Supplement: Bayesian Nonparametric Policy Search with Application to Periodontal Recall Intervals

1 Priors and MCMC details

In this section, we describe our choice of priors and MCMC details for both simulation study and the tailored model for HP data analysis.

1.1 Simulation study

Recall, the model described in Section 4 is

$$
\begin{aligned}
(\mathbf{X}_i^T, Y_{i0})^T|\Theta_i &\sim \text{Normal}(\boldsymbol{\theta}_{i0}, \boldsymbol{\Sigma}_0) \\
\log(\delta_{it})|\Theta_i, Y_{it-1}, A_{it} &\sim \text{Normal}(\mathcal{X}_{it}^T\boldsymbol{\theta}_{i1}, \sigma_1^2) \\
Y_{it}|\Theta_i, Y_{it-1}, A_{it}, \delta_{it} &\sim \text{Normal}(\mathcal{Z}_{it}^T\boldsymbol{\theta}_{i2}, \sigma_2^2)
\end{aligned}
$$

where $\mathcal{X}_{it} = [\mathbf{X}_i^T, Y_{it-1}, \log(A_{it}), \mathbf{X}_i^T \log(A_{it}), Y_{it-1}\log(A_{it})]^T$ and $\mathcal{Z}_{it} = (\mathbf{X}_i^T, Y_{it-1}, \delta_{it}, \mathbf{X}_i^T\delta_{it}, Y_{it-1}\delta_{it})^T$. Let g be the random effects density, such that $\Theta_i \overset{iid}{\sim} g(\Theta)$, and we assume g follows a Dirichlet Process prior $g \sim DP(\alpha_0, P_0)$, which can be written as $g(\Theta) = \sum_{l=1}^{\infty} \omega_l \mathbb{1}_{\Delta_l}(\Theta)$, where $\omega_l = V_l \prod_{h<l}(1 - V_h)$, $V_l \sim \text{Beta}(1, \alpha_0)$, $\Delta_l = (\boldsymbol{\theta}_{0l}^{\star T}, \boldsymbol{\theta}_{1l}^{\star T}, \boldsymbol{\theta}_{2l}^{\star T})^T \sim \text{Normal}(\boldsymbol{m}_b, \boldsymbol{\Sigma}_b)$, and $\mathbb{1}_{\Delta_l}(\cdot)$ is the indicator function with a point mass at Δ_l.. The covariance matrix $\boldsymbol{\Sigma}_b$ is taken to

be block diagonal with $\text{Cov}(\boldsymbol{\theta}^\star_{jl}) = \boldsymbol{\Sigma}_{bj}$ and $\text{Cov}(\boldsymbol{\theta}^\star_{jl}, \boldsymbol{\theta}^\star_{kl}) = 0$. For priors, we select $\boldsymbol{m}_b = (\boldsymbol{m}^T_{b0}, \boldsymbol{m}^T_{b1}, \boldsymbol{m}^T_{b2})^T \sim \text{Normal}(0, I_{p0+p1+p2})$ and $\boldsymbol{\Sigma}^{-1}_{bj} \sim \text{Wishart}(p_j + 1, 1/(p_j + 1)I_{p_j})$, where p_j is the dimension of $\boldsymbol{\theta}^\star_{jl}$, $j = 0, 1, 2$. We assume conjugate prior $\Sigma^{-1}_0 \sim \text{Wishart}(p_0+1, 1/(p_0+1)\Psi_0)$, $\sigma^{-2}_1 \sim \text{Gamma}(\epsilon_1, \epsilon_2)$, $\sigma^{-2}_2 \sim \text{Gamma}(\epsilon_3, \epsilon_4)$. In the simulation, we assume $\Psi_0 = I_{p0}$, $\epsilon_1 = \epsilon_2 = \epsilon_3 = \epsilon_4 = 1$ to make the priors non-informative.

For MCMC, we introduce a latent indicator variable γ_i such that $(\Theta_i|\gamma_i = l) = \Delta_l = (\boldsymbol{\theta}^{\star T}_{0l}, \boldsymbol{\theta}^{\star T}_{1l}, \boldsymbol{\theta}^{\star T}_{2l})^T$ and $Pr(\gamma_i = l) = \omega_l$. We let $l = 1, \ldots, L$ be unique values of γ_i of n realizations $\{\gamma_i, i = 1, \ldots, n\}$. Let $s_l = \sum_{i=1}^n I(\gamma_i = l)$ and N_i be the total number of visits for subject i. We sample from the posterior using Gibbs sampling. The algorithm begins with setting all parameters to initial values and drawing the parameters from their full conditional distributions in sequence. The full conditional distributions are given below.

$$\boldsymbol{\theta}^\star_{0l}|\text{rest} \sim \text{Normal}\Big((s_l\boldsymbol{\Sigma}^{-1}_0 + \boldsymbol{\Sigma}^{-1}_{b0})^{-1}\Big(\boldsymbol{\Sigma}^{-1}_0 \sum_{i:\gamma_i=l} \boldsymbol{X}_i + \boldsymbol{\Sigma}^{-1}_{b0}\boldsymbol{m}_{b0}\Big), (s_l\boldsymbol{\Sigma}^{-1}_0 + \boldsymbol{\Sigma}^{-1}_{b0})^{-1}\Big)$$

$$\boldsymbol{\Sigma}^{-1}_0|\text{rest} \sim \text{Wishart}\Big(n + p_0 + 1, \Big\{ \sum_{i=1}^n \big[(\boldsymbol{X}^T_i, Y_{i0}) - \boldsymbol{\theta}_{i0}\big]\big[(\boldsymbol{X}^T_i, Y_{i0}) - \boldsymbol{\theta}_{i0}\big]^T + (p_0 + 1)\Psi^{-1}_0\Big\}^{-1}\Big)$$

