Supporting Information for "A Bayesian multivariate mixture model for skewed longitudinal data with intermittent missing observations: An application to infant motor development" by

Carter Allen[1], Sara E. Benjamin-Neelon[2] and Brian Neelon[3*]

1: Department of Biomedical Informatics, The Ohio State University, Columbus, OH, U.S.A.

2: Department of Health, Behavior and Society, Johns Hopkins University, Baltimore, MD, U.S.A.

3: Medical University of South Carolina, Charleston, South Carolina, U.S.A.

*email: neelon@musc.edu

Web Appendix A: Proof of Proposition 1

For each cluster $k = 1, ..., K$, let $\mathbf{Y}_k = \mathbf{X}_k^* \mathbf{B}_k^* + \mathbf{E}_k$, where $\mathbf{Y}_k$ is an $n_k \times J$ response matrix, $\mathbf{X}_k^*$ is an $n_k \times (p+1)$ matrix of predictors and latent truncated normal random variables from equation (6) in the manuscript, $\mathbf{B}_k^*$ is a $(p+1) \times J$ matrix of regression and skewness coefficients from equation (6), and $\mathbf{E}_k$ is the $n_k \times J$ matrix of residuals associated with $\mathbf{Y}_k$. We assume $\mathbf{E}_k \sim \text{MatNorm}(\mathbf{0}, \mathbf{I}_{n_k}, \mathbf{\Sigma}_k)$, where $\mathbf{0}$ is an $n_k \times J$ matrix of 0's, $\mathbf{I}_{n_k}$ is the n_k dimensional identity matrix, and $\mathbf{\Sigma}_k$ is a $J \times J$ variance-covariance matrix. As prior distributions, we assume $\mathbf{B}_k^* | \mathbf{\Sigma}_k \sim \text{MatNorm}(\mathbf{B}_{0k}^*, \mathbf{L}_{0k}, \mathbf{\Sigma}_k)$ and $\mathbf{\Sigma}_k \sim \text{IW}(\nu_{0k}, \mathbf{V}_{0k})$. That is, $\mathbf{B}_k^*$ and $\mathbf{\Sigma}_k$ have a joint Matrix Normal–Inverse Wishart (IW) prior, denoted $\text{MatNorm–IW}_{(p+1)\times J}(\mathbf{B}_{0k}^*, \mathbf{L}_{0k}, \nu_{0k}, \mathbf{V}_{0k})$, of the form

$$\begin{aligned} \pi(\mathbf{B}_k^*, \mathbf{\Sigma}_k) &= \pi(\mathbf{B}_k^* | \mathbf{\Sigma}_k) \pi(\mathbf{\Sigma}_k) \\ &= \text{MatNorm}_{(p+1)\times J}(\mathbf{B}_{0k}^*, \mathbf{L}_{0k}, \mathbf{\Sigma}_k) \text{IW}(\nu_{0k}, \mathbf{V}_{0k}), \end{aligned}$$

where $\mathbf{B}_{0k}^*$ is a $(p+1) \times J$ prior mean matrix, $\mathbf{L}_{0k}$ and $\mathbf{V}_{0k}$ are, respectively, $(p+1) \times (p+1)$ and $J \times J$ prior scale matrices, and ν_{0k} denotes the prior degrees of freedom. Under this set-up, the full conditional distribution for $\mathbf{B}_k^*$ can be obtained as follows:

$$\begin{aligned} \mathbf{B}_k^* | \mathbf{\Sigma}_k, \mathbf{Y}_k &\propto \exp\left(-\frac{1}{2}\left[\text{tr}\{\mathbf{\Sigma}_k^{-1}(\mathbf{Y}_k - \mathbf{X}_k^*\mathbf{B}_k^*)^T(\mathbf{Y}_k - \mathbf{X}_k^*\mathbf{B}_k^*)\} + \text{tr}\{\mathbf{\Sigma}_k^{-1}(\mathbf{B}_k^* - \mathbf{B}_{0k}^*)^T \mathbf{L}_{0k}^{-1}(\mathbf{B}_k^* - \mathbf{B}_{0k}^*)\}\right]\right) \\ &\propto \exp\left(-\frac{1}{2}\left[\text{tr}\{\mathbf{\Sigma}_k^{-1}(\mathbf{Y}_k - \mathbf{X}_k^*\mathbf{B}_k^*)^T(\mathbf{Y}_k - \mathbf{X}_k^*\mathbf{B}_k^*) + (\mathbf{B}_k^* - \mathbf{B}_{0k}^*)^T \mathbf{L}_{0k}^{-1}(\mathbf{B}_k^* - \mathbf{B}_{0k}^*)\}\right]\right) \\ &\propto \exp\left(-\frac{1}{2}\left[\text{tr}\{\mathbf{\Sigma}_k^{-1}(\mathbf{B}_k^* - \mathbb{B}_k^*)^T \mathbf{L}_k^{-1}(\mathbf{B}_k^* - \mathbb{B}_k^*)\}\right]\right) \text{ after completing the square,} \end{aligned}$$

