\} \right].
\]

Similarly,

\[
E(Y|Y_1, Y_2, Y_3, \bar{X}) = \gamma^T \bar{X} + (1, t_5) \left[ \bar{\mu}_{1:2} + \bar{\Sigma}_{1:2,3:5} \bar{\Sigma}_{3:5,3:5}^{-1} \{(Y_1, Y_2, Y_3)^T - \bar{\mu}_{3:5}\} \right],
\]

\[
E(Y|Y_1, Y_2, \bar{X}) = \gamma^T \bar{X} + (1, t_5) \left[ \bar{\mu}_{1:2} + \bar{\Sigma}_{1:2,3:4} \bar{\Sigma}_{3:4,3:4}^{-1} \{(Y_1, Y_2)^T - \bar{\mu}_{3:4}\} \right],
\]

\[
E(Y|Y_1, \bar{X}) = \gamma^T \bar{X} + (1, t_5) \left[ \bar{\mu}_{1:2} + \bar{\Sigma}_{1:2,3:3} \bar{\Sigma}_{3:3,3:3}^{-1} (Y_1 - \bar{\mu}_{3:3}) \right].
\]

Next, we provide the derivation of the conditional expectations

\[
E(Y|Y_1, \ldots, Y_j, \bar{X}, \text{dis}_1, \text{dis}_2, \text{dis}_3, \text{dis}_4)
\]
for $j = 1, \ldots, 4$ implied by assumed linear mixed model used in the second, general coarsened data analysis in Section 5 of the main paper; i.e., we assumed that, for $r = 1, \ldots, 5$, the data follow the linear mixed model

$$Y_{ir} = \alpha_{0i} + \alpha_{1i}t_{ir} + \gamma^T \tilde{X}_i + \phi_1 I(r \geq 3)\text{dis}_{i2} + \phi_2 I(r = 5)\text{dis}_{i4} + e_{ir},$$

where the random effects and within-subject deviations are normal as above, and now $\tilde{X} = (\text{weight,karnof,symp})$.

Following the same logic as above, the distribution of $(\alpha_0, \alpha_1, Y_1, Y_2, Y_3, Y_4)^T$, conditional on $(\tilde{X}, \text{dis}_1, \text{dis}_2, \text{dis}_3, \text{dis}_4)$, follows multivariate normal distribution with mean $\bar{\mu}^* = A\mu + \bar{c}$ and variance $\bar{\Sigma} = A\Sigma A^T$, where $A, \mu, \Sigma, \bar{\Sigma}$ are the same as above, and

$$\bar{c} = \begin{pmatrix} 0_{1 \times 2}, \gamma^T \tilde{X}, \gamma^T \tilde{X}, \gamma^T \tilde{X} + \phi_1 \text{dis}_2, \gamma^T \tilde{X} + \phi_1 \text{dis}_2 \end{pmatrix}^T.$$

The conditional expectations are given as follows:

$$E(Y|Y_1, Y_2, Y_3, Y_4, \tilde{X}, \text{dis}_1, \text{dis}_2, \text{dis}_3, \text{dis}_4)$$

$$= \gamma^T \tilde{X} + \phi_1 \text{dis}_2 + \phi_2 \text{dis}_4 + (1, t_5) \left[ \bar{\mu}_{1:2}^* + \bar{\Sigma}_{1:2,3:6} \bar{\Sigma}_{3:6,3:6}^{-1} (Y_1, Y_2, Y_3, Y_4)^T - \bar{\mu}_{3:6}^* \right],$$

$$E(Y|Y_1, Y_2, Y_3, \tilde{X}, \text{dis}_1, \text{dis}_2, \text{dis}_3, \text{dis}_4)$$

$$= \gamma^T \tilde{X} + \phi_1 \text{dis}_2 + \phi_2 \text{dis}_4 + (1, t_5) \left[ \bar{\mu}_{1:2}^* + \bar{\Sigma}_{1:2,3:5} \bar{\Sigma}_{3:5,3:5}^{-1} (Y_1, Y_2, Y_3)^T - \bar{\mu}_{3:5}^* \right],$$

$$E(Y|Y_1, Y_2, \tilde{X}, \text{dis}_1, \text{dis}_2, \text{dis}_3, \text{dis}_4)$$

$$= \gamma^T \tilde{X} + \phi_1 \text{dis}_2 + \phi_2 \text{dis}_4 + (1, t_5) \left[ \bar{\mu}_{1:2}^* + \bar{\Sigma}_{1:2,3:4} \bar{\Sigma}_{3:4,3:4}^{-1} (Y_1, Y_2)^T - \bar{\mu}_{3:4}^* \right],$$