$$\boldsymbol{\theta}^\star_{1l}|\text{rest} \sim \text{Normal}\Bigg(\Big(\sigma^{-2}_1 \sum_{i:\gamma_i=l} \sum_{t=1}^{N_i} \mathcal{X}_{it}\mathcal{X}^T_{it} + \boldsymbol{\Sigma}^{-1}_{b1}\Big)^{-1}\Big[\sigma^{-2}_1 \sum_{i:\gamma_i=l} \sum_{t=1}^{N_i} \mathcal{X}_{it}\log(\delta_{it}) + \boldsymbol{\Sigma}^{-1}_{b1}\boldsymbol{m}_{b1}\Big], \Big(\sigma^{-2}_1 \sum_{i:\gamma_i=l} \sum_{t=1}^{N_i} \mathcal{X}_{it}\mathcal{X}^T_{it} + \boldsymbol{\Sigma}^{-1}_{b1}\Big)^{-1}\Bigg)$$

$$\sigma^{-2}_1|\text{rest} \sim \text{Gamma}\Big(\frac{1}{2}\sum_{i=1}^n N_i + \epsilon_1, \frac{1}{2}\sum_{i=1}^n \sum_{t=1}^{N_i} \Big\{[\log(\delta_{it})]^2 - 2\log(\delta_{it})\mathcal{X}^T_{it}\boldsymbol{\theta}_{i1} + \boldsymbol{\theta}^T_{i1}\mathcal{X}_{it}\mathcal{X}^T_{it}\boldsymbol{\theta}_{i1}\Big\} + \epsilon_2\Big)$$

$$\boldsymbol{\theta}_{2l}^{\star}|\text{rest} \sim \text{Normal}\Bigg(\Big(\sigma_2^{-2}\sum_{i:\gamma_i=l}\sum_{t=1}^{N_i}\mathcal{Z}_{it}\mathcal{Z}_{it}^T+\boldsymbol{\Sigma}_{b2}^{-1}\Big)^{-1}\Big(\sigma_2^{-2}\sum_{i:\gamma_i=l}\sum_{t=1}^{N_i}\mathcal{Z}_{it}Y_{it}+\boldsymbol{\Sigma}_{b2}^{-1}\boldsymbol{m}_{b2}\Big),$$
$$\Big(\sigma_2^{-2}\sum_{i:\gamma_i=l}\sum_{t=1}^{N_i}\mathcal{Z}_{it}\mathcal{Z}_{it}^T+\boldsymbol{\Sigma}_{b2}^{-1}\Big)^{-1}\Bigg)$$

$$\sigma_2^{-2}|\text{rest} \sim \text{Gamma}\Bigg(\frac{1}{2}\sum_{i=1}^{n}N_i+\epsilon_3,\frac{1}{2}\sum_{i=1}^{n}\sum_{t=1}^{N_i}\Big(Y_{it}^2-2Y_{it}\mathcal{Z}_{it}^T\boldsymbol{\theta}_{i2}+\boldsymbol{\theta}_{i2}^T\mathcal{Z}_{it}\mathcal{Z}_{it}^T\boldsymbol{\theta}_{i2}\Big)+\epsilon_4\Bigg)$$

$$\begin{aligned}\Pr(\gamma_i=l)|\text{rest} \propto \omega_l \times \exp\Big\{-\frac{1}{2}\Big[\boldsymbol{\theta}_{0l}^{\star T}\boldsymbol{\Sigma}_0^{-1}\boldsymbol{\theta}_{0l}^{\star}-2(\boldsymbol{X}_i^T,Y_{i0})\boldsymbol{\Sigma}_0^{-1}\boldsymbol{\theta}_{0l}^{\star}\Big]\Big\}\\ \times \exp\Big\{-\frac{1}{2}\sum_{t=1}^{N_i}\Big[\sigma_1^{-2}\boldsymbol{\theta}_{1l}^{\star T}\mathcal{X}_{it}\mathcal{X}_{it}^T\boldsymbol{\theta}_{1l}^{\star}-2\log(\delta_{it})\sigma_1^{-2}\mathcal{X}_{it}^T\boldsymbol{\theta}_{1l}^{\star}\Big]\Big\}\\ \times \exp\Big\{-\frac{1}{2}\sum_{t=1}^{N_i}\Big(\sigma_2^{-2}\boldsymbol{\theta}_{2l}^{\star T}\mathcal{Z}_{it}\mathcal{Z}_{it}^T\boldsymbol{\theta}_{2l}^{\star}-2Y_{it}\sigma_2^{-2}\mathcal{Z}_{it}^T\boldsymbol{\theta}_{2l}^{\star}\Big)\Big\}\end{aligned}$$

$$V_l|\text{rest} \sim \text{Beta}\Bigg(1+s_l,\alpha_0+\sum_{h>l}n_h\Bigg)$$

,

$$\omega_l = V_l\prod_{h<l}(1-V_h)$$

For $j=0,1,2$,

$$\boldsymbol{m}_{bj}|\text{rest} \sim \text{Normal}\Bigg(\Big(L\boldsymbol{\Sigma}_{bj}^{-1}+I_{p_j}\Big)^{-1}\Big(\boldsymbol{\Sigma}_{bj}\sum_{l=1}^{L}\boldsymbol{\theta}_{jl}^{\star}\Big),\Big(L\boldsymbol{\Sigma}_{bj}^{-1}+I_{p_j}\Big)^{-1}\Bigg)$$

$$\boldsymbol{\Sigma}_{bj}^{-1}|\text{rest} \sim \text{Wishart}\Bigg(L+p_j+1,\Big[\sum_{l=1}^{L}(\boldsymbol{\theta}_{jl}^{\star}-\boldsymbol{m}_{bj})(\boldsymbol{\theta}_{jl}^{\star}-\boldsymbol{m}_{bj})^T+(p_j+1)I_{p_j}\Big]^{-1}\Bigg)$$