where $\mathbb{B}_k^* = \mathbf{L}_k(\mathbf{L}_{0k}^{-1}\mathbf{B}_{0k}^* + \mathbf{X}_k^{*T}\mathbf{Y}_k)$ and $\mathbf{L}_k = (\mathbf{L}_{0k}^{-1} + \mathbf{X}_k^{*T}\mathbf{X}_k^*)^{-1}$. Similarly, we may express $f(\mathbf{\Sigma}_k | \mathbf{B}_k^*, \mathbf{Y}_k)$ as

$$\begin{aligned} f(\mathbf{\Sigma}_k | \mathbf{B}_k^*, \mathbf{Y}_k) &\propto f(\mathbf{Y}_k | \mathbf{B}_k^*, \mathbf{\Sigma}_k) \pi(\mathbf{B}_k^* | \mathbf{\Sigma}_k) \pi(\mathbf{\Sigma}_k), \text{ where} \\ f(\mathbf{Y}_k | \mathbf{B}_k^*, \mathbf{\Sigma}_k) &\propto |\mathbf{\Sigma}_k|^{-n_k/2} \exp\left(-\frac{1}{2}\left[\text{tr}\{\mathbf{\Sigma}_k^{-1}(\mathbf{Y}_k - \mathbf{X}_k^*\mathbf{B}_k^*)^T(\mathbf{Y}_k - \mathbf{X}_k^*\mathbf{B}_k^*)\}\right]\right), \\ \pi(\mathbf{B}_k^* | \mathbf{\Sigma}_k) &\propto |\mathbf{\Sigma}_k|^{-(p+1)/2} \exp\left(-\frac{1}{2}\left[\text{tr}\{\mathbf{\Sigma}_k^{-1}(\mathbf{B}_k^* - \mathbf{B}_{0k}^*)^T \mathbf{L}_{0k}^{-1}(\mathbf{B}_k^* - \mathbf{B}_{0k}^*)\}\right]\right), \text{ and} \\ \pi(\mathbf{\Sigma}_k) &\propto |\mathbf{\Sigma}_k|^{-(\nu_{0k}+J)/2} \exp\left[-\frac{1}{2}\text{tr}(\mathbf{\Sigma}_k^{-1}\mathbf{V}_{0k})\right]. \end{aligned}$$

Combining terms, we have

$$\begin{aligned} f(\mathbf{\Sigma}_k | \mathbf{B}_k^*, \mathbf{Y}_k) &\propto |\mathbf{\Sigma}_k|^{-\frac{n_k + \nu_{0k} + (p+1) + k + 1}{2}} \\ &\times \exp\left(-\frac{1}{2}\text{tr}\left[\mathbf{\Sigma}_k^{-1}\{\mathbf{V}_{0k} + (\mathbf{Y}_k - \mathbf{X}_k^*\mathbf{B}_k^*)^T(\mathbf{Y}_k - \mathbf{X}_k^*\mathbf{B}_k^*) + (\mathbf{B}_k^* - \mathbf{B}_{0k}^*)^T \mathbf{L}_{0k}^{-1}(\mathbf{B}_k^* - \mathbf{B}_{0k}^*)\}\right]\right) \end{aligned}$$

Thus, $\mathbf{\Sigma}_k|\mathbf{B}_k^*, \mathbf{Y}_k \sim \text{IW}(\nu_k, \mathbf{V}_k)$, where

$$\begin{aligned}\nu_k &= \nu_0 + n_k + p + 1, \quad \text{and} \\ \mathbf{V}_k &= \mathbf{V}_{0k} + (\mathbf{B}_k^* - \mathbf{B}_{0k}^*)^T \mathbf{L}_{0k}^{-1}(\mathbf{B}_k^* - \mathbf{B}_{0k}^*) + (\mathbf{Y}_k - \mathbf{X}_k^*\mathbf{B}_k^*)^T(\mathbf{Y}_k - \mathbf{X}_k^*\mathbf{B}_k^*),\end{aligned}$$

as outlined in Proposition 1 of the manuscript.