$$E(Y|Y_1, \tilde{X}, \text{dis}_1, \text{dis}_2, \text{dis}_3, \text{dis}_4)$$

$$= \gamma^T \tilde{X} + \phi_1 \text{dis}_2 + \phi_2 \text{dis}_4 + (1, t_5) \left[ \bar{\mu}_{1:2}^* + \bar{\Sigma}_{1:2,3:3} \bar{\Sigma}_{3:3,3:3}^{-1} (Y_1 - \bar{\mu}_{3:3}^*) \right].$$
Therefore, used in Section 6 of the main paper. The model implies that, in truth,

\[ E(Y|L) = E \{ E(Y|L_2, \alpha_0, \alpha_1)|L_2 \} = \gamma^TX + \mu_1(X, Y_1, Y_2) + t_3\mu_2(X, Y_1, Y_2), \]

where \( \mu_1(X, Y_1, Y_2) = E(\alpha_0|X, Y_1, Y_2) \), and \( \mu_2(X, Y_1, Y_2) = E(\alpha_1|X, Y_1, Y_2) \).

Thus, we need to calculate the conditional distribution of \( \alpha_0, \alpha_1 \) given \( X, Y_1, Y_2 \). The joint density of \( (\alpha_0, \alpha_1, X, Y_1, Y_2)^T \) is given by

\[ f(\alpha_0, \alpha_1, X, Y_1, Y_2) = f(Y_2|\alpha_0, \alpha_1, X, Y_1)f(Y_1|\alpha_0, \alpha_1, X)f(X)f(\alpha_0, \alpha_1) \]

Therefore,

\[ f(\alpha_0, \alpha_1|X, Y_1, Y_2) = \frac{f(\alpha_0, \alpha_1, X, Y_1, Y_2)}{\int f(\alpha_0, \alpha_1, X, Y_1, Y_2)d\alpha_0d\alpha_1} = \frac{f(Y_2|\alpha_0, \alpha_1, X, Y_1)f(Y_1|\alpha_0, \alpha_1, X)f(\alpha_0, \alpha_1)}{\int f(Y_2|\alpha_0, \alpha_1, X, Y_1)f(Y_1|\alpha_0, \alpha_1, X)f(\alpha_0, \alpha_1)d\alpha_0d\alpha_1}. \]

As a consequence,

\[ f(\alpha_0, \alpha_1|X, Y_1, Y_2) \propto f(Y_2|\alpha_0, \alpha_1, X, Y_1)f(Y_1|\alpha_0, \alpha_1, X)f(\alpha_0, \alpha_1). \]

After some algebra, it can be shown that, if we let \( a = \sigma_{22}/(\sigma_{11}\sigma_{22} - \sigma_{12}^2) \), \( b = -\sigma_{12}/(\sigma_{11}\sigma_{22} - \sigma_{12}^2) \), \( c = \sigma_{11}/(\sigma_{11}\sigma_{22} - \sigma_{12}^2) \), \( g_1(X, Y_1, Y_2) = a\mu_{ao} + b\mu_{ai} + (Y_2 + Y_1 - 2\gamma^TX)/\sigma_e^2 \), and \( g_2(X, Y_1, Y_2) = b\mu_{ao} + c\mu_{ai} + (Y_2 - \gamma^TX)/\sigma_e^2 \), then

\[ \mu_2(X, Y_1, Y_2) = E(\alpha_1|Z, Y_1, Y_2) = \frac{g_1(X, Y_1, Y_2) (b + 1/\sigma_e^2) - g_2(X, Y_1, Y_2) (a + 2/\sigma_e^2)}{(b + 1/\sigma_e^2)^2 - (c + 1/\sigma_e^2) (a + 2/\sigma_e^2)}, \]

\[ \mu_1(X, Y_1, Y_2) = E(\alpha_0|Z, Y_1, Y_2) = \frac{g_2(X, Y_1, Y_2) - \mu_2(X, Y_1, Y_2) (c + 1/\sigma_e^2)}{b + 1/\sigma_e^2}. \]
Similarly, we have

\[
E(Y|\mathcal{L}_1) = E\{E(Y|\mathcal{L}_1, \alpha_0, \alpha_1)|\mathcal{L}_1\} = \gamma^T X + \mu_3(X, Y_1) + t_3 \mu_4(X, Y_1),
\]

where \( \mu_3(X, Y_1) = E(\alpha_0|X, Y_1) \), and \( \mu_4(X, Y_1) = E(\alpha_1|X, Y_1) \). Letting \( d = b \mu_{\alpha_0} + c \mu_{\alpha_1} \),
\[ g_3(X, Y_1) = a \mu_{\alpha_0} + b \mu_{\alpha_1} + (Y_1 - \gamma^T X)/\sigma^2_e, \]
we have
\[
\begin{align*}
\mu_3(X, Y_1) &= E(\alpha_0|X, Y_1) = \frac{g_3(X, Y_1) \cdot c - d \cdot b}{(a + 1/\sigma^2_e) c - b^2}, \\
\mu_4(X, Y_1) &= E(\alpha_1|X, Y_1) = \frac{d - \mu_3(X, Y_1) \cdot b}{c}.
\end{align*}
\]