1.2 HP Data Analysis

The final DPM model we use for the data analysis described in Section 7 is:

$$
\begin{aligned}
(\mathbf{X}_i^{\star T}, Y_{i0}^{\star})^T|\Theta_i &\sim \text{Normal}(\boldsymbol{\theta}_{i0}, \boldsymbol{\Sigma}_0) \\
\log(\delta_{it})|\Theta_i, \mathbf{X}_i, Y_{it-1}, A_{it} &\sim \text{Normal}(\mathcal{X}_{it}^T \boldsymbol{\theta}_{i1}, \lambda_{it1}\sigma_1^2) \\
Y_{it}^{\star}|\Theta_i, \mathbf{X}_i, Y_{it-1}, \delta_{it} &\sim \text{Normal}(\mathcal{Z}_{it}^T \boldsymbol{\theta}_{i2}, \lambda_{it2}\sigma_2^2)
\end{aligned}
$$

where $\lambda_{it1} \sim$ Inv-Gamma$(\nu_1/2, \nu_1/2)$, $\lambda_{it2} \sim$ Inv-Gamma$(\nu_2/2, \nu_2/2)$; $X_{ij} = I(X_{ij}^{\star} > 0)$ for $j = 1, 2, 4, 5, 6$. The 1st, 2nd 4th, 5th, 6th diagonal elements in $\boldsymbol{\Sigma}_0$ are forced to 1. $\tilde{Y}_{it} = Y_{it}^{\star}$ if $0 \leq Y_{it}^{\star} \leq 1$, $\tilde{Y}_{it} = 0$ if $Y_{it}^{\star} < 0$ and, $\tilde{Y}_{it} = 1$ if $Y_{it}^{\star} > 1$. We introduce an indicator ξ_{it}, such that

$$
Y_{it}|(Y_{it-1} = y_{it-1}) = \begin{cases} y_{it-1} & \text{if } \xi_{it} = 0 \\ \tilde{Y}_{it} & \text{if } \xi_{it} = 1 \end{cases}
$$

The probability of $\xi_{it} = 1$ depends on all the covariates $\mathcal{Z}_{it}$. We use the probit regression to model the binary latent variable ξ_{it},

$$
p_{it} = \Pr(\xi_{it} = 1) = \Phi(\mathcal{Z}_{it}^T \boldsymbol{\theta}_{i3})
$$

where Φ is the standard normal CDF. We use another auxiliary variable $\xi_{it}^{\star} \sim$ Normal$(\mathcal{Z}_{it}^T \boldsymbol{\theta}_{i3}, 1)$ in order to get conjugate updates. $\xi_{it}^{\star}$ relates to the ξ_{it} via $\xi_{it} = \mathbb{1}(\xi_{it}^{\star} > 0)$.

In the tailored model for HP data, $\Theta_i = \{\boldsymbol{\theta}_{i0}, \boldsymbol{\theta}_{i1}, \boldsymbol{\theta}_{i2}, \boldsymbol{\theta}_{i3}\}$ and we give Dirichlet process prior to the density of Θ_i in a similar way. The hyperprameters Ψ_0, ϵ_1, ϵ_2, ϵ_3, ϵ_4 are chosen via empirical Bayes technique, based on some preliminary fits. Ψ_0 are assumed to be a diagonal matrix with in-

verse of sample variance of Y_{i0} as its last diagonal element and 1 to be the other diagonal elements. We obtain the empirical estimates of σ_1^2 and σ_2^2 via linear regression fitting and tune ϵ_1, ϵ_2, ϵ_3, ϵ_4 so that the prior mean of σ_j^2 is equal to the empirical estimate and prior variance is equal to 10 times the empirical estimate for $j = 1, 2$. We assume $\nu_j \sim \text{DiscreteUniform}(\{1, 1.01, 1.02, \ldots, 50\})$ for $j = 1, 2$.

Let rtNormal$(\cdot)$ be the truncated normal distribution. The full conditional distributions in MCMC are given below:

$$X_{ij}^{\star}|\text{rest} \sim \begin{cases} \text{rtNormal}\big(\boldsymbol{\theta}_{i0,j} + \boldsymbol{\Sigma}_{0j,-j}\boldsymbol{\Sigma}_{0-j,-j}^{-1}(\mathbf{X}_{i,-j} - \boldsymbol{\theta}_{i0,-j}), \boldsymbol{\Sigma}_{0j,j} - \boldsymbol{\Sigma}_{0j,-j}\boldsymbol{\Sigma}_{0-j,-j}^{-1}\boldsymbol{\Sigma}_{0-j,j}, -\infty, 0\big) & \text{if } X_{ij} = 0 \\ \text{rtNormal}\big(\boldsymbol{\theta}_{i0,j} + \boldsymbol{\Sigma}_{0j,-j}\boldsymbol{\Sigma}_{0-j,-j}^{-1}(\mathbf{X}_{i,-j} - \boldsymbol{\theta}_{i0,-j}), \boldsymbol{\Sigma}_{0j,j} - \boldsymbol{\Sigma}_{0j,-j}\boldsymbol{\Sigma}_{0-j,-j}^{-1}\boldsymbol{\Sigma}_{0-j,j}, 0, +\infty\big) & \text{if } X_{ij} = 1 \end{cases}$$

where $\boldsymbol{\Sigma}_{0j,-j}$ is the jth row of $\boldsymbol{\Sigma}_0$ without jth column, $\boldsymbol{\Sigma}_{0-j,j}$ is the jth column of $\boldsymbol{\Sigma}_0$ without jth row, $\boldsymbol{\Sigma}_{0-j,-j}$ is $\boldsymbol{\Sigma}_0$ without the jth row and and jth column, $\boldsymbol{\Sigma}_{0j,j}$ is the jth row and jth column element in $\boldsymbol{\Sigma}_0$; $\boldsymbol{\theta}_{i0,j}$ is the jth element in $\boldsymbol{\theta}_{i0}$, $\boldsymbol{\theta}_{i0,-j}$ is $\boldsymbol{\theta}_{i0}$ without the jth element.