Web Appendix B: MCMC Algorithm

In this section we outline Gibbs updates of all model parameters. For ease of notation, each parameter update is implicitly assumed to be conditional on the data and other model parameters. All notation is defined as in Section 3. The algorithm presented below is not necessarily optimized for computational efficiency.

1. *Update missing responses,* $\mathbf{y}_i^{miss}$. For $i = 1, ..., n$ and given $z_i = k$:
   (a) Partition $\boldsymbol{\mu}_{ki}$ and $\mathbf{\Sigma}_k$, the variance of $\boldsymbol{\epsilon}_{ki}$ in equation (3) of the manuscript, as
   $$\boldsymbol{\mu}_{ki} = \begin{pmatrix} \boldsymbol{\mu}_{ki}^{miss} \\ \boldsymbol{\mu}_{ki}^{obs} \end{pmatrix} \text{ and } \mathbf{\Sigma}_k = \begin{pmatrix} \mathbf{\Sigma}_{k11} & \mathbf{\Sigma}_{k12} \\ \mathbf{\Sigma}_{k21} & \mathbf{\Sigma}_{k22} \end{pmatrix};$$
   (b) Compute $\boldsymbol{\mu}_{ki}^{cond} = \boldsymbol{\mu}_{ki}^{miss} + \mathbf{\Sigma}_{k12}\mathbf{\Sigma}_{k22}^{-1}(\mathbf{y}_i^{obs} - \boldsymbol{\mu}_{ki}^{obs})$.
   (c) Compute $\mathbf{\Sigma}_k^{cond} = \mathbf{\Sigma}_{k11} - \mathbf{\Sigma}_{k12}\mathbf{\Sigma}_{k22}^{-1}\mathbf{\Sigma}_{k21}$.
   (d) Sample $\mathbf{y}_i^{miss}$ from $\text{N}_{J-q_i}(\boldsymbol{\mu}_{ki}^{cond}, \mathbf{\Sigma}_k^{cond})$.
2. *For* $k = 1, ..., K$, *use Pólya-Gamma data augmentation to update the logistic regression parameters,* $\boldsymbol{\gamma}_k, \mathbf{b}_k = (b_{ki}, \ldots, b_{kn_k})^T$, *and* σ_k^2 *for the missing data model described in equation (15) of Section 3.4.*
   (a) Compute $n_k = \sum_{i=1}^n \mathbb{1}_{(z_i=k)}$, the number of subjects in cluster k.
   (b) For $i = 1, ..., n_k$ and $j = 1, ..., J$:
      i. Compute $\text{logit}(\phi_{kij}) = \mathbf{x}_{ij}^T\boldsymbol{\gamma}_k + b_{ki}$ as in equation (15), where $\mathbf{x}_{ij}$ is an $m \times 1$ vector of covariates that may overlap with those used in the MSN model, and $\boldsymbol{\gamma}_k$ is an $m \times 1$ vector of regression parameters as in equation (15).
      ii. Update Pólya-Gamma weights w_{kij} from $\text{PG}\{1, \text{logit}(\phi_{kij})\}$.
      iii. Compute $h_{kij} = \frac{R_{kij}-1/2}{w_{kij}} - b_{ki}$, where R_{kij} is the binary indicator of whether the response for subject i in cluster k at timepoint j is missing, as in equation (14) of the manuscript, and b_{ki} is the random intercept for subject i in cluster k, as in equation (15).
   (c) Form the vector $\mathbf{h}_k = (h_{k11}, ..., h_{kn_kJ})^T$.
   (d) Form the diagonal matrix $\mathbf{O}_k$ with entries $(w_{k11}, ..., w_{kn_kJ})$.
   (e) Compute $\mathbf{G}_k = (\mathbf{G}_{0k}^{-1} + \mathbf{X}_k^T\mathbf{O}_k\mathbf{X}_k)^{-1}$, where $\mathbf{G}_{0k}$ is the $m \times m$ prior covariance of $\boldsymbol{\gamma}_k$.
   (f) Compute $\mathbf{g}_k = \mathbf{G}_k(\mathbf{G}_{0k}^{-1}\mathbf{g}_{0k} + \mathbf{X}_k^T\mathbf{O}_k\mathbf{h}_k)$, where $\mathbf{g}_{0k}$ is the prior mean of $\boldsymbol{\gamma}_k$.
   (g) Compute $\tau_k = 1/\sigma_k^2$, where σ_k^2 is the variance of b_{ki}.
   (h) For $i = 1, ..., n_k$:
      i. Compute $v_{ki} = (\tau_k + \sum_{j=1}^J w_{kij})^{-1}$
      ii. Compute $m_{ki} = v_{ki}\{\sum_{j=1}^J w_{kij}(h_{kij} - \mathbf{x}_{ij}^T\boldsymbol{\gamma}_k)\}$.
      iii. Update b_{ki} from $\text{N}(m_{ki}, v_{ki})$.
   (i) Update σ_k^2 from $\text{IG}\{\lambda_{1k}+n_k/2, \lambda_{2k}+(\sum_{i=1}^{n_k} b_{ki}^2)/2\}$, where σ_k^2 is assumed to have a $\text{IG}(\lambda_{1k}, \lambda_{2k})$ prior distribution. Alternatively, update τ_k from a $\text{Gamma}\{\lambda_{1k}+n_k/2, \lambda_{2k}+(\sum_{i=1}^{n_k} b_{ki}^2)/2\}$, where $\text{Gamma}(a, b)$ denotes a gamma distribution with shape parameter a and rate parameter b.
3. *Update the multinomial logit regression parameters for the cluster allocation model as described in Section 3.2.* For $k = 1, ..., K - 1$:

(a) For $i = 1, ...n$:

i. Define $U_{ki} = \mathbb{1}_{(z_i=k)}$.

ii. Compute $c_{ki} = \log(1 + \sum_{h \notin \{k,K\}} e^{\mathbf{w}_i^T \boldsymbol{\delta}_h})$ as described in Section 3.2.

iii. Compute $\eta_{ki} = \mathbf{w}_i^T \boldsymbol{\delta}_k - c_{ki}$ as in equation (10).

iv. Update ω_{ki} from PG$(1, \eta_{ki})$.

(b) Define $\mathbf{U}_k^* = \left(\frac{U_{k1}-1/2}{\omega_{k1}} + c_{k1}, ..., \frac{U_{kn}-1/2}{\omega_{kn}} + c_{kn}\right)^T$ as described in Section 3.2.

(c) Compute $\mathbf{S}_k = (\mathbf{S}_{0k}^{-1} + \mathbf{W}^T \mathbf{O}_k \mathbf{W})^{-1}$, where $\mathbf{O}_k$ is the diagonal matrix with entries $(\omega_{k1}, ..., \omega_{kn})$ and $\mathbf{S}_{0k}$ is the prior covariance of $\boldsymbol{\delta}_k$.

(d) Compute $\mathbf{d}_k = \mathbf{S}_k(\mathbf{S}_{0k}^{-1}\mathbf{d}_{0k} + \mathbf{W}^T \mathbf{O}_k \mathbf{U}_k^*)$, where $\mathbf{d}_{0k}$ is the prior mean of $\boldsymbol{\delta}_k$.

(e) Update $\boldsymbol{\delta}_k$ from $\mathrm{N}_r(\mathbf{d}_k, \mathbf{S}_k)$.

4. *Update cluster indicators $z_1, ..., z_n$.* For $i = 1, ..., n$, iterate through the following steps:

(a) For $k = 1, ..., K$:

i. Compute $p_{ki} = \mathrm{dnorm}(\mathbf{y}_i; \boldsymbol{\mu}_{ki}, \boldsymbol{\Sigma}_k)$, where $\mathrm{dnorm}(\mathbf{y}; \boldsymbol{\mu}, \boldsymbol{\Sigma})$ denotes the density of a multivariate normal random variable with mean $\boldsymbol{\mu}$ and covariance $\boldsymbol{\Sigma}$ evaluated at $\mathbf{y}$; and $\boldsymbol{\mu}_{ki} = \mathbf{X}_i \boldsymbol{\beta}_k + t_i \boldsymbol{\psi}_k$, where $\mathbf{X}_i$ is the $J \times Jp$ design matrix defined in equation (3) and $\boldsymbol{\beta}_k = \mathrm{vec}(\mathbf{B}_k) = (\beta_{k11}, \ldots, \beta_{k1p}, \ldots, \beta_{kJ1}, \ldots, \beta_{kJp})^T$ is a $Jp \times 1$ vector of cluster- and outcome-specific regression coefficients also defined as in equation (3). When covariates are not time dependent, we may simplify notation to $\boldsymbol{\mu}_{ki} = \mathbf{B}_k^{*T} \mathbf{x}_{ki}^*$, where $\mathbf{x}_{ki}^{*T}$ is given by the i^{th} row of the $n_k \times (p+1)$ matrix $\mathbf{X}_k^*$, where