Web Appendix G: Sample SAS Code for Implementation of the Analyses in Section 5

```sas
/*****************************/
/*  Fit discrete hazard models */
/*****************************/
ods listing close;
proc logistic data=actg2 descending;
  model C1=cd40 age wtkg hemo homo drugs karnof z30
          preanti race gender str2 symptom/selection=forward slentry=0.15;
  output out=prob1 (drop=_level_) predicted=phat1;
ods output ParameterEstimates=parestc1;
run;
ods listing;

proc sql noprint;
  select variable into :parc1 separated by ' ' 
from parestc1 where variable ne 'Intercept';
quit;
%put &parc1;
ods listing close;
proc logistic data=prob1 descending;
  model C2=cd40 cd420 age wtkg hemo homo drugs karnof z30
          preanti race gender str2 symptom /selection=forward slentry=0.15;
  output out=prob2 (drop=_level_) predicted=phat2;
```
ods output ParameterEstimates=parestc2;
run;
ods listing;
proc sql noprint;
select variable into :parc2 separated by ','
from parestc2 where variable ne 'Intercept';
quit;
%put &parc2;
ods listing close;
proc logistic data=prob2 descending;
model C3=cd40 cd420 cd440 age wtkg hemo homo drugs karnof z30
preanti race gender str2 symptom/selection=forward slentry=0.15;
output out=prob3 (drop=_level_) predicted=phat3;
ods output ParameterEstimates=parestc3;
run;
ods listing;
proc sql noprint;
select variable into :parc3 separated by ','
from parestc3 where variable ne 'Intercept';
quit;
%put &parc3;
ods listing close;
proc logistic data=prob3 descending;
model C4=cd40 cd420 cd440 cd460 age wtkg hemo homo drugs karnof
z30 preanti race gender str2 symptom/selection=forward slentry=0.15;
output out=prob4 (drop=_level_) predicted=phat4;
ods output ParameterEstimates=parestc4;
run;
ods listing;
proc sql noprint;
select variable into :parc4 separated by ','
from parestc4 where variable ne 'Intercept';
quit;
%put &parc4;
/****************************/
/*                        */
/*                   */
/****************************/
data IPW;
set prob4;
pi_5=(1-phat1)*(1-phat2)*(1-phat3)*(1-phat4);
ipi_5=1/pi_5;
\[
\begin{align*}
\pi_4 &= (1 - \text{phat1}) \times (1 - \text{phat2}) \times (1 - \text{phat3}) \\
\text{ipi}_4 &= 1 / \pi_4 \\
\pi_3 &= (1 - \text{phat1}) \times (1 - \text{phat2}) \\
\text{ipi}_3 &= 1 / \pi_3 \\
\pi_2 &= (1 - \text{phat1}) \\
\text{ipi}_2 &= 1 / \pi_2 \\
w_{cd496} &= \text{cd496} / \pi_5 \\
\text{proc means data=IPW sum noprint;}
\text{var wcd496;}
\text{output out=ipwsum sum=wsum_cd496;}
\text{run;}
\text{data ipwresults;}
\text{set ipwsum;}
w_{cd496} &= 500 \times \text{wsum_cd496} / \_\text{freq} \\
\text{run;}
\text{quit;}
\end{align*}
\]

/*********************/
/*
/* Bang-Robins estimator */
/*
/*********************/

/* Fit regression models */

ods listing close;
\text{proc reg data=ipw;}
\text{model cd496=cd40 cd420 cd440 cd460 wtkg karnof str2 symptom;}
\text{ods output ParameterEstimates=parm5;}
\text{run;}
\text{quit;}
ods listing;
*ods trace off;

\text{proc sql noprint;}
\text{select variable into :parm5 separated by , from parm5 where variable ne 'Intercept';}
\text{quit;}
%put &parm5;
ods listing close;
\text{proc reg data=ipw outest=parest5 ;}
\text{model cd496=ipi_5 &parm5;}
*where c3=0;
\text{output out=ipw2 predicted=h4;}

15
run;
quit;
ods listing;

ods listing close;
proc reg data=ipw2;
  model h4=cd40 cd420 cd440 wtkg karnof str2 symptom ;
  ods output ParameterEstimates=parm4;
run;
quit;
ods listing;

proc sql noprint;
  select variable into :parm4 separated by ', ' from parm4 where variable ne 'Intercept';
quit;
%put &parm4;
ods listing close;
proc reg data=ipw2 outest=parest4;
  model h4=ipi_4 &parm4;
  *where c2=0;
  output out=ipw3 predicted=h3;
run;
quit;
ods listing;

ods listing close;
proc reg data=ipw3;
  model h3=cd40 cd420 wtkg karnof str2 symptom;
  ods output ParameterEstimates=parm3;
run;
quit;
ods listing;