$$Y_{i0}^{\star}|\text{rest} \sim \begin{cases} \text{rtNormal}\big(\boldsymbol{\theta}_{i0,j} + \boldsymbol{\Sigma}_{0j,-j}\boldsymbol{\Sigma}_{0-j,-j}^{-1}(\mathbf{X}_{i} - \boldsymbol{\theta}_{i0,-j}), \boldsymbol{\Sigma}_{0j,j} - \boldsymbol{\Sigma}_{0j,-j}\boldsymbol{\Sigma}_{0-j,-j}^{-1}\boldsymbol{\Sigma}_{0-j,j}, -\infty, 0\big) & \text{if } Y_{i0} = 0 \\ Y_{i0} & \text{if } 0 < Y_{i0} < 1 \\ \text{rtNormal}\big(\boldsymbol{\theta}_{i0,j} + \boldsymbol{\Sigma}_{0j,-j}\boldsymbol{\Sigma}_{0-j,-j}^{-1}(\mathbf{X}_{i} - \boldsymbol{\theta}_{i0,-j}), \boldsymbol{\Sigma}_{0j,j} - \boldsymbol{\Sigma}_{0j,-j}\boldsymbol{\Sigma}_{0-j,-j}^{-1}\boldsymbol{\Sigma}_{0-j,j}, 0, +\infty\big) & \text{if } Y_{i0} = 1 \end{cases}$$

where $j = 7$.

$$\boldsymbol{\Sigma}_0^{\star -1}|\text{rest} \sim \text{Wishart}\left(n + p_0 + 1, \left\{\sum_{i=1}^{n}\left[(\boldsymbol{X}_i^T, Y_{i0}) - \boldsymbol{\theta}_{i0}\right]\left[(\boldsymbol{X}_i^T, Y_{i0}) - \boldsymbol{\theta}_{i0}\right]^T + (p_0+1)\Psi_0^{-1}\right\}^{-1}\right)$$

$$\boldsymbol{\Sigma}_{0j,\cdot} = \boldsymbol{\Sigma}^{\star}_{0j,\cdot}/\sqrt{\boldsymbol{\Sigma}^{\star}_{0j,j}}, \quad \boldsymbol{\Sigma}_{0\cdot,j} = \boldsymbol{\Sigma}^{\star}_{0\cdot,j}/\sqrt{\boldsymbol{\Sigma}^{\star}_{0j,j}} \quad \text{for } j = 1,2,4,5,6$$

The second step sets the variance of binary covariates to be 1.

$$\boldsymbol{\theta}^{\star\star}_{0l}|\text{rest} \sim \text{Normal}\Big((s_l\boldsymbol{\Sigma}^{\star -1}_0 + \boldsymbol{\Sigma}^{-1}_{b0})^{-1}\Big(\boldsymbol{\Sigma}^{\star -1}_0 \sum_{i:\gamma_i=l} \boldsymbol{X}_i + \boldsymbol{\Sigma}^{-1}_{b0}\boldsymbol{m}_{b0}\Big), (s_l\boldsymbol{\Sigma}^{\star -1}_0 + \boldsymbol{\Sigma}^{-1}_{b0})^{-1}\Big)$$

$$\boldsymbol{\theta}^{\star}_{0l,j} = \boldsymbol{\theta}^{\star\star}_{0l,j}/\sqrt{\boldsymbol{\Sigma}^{\star}_{0j,j}} \quad \text{for } j = 1,2,4,5,6$$

$$\boldsymbol{\theta}^{\star}_{1l}|\text{rest} \sim \text{Normal}\Bigg(\Big(\sigma_1^{-2}\sum_{i:\gamma_i=l}\sum_{t=1}^{N_i}\mathcal{X}_{it}\lambda^{-1}_{it1}\mathcal{X}^T_{it} + \boldsymbol{\Sigma}^{-1}_{b1}\Big)^{-1}\Big[\sigma_1^{-2}\sum_{i:\gamma_i=l}\sum_{t=1}^{N_i}\mathcal{X}_{it}\lambda^{-1}_{it1}\log(\delta_{it}) + \boldsymbol{\Sigma}^{-1}_{b1}\boldsymbol{m}_{b1}\Big], \Big(\sigma_1^{-2}\sum_{i:\gamma_i=l}\sum_{t=1}^{N_i}\mathcal{X}_{it}\lambda^{-1}_{it1}\mathcal{X}^T_{it} + \boldsymbol{\Sigma}^{-1}_{b1}\Big)^{-1}\Bigg)$$

$$\sigma_1^{-2}|\text{rest} \sim \text{Gamma}\Bigg(\frac{1}{2}\sum_{i=1}^{n}N_i + \epsilon_1, \frac{1}{2}\sum_{i=1}^{n}\sum_{t=1}^{N_i}\Big\{\lambda^{-1}_{it1}[\log(\delta_{it})]^2 - 2\log(\delta_{it})\lambda^{-1}_{it1}\mathcal{X}^T_{it}\boldsymbol{\theta}_{i1} + \boldsymbol{\theta}^T_{i1}\mathcal{X}_{it}\lambda^{-1}_{it1}\mathcal{X}^T_{it}\boldsymbol{\theta}_{i1}\Big\} + \epsilon_2\Bigg)$$

$$\lambda^{-1}_{it1}|\text{rest} \sim \text{Gamma}\Bigg(\frac{1}{2} + \frac{\nu_1}{2}, \frac{1}{\sigma_1^2}[\log(\delta_{it}) - \mathcal{X}_{it}\boldsymbol{\theta}_{i1}]^2 + \frac{\nu_1}{2}\Bigg)$$