$$\underset{n_k \times (p+1)}{\mathbf{X}_k^*} = \begin{pmatrix} x_{11} & \ldots & x_{1p} & t_{ki} \\ \vdots & \ddots & \vdots & \vdots \\ x_{n_k 1} & \ldots & x_{n_k p} & t_{kn_k} \end{pmatrix} \quad \text{and} \quad \underset{(p+1) \times J}{\mathbf{B}_k^*} = \begin{pmatrix} \beta_{k11} & \ldots & \beta_{kJ1} \\ \vdots & \ddots & \vdots \\ \beta_{k1p} & \ldots & \beta_{kJp} \\ \psi_{k1} & \ldots & \psi_{kJ} \end{pmatrix},$$

as in Section 3.1.

ii. From equation (15) in the manuscript, compute $\phi_{ki} = \mathrm{logit}^{-1}(\mathbf{x}_{ij}^T \boldsymbol{\gamma}_k + b_i)$.

iii. Compute $\rho_{ki} = \prod_{j=1}^{J} \mathrm{dbern}(R_{ij}; \phi_{ki})$, where dbern() denotes the Bernoulli distribution function.

(b) Compute $\boldsymbol{\pi}_i = (\pi_{1i}, ..., \pi_{Ki})$, where $\pi_{ki} = \frac{e^{\mathbf{w}_i^T \boldsymbol{\delta}_k}}{\sum_{h=1}^{K} e^{\mathbf{w}_i^T \boldsymbol{\delta}_h}}$ for $k = 1, ..., K$, as denoted in equation (7) of the manuscript. Recall that cluster K serves as the reference category, implying that $\boldsymbol{\delta}_K = \mathbf{0}$.

(c) Compute the posterior probability $\Pr(z_i = k) = \frac{\pi_{ki} p_{ki} \rho_{ki}}{\sum_{l=1}^{K} \pi_{li} p_{li} \rho_{li}}$, for $k = 1, ..., K$. Note that under (marginal) MAR imputation, ρ is left out of this equation, as the missing data model is fully ignorable in this case.

(d) Update z_i from Categorical$\{\Pr(z_i = 1), ..., \Pr(z_i = K)\}$.

5. *Update the multivariate skew normal regression parameters as described in Section 3.1.* We first consider the case where there are no time-dependent covariates. We then consider time-varying designs.

(a) Time-Invariant Designs:

i. For $i = 1, \ldots, n$ and given $z_i = k$, update t_i from its $\mathrm{N}_+(a_{ki}, A_k)$ full conditional, where $\mathrm{N}_+()$ denotes a truncated normal random variable restricted to the positive real line,

$$\begin{aligned} A_k &= (1 + \boldsymbol{\psi}_k^T \boldsymbol{\Sigma}_k^{-1} \boldsymbol{\psi}_k)^{-1}, \\ a_{ki} &= A_k \boldsymbol{\psi}_k^T \boldsymbol{\Sigma}_k^{-1} (\mathbf{y}_i - \mathbf{B}_k^T \mathbf{x}_{ki}), \end{aligned}$$

$\mathbf{y}_i = (y_{i1}, \ldots, y_{iJ})^T$, $\boldsymbol{\psi}_k = (\psi_{k1}, \ldots, \psi_{kJ})^T$, $\mathbf{B}_k$ is the $p \times J$ defined below, and $\mathbf{x}_{ki}$ is the $p \times 1$ vector formed from the i-th row of $\mathbf{X}_k$ described below.

$$\mathbf{X}_k = \begin{pmatrix} x_{11} & \dots & x_{1p} \\ \vdots & \ddots & \vdots \\ x_{n_k1} & \dots & x_{n_kp} \end{pmatrix} \text{ and } \mathbf{B}_k = \begin{pmatrix} \beta_{k11} & \dots & \beta_{k1J} \\ \vdots & \ddots & \vdots \\ \beta_{kp1} & \dots & \beta_{kpJ} \end{pmatrix}.$$

ii. For $k = 1, \dots, K$, draw $\mathbf{B}_k^*$ from $\text{MatNorm}_{(p+1)\times J}(\mathbb{B}_k^*, \mathbf{L}_k, \boldsymbol{\Sigma}_k)$ as described in Proposition 1. Note that the $(p+1)^{th}$ row of $\mathbf{B}_k^*$ contains $\boldsymbol{\psi}_k = (\psi_{k1}, \dots, \psi_{kJ})^T$. Therefore, we vectorize $\mathbf{B}_k^*$ into $Jp \times 1$ vector $\boldsymbol{\beta}_k$ and $J \times 1$ vector $\boldsymbol{\psi}_k$ to perform the back transformations described above. To draw from the matrix normal density, make use of the R package `matrixsample` (Laurent, 2018).