proc sql noprint;
  select variable into :parm3 separated by ', ' from parm3 where variable ne 'Intercept';
quit;
%put &parm3;
ods listing close;
proc reg data=ipw3 outest=parest3;
  model h3=ipi_3 &parm3;
  *where c1=0;
  output out=ipw4 predicted=h2;
run;
quit;
ods listing;
ods listing close;
proc reg data=ipw4;
model h2= cd40 wtkg karnof str2 symptom;
ods output ParameterEstimates=parm2;
run;
quit;
ods listing;
proc sql noprint;
select variable into :parm2 separated by ', ' from parm2 where variable ne 'Intercept';
quit;
%put &parm2;
ods listing close;
proc reg data=ipw4 outest=parest2;
model h2=ipi_2 &parm2;
output out=ipw5 predicted=h1;
run;
quit;
ods listing;
proc means data=ipw5 noprint;
var h1;
output out=meany mean=mu_br;
run;
data meany;
set meany;
mu_br=mu_br*500;
run;
proc print data=meany;
run;
quit;

/***************************************
/*
/*  New estimator
/*
/*
***************************************/
ods listing close;
proc logistic data=actg2 descending;
model C1=cd40 age wtkg hemo homo drugs karnof z30 preanti race gender str2 symptom/selection=forward slentry=0.15;
output out=prob1 (drop=_level_) predicted=phat1;
ods output ParameterEstimates=parestc1;
run;
ods listing;

proc sql noprint;
   select variable into :parc1 separated by ', ' from parestc1 where variable ne 'Intercept';
quit;
%put &parc1;
ods listing close;
proc logistic data=prob1 descending;
   model C2=cd40 cd420 age wtkg hemo homo drugs karnof z30 preanti race gender str2 symptom /selection=forward slentry=0.15;
   output out=prob2 (drop=_level_) predicted=phat2;
ods output ParameterEstimates=parestc2;
run;
ods listing;

proc sql noprint;
   select variable into :parc2 separated by ', ' from parestc2 where variable ne 'Intercept';
quit;
%put &parc2;
ods listing close;
proc logistic data=prob2 descending;
   model C3=cd40 cd420 cd440 age wtkg hemo homo drugs karnof z30 preanti race gender str2 symptom/selection=forward slentry=0.15;
   output out=prob3 (drop=_level_) predicted=phat3;
ods output ParameterEstimates=parestc3;
run;
ods listing;

proc sql noprint;
   select variable into :parc3 separated by ', ' from parestc3 where variable ne 'Intercept';
quit;
%put &parc3;
ods listing close;
proc logistic data=prob3 descending;
   model C4=cd40 cd420 cd440 cd460 age wtkg hemo homo drugs karnof z30 preanti race gender str2 symptom/selection=forward slentry=0.15;
   output out=prob4 (drop=_level_) predicted=phat4;
ods output ParameterEstimates=parestc4;
run;
ods listing;
PROC SQL NOPRINT;
SELECT VARIABLE INTO :PARC4 SEPARATED BY ' ' FROM PARESTC4 WHERE VARIABLE NE 'Intercept';
QUIT;
%PUT &PARC4;

DATA IPW;
SET PROB4;
PI_5=(1-PHAT1)*(1-PHAT2)*(1-PHAT3)*(1-PHAT4);
IPI_5=1/PI_5;
PI_4=(1-PHAT1)*(1-PHAT2)*(1-PHAT3);
IPI_4=1/PI_4;
PI_3=(1-PHAT1)*(1-PHAT2);
IPI_3=1/PI_3;
PI_2=(1-PHAT1);
IPI_2=1/PI_2;

WCd496=CD496/PI_5;
RUN;

PROC MEANS DATA=IPW SUM NOPRINT;
VAR WCd496;
OUTPUT OUT=IPWSUM SUM=WSUM_CD496;
RUN;

DATA IPWRESULTS;
SET IPWSUM;
CD496=500*WSUM_CD496/_FREQ_;;
RUN;

PROC PRINT DATA=IPWRESULTS;
RUN;

PROC IML SYMSIZE=82920000 WORKSIZE=829200000;
USE IPW;
READ ALL VAR{CD40 CD420 CD440 CD460 CD496 C1 C2 C3 C4 PHAT1 PHAT2 PHAT3 PHAT4 PI_2 PI_3 PI_4 PI_5 DIS1 DIS2 DIS4 WTKG KARNOF OPRIOR SYMPTOM AGE DRUGS PREATN STR2 HOMO HEMO RACE} INTO FULLDATA;

Y1=FULLDATA[,1];
Y2=FULLDATA[,2];
Y3=FULLDATA[,3];
Y4=FULLDATA[,4];
y5=fulldata[,5];
/* missingness indicator */
c1=fulldata[,6];
c2=fulldata[,7];
c3=fulldata[,8];
c4=fulldata[,9];
/* discrete and cumulative hazard function */
lambda1=fulldata[,10];
lambda2=fulldata[,11];
lambda3=fulldata[,12];
lambda4=fulldata[,13];
k1=fulldata[,14];
k2=fulldata[,15];
k3=fulldata[,16];
k4=fulldata[,17];
n=nrow(y1); /* number of observations */
a=j(n,1,1);