$$\Pr(\nu_1 = v)|\text{rest} \propto \exp\Bigg\{\sum_{i=1}^{n}N_i\log\Big[\frac{(v/2)^{v/2}}{\Gamma(v/2)}\Big] + (v/2-1)\sum_{i=1}^{n}N_i - v/2\sum_{i=1}^{n}\sum_{t=1}^{N_i}\lambda^{-1}_{it1}\Bigg\}$$

$$\boldsymbol{\theta}^{\star}_{2l}|\text{rest} \sim \text{Normal}\Bigg(\Big(\sigma_2^{-2}\sum_{i:\gamma_i=l}\sum_{t:\xi_{it}=1}\mathcal{Z}_{it}\lambda^{-1}_{it2}\mathcal{Z}^T_{it} + \boldsymbol{\Sigma}^{-1}_{b2}\Big)^{-1}\Big(\sigma_2^{-2}\sum_{i:\gamma_i=l}\sum_{t:\xi_{it}=1}\mathcal{Z}_{it}\lambda^{-1}_{it2}Y^{\star}_{it} + \boldsymbol{\Sigma}^{-1}_{b2}\boldsymbol{m}_{b2}\Big), \Big(\sigma_2^{-2}\sum_{i:\gamma_i=l}\sum_{t:\xi_{it}=1}\mathcal{Z}_{it}\lambda^{-1}_{it2}\mathcal{Z}^T_{it} + \boldsymbol{\Sigma}^{-1}_{b2}\Big)^{-1}\Bigg)$$

$$\sigma_2^{-2}|\text{rest} \sim \text{Gamma}\Bigg(\frac{1}{2}\sum_{i=1}^{n}\sum_{t=1}^{N_i}\xi_{it} + \epsilon_3, \frac{1}{2}\sum_{i=1}^{n}\sum_{t:\xi_{it}=1}\Big(\lambda^{-1}_{it2}Y^{\star 2}_{it} - 2Y^{\star}_{it}\lambda^{-1}_{it2}\mathcal{Z}^T_{it}\boldsymbol{\theta}_{i2} + \boldsymbol{\theta}^T_{i2}\mathcal{Z}_{it}\lambda^{-1}_{it2}\mathcal{Z}^T_{it}\boldsymbol{\theta}_{i2}\Big) + \epsilon_4\Bigg)$$

$$Y_{it}^{\star}|\text{rest} \sim \text{Normal}(\mathcal{Z}_{it}^T \boldsymbol{\theta}_{i2}, \lambda_{it2}\sigma_2^2) \qquad \text{for } t : \xi_t = 0$$

$$\lambda_{it2}^{-1}|\text{rest} \sim \text{Gamma}\left(\frac{1}{2} + \frac{\nu_2}{2}, \frac{1}{\sigma_2^2}(Y_{it}^{\star} - \mathcal{Z}_{it}\boldsymbol{\theta}_{i2})^2 + \frac{\nu_2}{2}\right)$$

$$\Pr(\nu_2 = v)|\text{rest} \propto \exp\left\{ \sum_{i=1}^{n} N_i \log\left[\frac{(v/2)^{v/2}}{\Gamma(v/2)}\right] + (v/2-1)\sum_{i=1}^{n} N_i - v/2 \sum_{i=1}^{n}\sum_{t=1}^{N_i} \lambda_{it2}^{-1}\right\}$$

$$\boldsymbol{\theta}_{3l}^{\star}|\text{rest} \sim \text{Normal}\Bigg(\Big(\sum_{i:\gamma_i=l}\sum_{t=1}^{N_i} \mathcal{Z}_{it}\mathcal{Z}_{it}^T + \boldsymbol{\Sigma}_{b3}^{-1}\Big)^{-1}\Big(\sum_{i:\gamma_i=l}\sum_{t=1}^{N_i} \mathcal{Z}_{it}\xi_{it}^{\star} + \boldsymbol{\Sigma}_{b3}^{-1}\boldsymbol{m}_{b3}\Big), \Big(\sum_{i:\gamma_i=l}\sum_{t=1}^{N_i} \mathcal{Z}_{it}\mathcal{Z}_{it}^T + \boldsymbol{\Sigma}_{b3}^{-1}\Big)^{-1}\Bigg)$$

$$\xi_{it}|\text{rest} \sim \text{Binomial}\left(1, \frac{\Pr(y = Y_{it}|\mathcal{Z}_{it}^T\boldsymbol{\theta}_{i2}, \sigma_2^2\lambda_{it2}) * (1-p_{it})}{I(Y_{it} = Y_{it-1}) * p_{it} + \Pr(y = Y_{it}|\mathcal{Z}_{it}^T\boldsymbol{\theta}_{i2}, \sigma_2^2\lambda_{it2}) * (1-p_{it})}\right)$$

$$\text{where } p_{it} = \Phi(\mathcal{Z}_{it}^T\boldsymbol{\theta}_{i3}), \Pr(y = Y_{it}|\mathcal{Z}_{it}^T\boldsymbol{\theta}_{i2}, \sigma_2^2\lambda_{it1}) = \begin{cases} \Phi(\frac{-\mathcal{Z}_{it}^T\boldsymbol{\theta}_{i2}}{\sigma_2\sqrt{\lambda_{it2}}}) & \text{if } Y_{it} = 0 \\ 0 & \text{if } 0 < Y_{it} < 1 \\ 1 - \Phi(\frac{1-\mathcal{Z}_{it}^T\boldsymbol{\theta}_{i2}}{\sigma_2\sqrt{\lambda_{it2}}}) & \text{if } Y_{it} = 1 \end{cases}$$

$$\xi_{it}^{\star}|\text{rest} \sim \begin{cases} \text{rtNormal}(\mathcal{Z}_{it}^T\boldsymbol{\theta}_{i3}, 1, 0, \infty) & \text{if } \xi_{it} = 1 \\ \text{rtNormal}(\mathcal{Z}_{it}^T\boldsymbol{\theta}_{i3}, 1, -\infty, 0) & \text{if } \xi_{it} = 0 \end{cases}$$