iii. For $k = 1, \dots, K$, update $\underset{J\times J}{\boldsymbol{\Sigma}_k}$ from $\text{IW}(\nu_k, \mathbf{V}_k)$ as described in Proposition 1.

iv. To backtransform to original MSN representations, use.

$$\begin{aligned} \boldsymbol{\zeta}_{ki} &= \mathbf{X}_i \boldsymbol{\beta}_k, \\ \boldsymbol{\Omega}_k &= \boldsymbol{\Sigma}_k + \boldsymbol{\psi}_k \boldsymbol{\psi}_k^T, \\ \boldsymbol{\alpha}_k &= \frac{1}{\sqrt{1 - \boldsymbol{\psi}_k^T \boldsymbol{\Omega}_k^{-1} \boldsymbol{\psi}_k}} \boldsymbol{\omega}_k \boldsymbol{\Omega}_k^{-1} \boldsymbol{\psi}_k, \quad \text{and} \\ \boldsymbol{\omega}_k &= \text{Diag}(\boldsymbol{\Omega}_k)^{1/2}, \end{aligned}$$

where $\text{Diag}(\boldsymbol{\Omega}_k)$ extracts the diagonal elements of $\boldsymbol{\Omega}_k$ as a vector.

(b) Time-Varying Designs: For designs that include time-varying covariates, we work with equation (3) in the manuscript.

i. For $i = 1, \dots, n$ and given $z_i = k$, update t_i: To update t_i given $z_i = k$, we create a $J \times Jp$ design matrix $\mathbf{X}_i$ and $Jp \times 1$ vector $\boldsymbol{\beta}_k$ of the form

$$\begin{aligned} \underset{J\times Jp}{\mathbf{X}_i} &= \begin{pmatrix} x_{i11} & \dots & x_{i1p} & 0 & \dots & 0 & \dots & 0 \\ & & \vdots & \ddots & \vdots & & & \\ 0 & \dots & 0 & 0 & \dots & x_{iJ1} & \dots & x_{iJp} \end{pmatrix} \\ \underset{Jp\times 1}{\boldsymbol{\beta}_k} &= (\beta_{k11}, \dots, \beta_{k1p}, \dots, \beta_{kJ1}, \dots, \beta_{kJp})^T. \end{aligned}$$

Next, we draw $t_i | (z_i = k)$ from its $\text{N}_+(a_{ki}, A_k)$ full conditional, where

$$\begin{aligned} A_k &= (1 + \boldsymbol{\psi}_k^T \boldsymbol{\Sigma}_k^{-1} \boldsymbol{\psi}_k)^{-1}, \\ a_{ki} &= A_k \boldsymbol{\psi}_k^T \boldsymbol{\Sigma}_k^{-1} (\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta}_k), \end{aligned}$$

$\mathbf{y}_i = (y_{i1}, \dots, y_{iJ})^T$, and $\boldsymbol{\psi}_k = (\psi_{k1}, \dots, \psi_{kJ})^T$.

ii. For $k = 1, \dots, K$, update $\boldsymbol{\beta}_k$, $\boldsymbol{\psi}_k$: To update the regression parameters, we create an augmented $J \times J(p+1)$ design matrix $\mathbf{X}_i^*$ and $J(p+1) \times 1$ vector $\boldsymbol{\beta}_k^*$ of the form

$$\begin{aligned} \underset{J\times J(p+1)}{\mathbf{X}_i^*} &= \begin{pmatrix} x_{i11} & \dots & x_{i1p} & t_i & 0 & \dots & 0 & \dots & 0 & 0 \\ & & & \vdots & \ddots & \vdots & & & & \\ 0 & \dots & 0 & 0 & 0 & \dots & x_{iJ1} & \dots & x_{iJp} & t_i \end{pmatrix} \\ \underset{J(p+1)\times 1}{\boldsymbol{\beta}_k^*} &= (\beta_{k11}, \dots, \beta_{k1p}, \psi_1, \dots, \beta_{kJ1}, \dots, \beta_{kJp}, \psi_J)^T. \end{aligned}$$