X=fulldata[,21:22] || fulldata[,28] || fulldata[,24];
p=ncol(X); /* number of columns for X, p=4 */

use ipw;
read all var{ &parc1 } into x_lam1;
read all var{ &parc2 } into x_lam2;
read all var{ &parc3 } into x_lam3;
read all var{ &parc4 } into x_lam4;

x_lam1=a || x_lam1; x_lam2=a || x_lam2;
x_lam3=a || x_lam3; x_lam4=a || x_lam4;

p1=ncol(x_lam1);
p2=ncol(x_lam2);
p3=ncol(x_lam3);
p4=ncol(x_lam4);

tp=6+p+p1+p2+p3+p4; /*mu_alpha0, mu_alpha1, sigma1^2, sigma2^2,
sigma12, sigma^2 */
time={0 1 2 3 4.8};

rindex2=0;
do i=1 to n;
   if c1[i]=0 then rindex2=rindex2//i;
end;
rindex2=rindex2[2:nrow(rindex2),];

    rindex3=0;
    do i=1 to n;
        if c2[i]=0 then rindex3=rindex3//i;
    end;
    rindex3=rindex3[2:nrow(rindex3),];

rindex4=0;
    do i=1 to n;
        if c3[i]=0 then rindex4=rindex4//i;
    end;
    rindex4=rindex4[2:nrow(rindex4),];

y12=y1 || y2;
y123=y1 || y2 || y3;
y1234=y1 || y2 || y3 || y4;

/* define the module for the mean of the whole vector alpha beta y1-y4 */
/* theta=alpha0,beta0,d11,d12,d22,sigma^-2,gamma */

start meanfunc(theta) global(time,x,p,n);
    alpha0=theta[1];
    beta0=theta[2];
    d11=theta[3];
    d12=theta[4];
    d22=theta[5];
    sigma2=theta[6];
    gamma1=theta[7:(6+p)];

    coeff2=1 || 0 || 1 || 1 || 1 || 1;

    prod=alpha0*coeff2 + beta0 * coeff;,
    stem=prod;
    do i=2 to n;
        stem=stem//prod;
    end;

    temp2=x*gamma1; /*n*1 vector*/
    fterm=j(n,1,0) || j(n,1,0) || temp2 || temp2 ||temp2 ||temp2 ;

    mean=fterm+stem;
    return(mean);
finish meanfunc;

start varfunc(theta) global(time, n,p);
    alpha0=theta[1];
beta0=theta[2];
d11=theta[3];
d12=theta[4];
d22=theta[5];
sigma2=theta[6];
gamma=theta[7:(6+p)];

coeff=(I(2) || j(2,4,0)) // (j(4,1,1) || time[1:4] || I(4));
  variance=j(6,6,0);
variance[1,1]=d11;
variance[1,2]=d12;
variance[2,1]=d12;
variance[2,2]=d22;
  variance[3:6,3:6]=sigma2*I(4);

var=coeff*variance*coeff';
return(var);
finish varfunc;

start hfunc(theta) global(time,x,p,n,y1,y2,y3,y4,y5,c1,c2,c3,
c4,rindex2,rindex3,rindex4,y12,y123,y1234);
  alpha0=theta[1];
  beta0=theta[2];
  d11=theta[3];
  d12=theta[4];
  d22=theta[5];
  sigma2=theta[6];
  gamma=theta[7:(6+p)];
  mean=meanfunc(theta); /* n x 6 matrix: mean of alpha beta y1-y4 */
  var=varfunc(theta); /* 6 x 6 matrix */
  fixed=x*gamma;
  h1=j(n,1,..);
  h2=j(n,1,..);
    h3=j(n,1,..);
  h4=j(n,1,..);

  h1=mean[,1] +(y1-mean[,5])*inv(var[5,5])*var[1,5]'
  /* conditional mean of y5 given y1 */
  temp1=mean[,1:2] + (y1-mean[,3])*inv(var[3,3])*var[3,1:2];
    h1=fixed+ temp1*(1//time[5]); /*conditional mean of y5 given y1*/
  temp2=j(n,2,..);
    temp2[rindex2,]=mean[rindex2,1:2] + (y12[rindex2,]-
mean[rindex2,3:4])*inv(var[3:4,3:4])*var[3:4,1:2];
h2[rindex2,]=(fixed[rindex2,] + temp2[rindex2,]*(1//time[5]);
temp3=j(n,2,..);
    temp3[rindex3,]=mean[rindex3,1:2] + (y123[rindex3,]-$\overline{y_{123}}[rindex3,]*inv(var[3:5,3:5])*var[3:5,1:2];
h3[rindex3,]=fixed[rindex3,] + temp3[rindex3,]*(1//time[5]);