$$\Pr(\gamma_i = l)|\text{rest} \propto \omega_l \times \exp\Big\{ -\frac{1}{2}\Big[\boldsymbol{\theta}_{0l}^{\star T}\boldsymbol{\Sigma}_0^{-1}\boldsymbol{\theta}_{0l}^{\star} - 2(\boldsymbol{X}_i^T, Y_{i0})\boldsymbol{\Sigma}_0^{-1}\boldsymbol{\theta}_{0l}^{\star}\Big]\Big\}$$
$$\times \exp\Big\{ -\frac{1}{2}\sum_{t=1}^{N_i}\Big[\sigma_1^{-2}\boldsymbol{\theta}_{1l}^{\star T}\mathcal{X}_{it}\lambda_{it1}^{-1}\mathcal{X}_{it}^T\boldsymbol{\theta}_{1l}^{\star} - 2\log(\delta_{it})\sigma_1^{-2}\lambda_{it1}^{-1}\mathcal{X}_{it}^T\boldsymbol{\theta}_{1l}^{\star}\Big]\Big\}$$
$$\times \exp\Big\{ -\frac{1}{2}\sum_{t=1}^{N_i}\Big(\sigma_2^{-2}\boldsymbol{\theta}_{2l}^{\star T}\mathcal{Z}_{it}\lambda_{it2}^{-1}\mathcal{Z}_{it}^T\boldsymbol{\theta}_{2l}^{\star} - 2Y_{it}^{\star}\sigma_2^{-2}\lambda_{it2}^{-1}\mathcal{Z}_{it}^T\boldsymbol{\theta}_{2l}^{\star}\Big)$$
$$\times \exp\Big[-\frac{1}{2}\sum_{t=1}^{N_i}\Big(\boldsymbol{\theta}_{3l}^{\star T}\mathcal{Z}_{it}\mathcal{Z}_{it}^T\boldsymbol{\theta}_{3l}^{\star} - 2\xi_{it}^{\star}\mathcal{Z}_{it}^T\boldsymbol{\theta}_{3l}^{\star}\Big)\Big]$$

$$V_l|\text{rest} \sim \text{Beta}\Big(1 + s_l, \alpha_0 + \sum_{h>l} n_h\Big)$$

,

$$\omega_l = V_l \prod_{h<l}(1 - V_h)$$

For $j = 0, 1, 2, 3$,

$$\boldsymbol{m}_{bj}|\text{rest} \sim \text{Normal}\Big(\Big(L\boldsymbol{\Sigma}_{bj}^{-1} + I_{p_j}\Big)^{-1}\Big(\boldsymbol{\Sigma}_{bj}\sum_{l=1}^{L}\boldsymbol{\theta}_{jl}^{\star}\Big), \Big(L\boldsymbol{\Sigma}_{bj}^{-1} + I_{p_j}\Big)^{-1}\Big)$$

$$\boldsymbol{\Sigma}_{bj}^{-1}|\text{rest} \sim \text{Wishart}\Big(L + p_j + 1, \Big[\sum_{l=1}^{L}(\boldsymbol{\theta}_{jl}^{\star} - \boldsymbol{m}_{bj})(\boldsymbol{\theta}_{jl}^{\star} - \boldsymbol{m}_{bj})^T + (p_j + 1)I_{p_j}\Big]^{-1}\Big)$$

Figure 1: Estimated optimal feature weights $\boldsymbol{\alpha}_{opt}$ for the simulation study with different sample sizes. The boxplots show the estimated $\boldsymbol{\alpha}_{opt}$ over the 100 simulated datasets for sample size $n = 500$ and sample size $n = 1000$; the solid points represent $\boldsymbol{\alpha}_{opt}$ for the Oracle policy. The risk score is a linear combination the two baseline covariates (X_1 and X_2), non-compliance ("Non-comp"; $\log(|\delta_{t-1} - A_{t-1}| + 1)$), and disease status ("Cur Y"; Y_{t-1}):

2 Additional simulations

2.1 Considering binary covariates, quadratic effects, and non-constant variances

In addition to the simulation with sample size $n = 1000$, we also conduct a simulation with a smaller sample size $n = 500$. We only use reduction utility function as illustration. Figure 1 shows the estimated $\boldsymbol{\alpha}_{opt}$ over the 100 simulated datasets with either sample size $n = 500$ or $n = 1000$ for each of the datasets. With smaller sample size $n = 500$, the method still works well and the estimated values have slightly larger variance.

In HP dataset, some of the baseline covariates are binary (gender, race, etc). To make the

simulation more similar to the HP analysis we have added a scenario ("Scenario 1") assuming X_2 is binary. We introduce latent continuous variable $X_i^\star$ to link the binary covariate to the DPM model, $X_{i2} = I(X_{i2}^\star > 0)$. Subjects with latent variable $X_{i2}^\star$ are generated from the two groups (6) and (7) with Prob$(G_i = 1) = 0.8$, and X_{i2} are generated using the link function. For this scenario, we only consider reduction utility function. Figure 2 ("Scenario 1") shows the sampling distribution of the estimated optimal feature weights $\boldsymbol{\alpha}$. The estimated optimal feature weights under the DPM model agree with those under the oracle model.