Next, for all k, we assign independent multivariate normal and IW priors to $\boldsymbol{\beta}_k^*$ and $\boldsymbol{\Sigma}_k$:

$$\begin{aligned} \underset{J(p+1)\times 1}{\boldsymbol{\beta}_k^*} &\sim \text{N}_{J(p+1)}\left(\boldsymbol{\beta}_0, \mathbf{T}_0^{-1}\right) \quad \text{and} \\ \underset{J\times J}{\boldsymbol{\Sigma}_k} &\sim \text{IW}(\nu_0, \mathbf{S}_0), \end{aligned}$$

where $\mathbf{T}_0$ is a $J(p+1) \times J(p+1)$ prior precision matrix and, in this context, $\mathbf{S}_0$ is a $J \times J$ prior scale matrix. Following standard algebraic routines for conjugate multivariate normal priors, we arrive at the following full conditional for $\boldsymbol{\beta}_k^*$:

$$\begin{aligned}
\underset{J(p+1)\times 1}{\boldsymbol{\beta}_k^*} &\sim \mathrm{N}_{J(p+1)}\left(\mathbf{m}_k, \mathbf{V}_k\right), \quad \text{where} \\
\underset{J(p+1)\times J(p+1)}{\mathbf{V}_k} &= \left\{\mathbf{T}_0 + \mathbf{X}_k^{*T}\left(\mathbf{I}_{n_k} \otimes \boldsymbol{\Sigma}_k^{-1}\right)\mathbf{X}_k^*\right\}^{-1} \quad \text{and} \\
\underset{J(p+1)\times 1}{\mathbf{m}_k} &= \mathbf{V}_k\left\{\mathbf{T}_0\boldsymbol{\beta}_0 + \mathbf{X}_k^{*T}\left(\mathbf{I}_{n_k} \otimes \boldsymbol{\Sigma}_k^{-1}\right)\mathbf{y}_k\right\}.
\end{aligned}$$

Here, $\mathbf{y}_k$ denotes the $Jn_k \times 1$ vector of responses for each observation in cluster k after imputation, and here $\mathbf{X}_k^*$ denotes an $Jn_k \times J(p+1)$ matrix formed by stacking $\mathbf{X}_i^*$ for all subjects in cluster k. To perform the back transformations described in equation (4) of the manuscript, we extract the $Jp \times 1$ vector $\boldsymbol{\beta}_k$ and the $J \times 1$ vector $\boldsymbol{\psi}_k = (\psi_{k1}, \ldots, \psi_{kJ})^T$ from $\boldsymbol{\beta}_k^*$.

iii. Finally, for $k = 1, \ldots, K$, we update $\boldsymbol{\Sigma}_k$ from $\mathrm{IW}(\nu_k, \mathbf{S}_k)$ where

$$\begin{aligned}
\nu_k &= \nu_0 + n_k \quad \text{and} \\
\underset{J\times J}{\mathbf{S}_k} &= \mathbf{S}_0 + \mathbf{R}_k^T\mathbf{R}_k,
\end{aligned}$$

where $\mathbf{R}_k$ is an $n_k \times J$ matrix with i-th row equal to $(\mathbf{y}_i - \mathbf{X}_i^*\boldsymbol{\beta}_k^*)^T$ for all i in cluster k.

(c) For both time-varying and time-invariant designs, we back transform to obtain $\boldsymbol{\alpha}_k$ and $\boldsymbol{\Omega}_k$ as described in equation (4) of the manuscript. In the time-invariant case, we vectorize the $(p+1) \times J$ matrix $\mathbf{B}_k^*$ into $Jp \times 1$ vector $\boldsymbol{\beta}_k$ and $J \times 1$ vector $\boldsymbol{\psi}_k$ prior to back transforming. In the time-varying setting, we extract the $Jp \times 1$ vector $\boldsymbol{\beta}_k$ and the $J \times 1$ vector $\boldsymbol{\psi}_k = (\psi_{k1}, \ldots, \psi_{kJ})^T$ from $\boldsymbol{\beta}_k^*$, and use these to perform the back transformations.