    temp4=j(n,2,..);
    temp4[rindex4,]=mean[rindex4,1:2] + (y1234[rindex4,]-$\overline{y_{1234}}[rindex4,]*inv(var[3:6,3:6])*var[3:6,1:2];
h4[rindex4,]=fixed[rindex4,] + temp4[rindex4,]*(1//time[5]);

h=h1 || h2 || h3 || h4;
* print h;
    return(h);
finish hfunc;

start htilda(theta) global (time,x,p,tp,p1,p2,p3,p4,n,y1,y2,y3,y4,y5,x2, lambda1, lambda2, lambda3, lambda4, k1,k2,k3,k4,a,c1,c2,c3,c4, x_lam1, x_lam2, x_lam3, x_lam4, rindex2, rindex3, rindex4, y12, y123, y1234);
    alpha0=theta[1];
    beta0=theta[2];
    d11=theta[3];
    d12=theta[4];
    d22=theta[5];
    sigma2=theta[6];
    gamma=theta[7:(6+p)];
    c=theta[(7+p):tp]; /*c is the theta on page 21: lambda1 lambda2 lambda3 lambda4*/
/* define the design matrix for lambda1, lambda2, lambda3, lambda4 */
    delta1=c[1:p1];
    delta2=c[(p1+1):(p1+p2)];
    delta3=c[(p1+p2+1):(p1+p2+p3)];
    delta4=c[(p1+p2+p3+1):(p1+p2+p3+p4)];

h=hfunc(theta[1:(6+p)]);
    h1=h[,1];
    h2=h[,2];
    h3=h[,3];
    h4=h[,4];

h1_tilda=j(n,1,..);
    h2_tilda=j(n,1,..);
    h3_tilda=j(n,1,..);
    h4_tilda=j(n,1,..);

h1_tilda=h1-(1-lambda1)#(x_lam1*delta1);
h2_tilda[rindex2,]=h2[rindex2,]-(1-lambda2[rindex2,])
  #k1[rindex2,]#(x_lam2[rindex2,]*delta2);
  h3_tilda[rindex3,]=h3[rindex3,]-(1-lambda3[rindex3,])
  #k2[rindex3,]#(x_lam3[rindex3,]*delta3);
  h4_tilda[rindex4,]=h4[rindex4,]-(1-lambda4[rindex4,])
  #k3[rindex4,]#(x_lam4[rindex4,]*delta4);

  h_tilda=h1_tilda || h2_tilda || h3_tilda || h4_tilda;
  * print htilda;
  return(h_tilda);
  finish htilda;

  /* using numerical derivatives */

  start hder(theta) global(time,x,tp1,p2,p3,p4,n,y1,y2,y3, y4,y5,x,lambda1,
    lambda2,lambda3,lambda4,k1,k2,k3,k4,a,c1,c2,c3,c4,
    x_lam1,x_lam2,x_lam3,x_lam4, rindex2,rindex3, rindex4,y12,y123,y1234);
  alpha0=theta[1];
  beta0=theta[2];
  d11=theta[3];
  d12=theta[4];
  d22=theta[5];
  sigma2=theta[6];
  gamma=theta[7:(6+p)];
  c=theta[(7+p):tp]; /*c is the theta on page 21: lambda1 lambda2 lambda3 lambda4*/

  theta2=theta[1:(6+p)];
  h=hfunc(theta2);
  h1=h[,1];
  h2=h[,2];
  h3=h[,3];
  h4=h[,4];

  mat=I(6+p);
  eps=0.000001;
  h1_der=j(n,tp,0); h2_der=j(n,tp,.); h3_der=j(n,tp,.); h4_der=j(n,tp,.);

  do i=1 to (6+p);
    theta_temp=theta2+eps*mat[,i]';
    hplus=hfunc(theta_temp);
    h1_der[,i]=(hplus[,1]-h1)/eps;
  end;
\[ h_{2,\text{der}}[i] = (h_{\text{plus},2} - h_2)/\varepsilon; \]
\[ h_{3,\text{der}}[i] = (h_{\text{plus},3} - h_3)/\varepsilon; \]
\[ h_{4,\text{der}}[i] = (h_{\text{plus},4} - h_4)/\varepsilon; \]

\[ \text{end;} \]

\[ h_{1,\text{der}}[(7+p):(6+p+p1)] = -(1-\lambda_1) \# x_{\text{lam}1}; \]
\[ h_{2,\text{der}}[\text{rindex2},(7+p):tp] = 0; \]
\[ h_{3,\text{der}}[\text{rindex3},(7+p):tp] = 0; \]
\[ h_{4,\text{der}}[\text{rindex4},(7+p):tp] = 0; \]