The DPM model is very flexible and so we anticipate that the results under DPM model should be fairly robust to model misspcification. To support this claim, we conduct two more simulation studies in which the true generative model is different from the DPM model we propose. First, we add a quadratic effect of δ_t (actual time between visits) on disease progression in the first group when generating data ("Scenario 2"):

$$
\begin{aligned}
&Y_{it}|Y_{it-1}, \delta_{it}, G_i = 1 \\
&\qquad \sim \text{Normal}\left[-2 + 0.2X_{i2} - 0.3(\delta_{it} - 6) + 0.9Y_{it-1} + 0.02(\delta_{it} - 6)Y_{it-1} + 0.06\delta_{it}^2, \sigma_2^2\right]
\end{aligned}
$$

but fit original proposed model which only considers the linear effect of δ_t (conditioned on the Dirichlet process group allocation). Figure 2 ("Scenario 2") shows that the estimated optimal feature weights under the proposed DPM model are still very close to the optimal feature weights estimated under the oracle model. Second, we assume non-constant variance in the disease status response ("Scenario 3"). We generate data assuming $Var(Y_t) = 0.2 * exp(1 + 0.2 * Y_{t-1})$, but fit the original proposed model with constant variance to the simulated datasets. Figure 2 ("Scenario

Figure 2: Estimated optimal feature weights $\boldsymbol{\alpha}_{opt}$ for the simulation study scenario 1, 2 and 3. The boxplots show the estimated $\boldsymbol{\alpha}_{opt}$ over the 100 simulated datasets; the solid points represent $\boldsymbol{\alpha}_{opt}$ for the Oracle policy. The risk score is a linear combination the two baseline covariates (X_1 and X_2), non-compliance ("Non-comp"; $\log(|\delta_{t-1} - A_{t-1}| + 1)$), and disease status ("Cur Y"; Y_{t-1}):

Table 1: Simulation study results comparing the baseline model with 6-month recommendation for all subjects, policy search methods based on Dirichlet process mixture (DPM) fits, and the oracle model which uses the true data-generating model to estimate the policy.

	Value (larger is preferred)		
Simulation setting	Baseline	DPM	Oracle
Scenario 1	-1.52	-0.97	-0.96
Scenario 2	-1.48	-1.15	-1.15
Scenario 3	-1.37	-0.83	-0.83

	Average recommended recall time		
Simulation setting	Baseline	DPM	Oracle
Scenario 1	6.00	6.00	6.00
Scenario 2	6.00	5.99	6.00
Scenario 3	6.00	6.00	6.00

3") shows that the optimal policy estimation is unaffected by this type of model misspecification. In addition to estimating the alpha's well, Table 1 shows that the DPM approach gives nearly the optimal value for all three scenarios.

In summary, while G-computing methods are reliant on a model that fits the data reasonably well, our proposed non-parametric Bayesian model is very flexible and is able to accommodate complicated data structures. Of course, no single model can fit every data set and so we emphasize that goodness-of-fit testing remains an important step in our approach.

2.2 Comparison with the overall oracle policy in a simple setting

We conduct an additional simulation study for a simplified problem to evaluate how close the optimal policy estimated within our proposed class of policies is to the oracle policy. In this simulation setting, we consider three stages of follow-up, and a binary treatment Z_{it} at each stage. Subjects are generated from 2 groups, with the group identifier for subject i denoted by $G_i \in \{1, 2\}$

and Prob$(G_i = 1) = 0.8$. For $t = 1, 2, 3$, subjects from the 1st group are generated as

$$
\begin{aligned}
&(\mathbf{X}_i^T, Y_{i0})^T | G_i = 1 \sim \text{Normal}(0, \mathbf{\Sigma}_0) \\
&Y_{it} | Y_{it-1}, Z_{it}, G_i = 1 \sim \\
&\text{Normal}\left[0.1 + 0.2X_{i2} + 0.2Z_{it} + 0.9Y_{it-1} + 0.2X_{i1}Z_{it} + 0.2X_{i2}Z_{it} + 0.3Y_{it-1}Z_{it}, \sigma^2\right];
\end{aligned}
$$

subjects from the second group are generated as

$$
\begin{aligned}
&(\mathbf{X}_i^T, Y_{i0})^T | G_i = 2 \sim \text{Normal}\left[(1, 0, 0)^T, \mathbf{\Sigma}_0\right] \\
&Y_{it} | Y_{it-1}, Z_{it}, G_i = 2 \sim \\
&\text{Normal}\left[0.1 + 0.3X_{i1} - 0.2Z_{it} + 0.9Y_{it-1} + 0.1X_{i1}Z_{it} + 0.3X_{i2}Z_{it} + 0.2Y_{it-1}Z_{it}, \sigma^2\right],
\end{aligned}
$$

where $\sigma_1 = 0.1$, $\sigma_2 = 0.5$, and $\mathbf{\Sigma}_0$ is the correlation matrix with 0.5 for all off-diagonal elements.

For this simplified problem, we assume there is no cost constraint. Our policy takes the form $Z_t(\boldsymbol{\alpha}) = \mathbb{I}(\alpha_0 + \alpha_1 X_1 + \alpha_2 X_2 + \alpha_3 Y_{t-1} > 0)$, where $\boldsymbol{\alpha} = (\alpha_0, \alpha_1, \alpha_2, \alpha_3)$ are the policy feature weights subject to the restriction $||\boldsymbol{\alpha}|| = (\sum_{j=0}^{3} \alpha_j^2)^{1/2} = 1$. The feature weights $\boldsymbol{\alpha}$ are optimized to maximize the reward of interest $V(\boldsymbol{\alpha}) = \mathbb{E}_{\boldsymbol{\alpha}}(Y_0 - Y_3)$. We calculate the expected value of the reward $\mathbb{E}(Y_{i0} - Y_{i3})$ for each of the eight treatment trajectories for each subject, and the one with highest expected reward is the oracle optimal policy for that subject, denoted as $\mathbf{Z}_i^o = (Z_{i1}^o, Z_{i2}^o, Z_{i3}^o)$.

We simulated $20,000$ subjects, and evaluate their expected reward under our proposed policy with optimal feature weights and under the oracle policy. Under these settings, the expected reward value from our policy that uses the optimal feature weights $\hat{\boldsymbol{\alpha}}$ is 0.10, and is very close to the

expected reward value under the oracle policy, 0.11. Furthermore, we also compute the proportion of the $20,000$ subjects that are given the same first-stage treatment under both policies. We observe that the treatment assignment at the first stage from our policy agrees with the oracle policy for 91% of the subjects.