\[ h_{2,\text{der}}[\text{rindex2},(7+p+p1):(6+p+p1+p2)] = -(1-\lambda_{2[\text{rindex2},\})} \# k_{1[\text{rindex2},\}] \# x_{\text{lam2}[\text{rindex2},\]}; \]
\[ h_{3,\text{der}}[\text{rindex3},(7+p+p1+p2):(6+p+p1+p2+p3)] = -(1-\lambda_{3[\text{rindex3},\})} \# k_{2[\text{rindex3},\}] \# x_{\text{lam3}[\text{rindex3},\]}; \]
\[ h_{4,\text{der}}[\text{rindex4},((7+p+p1+p2+p3)):tp] = -(1-\lambda_{4[\text{rindex4},\})} \# k_{3[\text{rindex4},\}] \# x_{\text{lam4}[\text{rindex4},\]}; \]

\[ \text{h\_der}=h_{1,\text{der}}//h_{2,\text{der}}//h_{3,\text{der}}//h_{4,\text{der}}; \]
\* \text{print h\_der;}
\* \text{return(h\_der);}
\text{finish h\_der;}

\text{start objfunc(theta) global(time, x, p, tp, p1, p2, p3, p4, n, y1, y2, y3, y4, y5, x2, lambda1, lambda2, lambda3, lambda4, k1, k2, k3, k4, a, c1, c2, c3, c4, x\_lam1, x\_lam2, x\_lam3, x\_lam4, rindex2, rindex3, rindex4, y12, y123, y1234);}
\text{alpha0=theta[1];}
\text{beta0=theta[2];}
\text{d11=theta[3];}
\text{d12=theta[4];}
\text{d22=theta[5];}
\text{sigma2=theta[6];}
\text{gamma=theta[7:(6+p)];}
\text{c=theta[(7+p):tp]; /* is the theta on page 21: lambda1 lambda2 lambda3 lambda4*/}
\text{htilda=htilda(theta);}
\text{h1\_tilde=htilda[1];}
\text{h2\_tilde=htilda[2];}
\text{h3\_tilde=htilda[3];}
\text{h4\_tilde=htilda[4];}
\text{hder=hder(theta);}
\text{h1\_der=hder[1:n,];}
\text{h2\_der=hder[(n+1):(2*n),];}
\text{h3\_der=hder[(2*n+1):(3*n),];}
\text{h4\_der=hder[(3*n+1):(4*n),];}
\text{object=j(tp,1,0);}

25
do i=1 to n;
    if c1[i]=0 then
do;
    temp1=lambda1[i]*h1_der[i,]’/k1[i];
    q1=-temp1/k1[i];
    object=object+(1-c1[i])*q1*(h2_tilda[i]-h1_tilda[i]);
end;

    if c2[i]=0 then
do;
    temp2=temp1+lambda2[i]*h2_der[i,]’/k2[i];
    q2=-temp2/k2[i];
    object=object+(1-c2[i])*q2*(h3_tilda[i]-h2_tilda[i]);
end;

    if c3[i]=0 then
do;
    temp3=temp2+lambda3[i]*h3_der[i,]’/k3[i];
    q3=-temp3/k3[i];
    object=object+(1-c3[i])*q3*(h4_tilda[i]-h3_tilda[i]);
end;

    if c4[i]=0 then
do;
    temp4=temp3+lambda4[i]*h4_der[i,]’/k4[i];
    q4=-temp4/k4[i];
    object=object+(1-c4[i])*q4*(y5[i]-h4_tilda[i]);
end;
end;
obj=object’*object;
obj=log(obj);
return(obj);
finish objfunc;

start objfunc2(theta) global(time,x,p,tp,p1,p2,p3,p4,n,y1,y2,y3,y4,y5,x2,
lambda1,lambda2,lambda3,lambda4,k1,k2,k3,k4,a,c1,c2,c3,c4,
x_lam1,x_lam2,x_lam3,x_lam4, rindex2,rindex3,rindex4,y12,y123,y1234);
    alpha0=theta[1];
    beta0=theta[2];
    d11=theta[3];
    d12=theta[4];
    d22=theta[5];
    sigma2=theta[6];
    gamma=theta[7:(6+p)];
    c=theta[(7+p):tp]; /*c is the theta on page 21: lambda1 lambda2 lambda3 lambda4*/
htilda=htilda(theta);
h1_tilda=htilda[,1];
h2_tilda=htilda[,2];
h3_tilda=htilda[,3];
h4_tilda=htilda[,4];

hder=hder(theta);
h1_der=hder[1:n,];
h2_der=hder[(n+1):(2*n),];
h3_der=hder[(2*n+1):(3*n),];
h4_der=hder[(3*n+1):(4*n),];

object=j(tp,1,0);
do i=1 to n;
   if c1[i]=0 then do;
   temp1=lambda1[i]*h1_der[i]'/k1[i];
   q1=-temp1/k1[i];
   object=object+(1-c1[i])q1*(h2_tilda[i]-h1_tilda[i]);
   end;
   if c2[i]=0 then do;
   temp2=temp1+lambda2[i]*h2_der[i]'/k2[i];
   q2=-temp2/k2[i];
   object=object+(1-c2[i])q2*(h3_tilda[i]-h2_tilda[i]);
   end;
   if c3[i]=0 then do;
   temp3=temp2+lambda3[i]*h3_der[i]'/k3[i];
   q3=-temp3/k3[i];
   object=object+(1-c3[i])q3*(h4_tilda[i]-h3_tilda[i]);
   end;
   if c4[i]=0 then do;
   temp4=temp3+lambda4[i]*h4_der[i]'/k4[i];
   q4=-temp4/k4[i];
   object=object+(1-c4[i])q4*(y5[i]-h4_tilda[i]);
   end;
* print object;
end;

return(object);
finish objfunc2;

start obji(theta) global(s,time,x,p,tp,p1,p2,p3,p4,n,y1,y2,
y3,y4,y5,x2,lambda1,
lambda2, lambda3, lambda4, k1, k2, k3, k4, a, c1, c2, c3, c4,
x_lam1, x_lam2, x_lam3, x_lam4, rindex2, rindex3, rindex4, y12,
y123, y1234, h1_tilda,
h2_tilda, h3_tilda, h4_tilda, h1_der, h2_der, h3_der, h4_der);
alpha0=theta[1];
beta0=theta[2];
d11=theta[3];
d12=theta[4];
d22=theta[5];
sigma2=theta[6];
gamma=theta[7:(6+p)];
c=theta[(7+p):tp]; /*c is the theta on page 21: lambda1
lambda2 lambda3 lambda4*/
i=s;
object=j(tp,1,0);
if c1[i]=0 then
  do;
    temp1=lambda1[i]*h1_der[i,j]/k1[i];
    q1=-temp1/k1[i];
    object=object+(1-c1[i])*q1*(h2_tilda[i]-h1_tilda[i]);
  end;
if c2[i]=0 then
  do;
    temp2=temp1+lambda2[i]*h2_der[i,j]/k2[i];
    q2=-temp2/k2[i];
    object=object+(1-c2[i])*q2*(h3_tilda[i]-h2_tilda[i]);
  end;
if c3[i]=0 then
  do;
    temp3=temp2+lambda3[i]*h3_der[i,j]/k3[i];
    q3=-temp3/k3[i];
    object=object+(1-c3[i])*q3*(h4_tilda[i]-h3_tilda[i]);
  end;
if c4[i]=0 then
  do;
    temp4=temp3+lambda4[i]*h4_der[i,j]/k4[i];
    q4=-temp4/k4[i];
    object=object+(1-c4[i])*q4*(y5[i]-h4_tilda[i]);
  end;
  * print object;
return(object);
finish obji;
x0=j(tp,1,0.12);
x0[1:(6+p)]=[ 0.5309, -0.00887, 0.03424, 0.004190, 0.001460, 0.03023, 0.03508, 0.2306, -0.09488, -0.07592];

print x0;
nrow=nrow(x0);
print nrow;

      optn2=J(1,2,.);  optn2[1]=0;  optn2[2]=1;
tc=J(1,10,.);
tc[1]=200;
*  tc[2]=500;
tc[2]=1000;

call nlpqn(rc,xres,"objfunc",x0,optn2) tc=tc;

/* calculate the estimated mean at the last time point */
/* C: coarsening variable */
C=J(n,1,1);
do i = 1 to n;
   if c1[i]=0 then c[i]=2;
   if c2[i]=0 then c[i]=3;
   if c3[i]=0 then c[i]=4;
   if c4[i]=0 then c[i]=5;
end;

h=hfnc(xres[1:(6+p)']);
h1=h[,1]; h2=h[,2]; h3=h[,3]; h4=h[,4];

mkh=0;
mk=0;
do i=1 to n;
   temp1=((c[i]=1)-lambda1[i]*(c[i]>=1))/k1[i];
   temp2=((c[i]=2)-lambda2[i]*(c[i]>=2))/k2[i];
   temp3=((c[i]=3)-lambda3[i]*(c[i]>=3))/k3[i];
   temp4=((c[i]=4)-lambda4[i]*(c[i]>=4))/k4[i];
   mkh=mkh+temp1*h1[i]+(temp2*h2[i])*(c[i]>=2)+(temp3*h3[i])*(c[i]>=3)+(temp4*h4[i])*(c[i]>=4);
   mk=mk+temp1*(c[i]>=1)+temp2*(c[i]>=2)+temp3*(c[i]>=3)+temp4*(c[i]>=4);
end;

numfirst=sum(y5/k4);
testipw=numfirst/n;
print testipw;
denfirst=sum((c=5)/k4);
print denfirst;

print mkh mk;
\[
\mu = \frac{\text{numfirst} + \text{mkh}}{\text{denfirst} + \text{mk}};
\]
print \mu;

quilt;