From these simple simulation results, we conclude that our proposed policy-search method to estimate the optimal policy under the pre-specified class of polices is a very close approximation to the overall optimal policy. As part of the simple design, we assume the treatment to be binary, and not 'time until next visit'. Also, our policy depends only on data available at baseline, with no cost-constraints, and with data generated under perfect compliance and noninformative time between visits. Estimating the oracle policy under these features can be complex; hence, we resorted to the simple design above to compare the two.

On the overall, our policy has the potential to be applied to various complicated scenarios.

3 Model comparison and goodness of fit diagnostics

Because our dataset is large, we use conventional validation methods to check our model fit. 80% of the subjects are randomly assigned to the training dataset and the remaining 20% to the testing dataset. We apply the models to the training dataset to obtain posterior samples of the random effects and other parameters. Given posterior samples, and conditioning on the observed baseline information $\mathbf{X}_i$ and Y_{i0} and observed time between visits δ_{it} for subjects in the testing dataset, we can predict the disease progression of each subject in the testing dataset, and compare the predicted and observed responses at the final visit. We compare the validation results of six models: Gaussian model (DPM model with $L = 1$) with all adjustments in Section 7.1; DPM model with

$L = 5, 10, 25$ and all adjustments in Section 7.1; DPM model with $L = 10$ and adjustments in Section 7.1 but no scale mixture of normals; DPM with $L = 10$ and adjustments in Section 7.1 but no excess probability on carrying the previous response forward.

Several metrics are used to assess the predictive performance of those models: mean squared error (MSE), mean absolute error (MAE), continuous ranked probability score (CRPS) (Matheson and Winkler, 1976), probability integral transforms (PIT) value, and the coverage of 95% posterior predictive credible intervals. The first two metrics only consider the predictive posterior means, and compare them with the observed values. CRPS generalizes MAE to probabilistic forecasts by comparing the full predictive posterior distribution with the observation (Gneiting et al., 2007). CRPS is defined as $\text{CRPS} = 1/n_{\text{test}} \sum_{i=1}^{n_{\text{test}}} \int [F_i(y) - \mathbf{1}(y \geq y_i)]^2 dy$, where n_{test} is the number of subjects in the testing dataset, y_i is the observed response value at the last visit for subject i in the testing dataset and F_i is the corresponding predictive cumulative distribution function. PIT is defined as $\text{PIT}_i = \Pr(y_i^{\text{rep}} < y_i | y_{\text{training}}) + 1/2 \Pr(y_i^{\text{rep}} = y_i | y_{\text{training}})$, where y_i is the observed response value at the last visit for subject i in the testing dataset. The second term in the definition is used to accommodate the non-zero mass on the response value of 0 or 1 (Gelman et al., 2014). If the sequence of predictive distribution coincides with the associated true data-generating process, then the PIT values will have the uniform distribution (Gneiting et al., 2007). Therefore, we compare the PIT values with the uniform distribution to check the performance of the model prediction. We quantify this by estimating area under the curve (AUC) of absolute difference between theoretical quantiles and empirical quantiles of PIT values and above the zero horizontal line. For the first four metrics, smaller value indicates better prediction performance.

Results are given in Table 2 and Figure 3. DPM models with adjustments in Section 7.1 have

Table 2: Validation results comparing six models in terms of mean square error (MSE), mean absolute error (MAE), continuous ranked probability score (CRPS), area under the curve (AUC) of the absolute difference between observed and expected PIT curves, and the coverage of the credible intervals (Coverage). "Gaussian" indicates Gaussian model with all adjustments in Section 7.1; "DPM, $L = l$" indicates Dirichlet process mixture model with number of mixture components $L = l$ and all adjustments in Section 7.1; "DPM, $L = 10$, no scale mixture" indicates DPM model with $L = 10$ and adjustments in Section 7.1 but no scale mixture of normals; "DPM, $L = 10$, no excess prob" indicates DPM model with $L = 10$ and adjustments in Section 7.1 but no excess probability on carrying the previous response. All the values shown in the table are the original result values $\times 100$.

	MSE	MAE	CRPS	AUC	Coverage
Gaussian	0.36	3.94	2.87	2.86	93.38
DPM, $L = 5$	0.36	3.93	2.85	2.48	93.92
DPM, $L = 10$	0.36	3.93	2.85	2.21	94.02
DPM, $L = 25$	0.36	3.96	2.85	2.15	94.44
DPM, $L = 10$, no scale mixture	0.40	4.16	3.00	4.32	96.58
DPM, $L = 10$, no excess prob	0.55	4.27	3.35	6.42	92.96

Figure 3: Theoretical quantiles of the uniform distribution versus empirical quantiles of PIT values and their absolute differences.

smaller CRPS, better coverage and the distributions of their PIT values are closer to the uniform distribution compared to the Gaussian model, especially when $L = 10$ and $L = 25$. The adjustment of adding excess probability that responses are identical to the previous response and using scale mixture of normals substantially improve the prediction performance. It seems that increasing the number of mixture components from 10 to 25 does not substantially improve the performance. As a result, we choose DPM model with $L = 10$ mixture components and all adjustments in Section 7.1 as our final model for disease progression.

References

Gelman, A., Carlin, J. B., Stern, H. S., and Rubin, D. B. (2014). *Bayesian Data Analysis*, volume 2. Chapman & Hall/CRC Boca Raton, FL, USA.

Gneiting, T., Balabdaoui, F., and Raftery, A. E. (2007). Probabilistic forecasts, calibration and sharpness. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **69,** 243–268.

Matheson, J. E. and Winkler, R. L. (1976). Scoring rules for continuous probability distributions. *Management Science* **22,** 1087–1096.