6. (Optional) *Update latent scaling terms for extension to skew-t model as described in Section 3.3.*

(a) Time invariant designs:

i. Compute $s_1 = \frac{\xi+J+1}{2}$, where ξ is the pre-specified degrees of freedom parameter.

ii. For $i = 1, ..., n$, compute $s_{2i} = \frac{\xi + t_i^2 + (\mathbf{y}_i - \mathbf{B}_k^{*T}\mathbf{x}_{ki}^*)^T\boldsymbol{\Sigma}_k^{-1}(\mathbf{y}_i - \mathbf{B}_k^{*T}\mathbf{x}_{ki}^*)}{2}$, where $\mathbf{B}_k^*$ is the $(p+1) \times J$ parameter matrix defined in equation (6) of the manuscript and $\mathbf{x}_{ki}^*$ is a $(p+1) \times 1$ vector formed from the i-th row of $\mathbf{X}_k^*$ in (7).

iii. For $i = 1, ..., n$, update d_i from $\mathrm{Gamma}(s_1, s_{2i})$.

iv. For $k = 1, \ldots, K$, form the $n_k \times J$ scaled matrix $\tilde{\mathbf{Y}}_k = \sqrt{\mathbf{d}_k} \circ \mathbf{Y}_k$, where "$\circ$" denotes the Hadamard product. Use $\tilde{\mathbf{Y}}_k$ in place of $\mathbf{Y}_k$ for all remaining updates.

v. For $k = 1, \ldots, K$, form the $n_k \times (p+1)$ scaled matrix $\tilde{\mathbf{X}}_k = \sqrt{\mathbf{d}_k} \circ \mathbf{X}_k^*$. Use $\tilde{\mathbf{X}}_k$ in place of $\mathbf{X}_k$ for all remaining updates.

(b) Time-varying designs

i. Compute $s_1 = \frac{\xi+J+1}{2}$, where ξ is the pre-specified degrees of freedom parameter.

ii. For $i = 1, ..., n$, compute $s_{2i} = \frac{\xi + t_i^2 + (\mathbf{y}_i - \mathbf{X}_i^*\boldsymbol{\beta}_k^*)^T\boldsymbol{\Sigma}_k^{-1}(\mathbf{y}_i - \mathbf{X}_i^*\boldsymbol{\beta}_k^*)}{2}$, where $\mathbf{X}_i^*$ is the $J \times J(p+1)$ matrix and $\boldsymbol{\beta}_k^*$ is the $J(p+1)$ parameter vector each defined in Step 5(b) above.

iii. For $i = 1, ..., n$, update d_i from $\mathrm{Gamma}(s_1, s_{2i})$.

iv. Use d_i to scale the $J \times 1$ response vector $\mathbf{y}_i$ and the $J \times J(p+1)$ matrix $\mathbf{X}_i^*$ from Step 5(b). Next, for $k = 1, \ldots, K$, combine data for all subjects in cluster k to form the $Jn_k \times 1$ scaled vector $\tilde{\mathbf{y}}_k$ and $Jn_k \times J(p+1)$ scaled matrix $\tilde{\mathbf{X}}_k^*$. Use the scaled data for all remaining updates.

Web Appendix C: Web Tables

Web Table 1: Sample characteristics from the Nurture study ($n = 560$, $N = 1769$).

	n (%)
No. Missing Observations†	471 (21.0)
Missing 3 mo.	113 (20.2)
Missing 6 mo.	112 (20.00)
Missing 9 mo.	131 (23.4)
Missing 12 mo.	115 (20.5)
Food Insecure	216 (38.6)
Infant Gender (Female)	277 (49.5)
Infant Race (Black)	378 (67.5)
Any Breastfeeding	213 (38.0)
	Median (IQR)
Bayley composite score at 3 mo.	110.0 (15.0)
Bayley composite score at 6 mo.	103.0 (18.0)
Bayley composite score at 9 mo.	100.0 (19.0)
Bayley composite score at 12 mo.	97.0 (16.0)
	Mean (SD)
Birth weight for gestational age z-score	0.1 (1.0)
Total number of children in household	2.5 (1.5)

† Number missing out of $560 \times 4 = 2240$ possible observations.

Web Table 2: Estimated correlation matrix from repeated measures model with unstructured correlation.

	3 mo.	6 mo.	9 mo.	12 mo.
3 mo.	1.00			
6 mo.	0.23	1.00		
9 mo.	0.15	0.24	1.00	
12 mo.	0.15	0.25	0.27	1.00

Web Table 3: Model results for Simulation 1: Multivariate skew normal (MSN) and multivariate normal (MVN) models fit to MSN data with varying skewness settings, $n = 1000$, $J = 4$, $p = 2$, $K = 3$, and $r = 2$. Cluster 3 corresponds to data generated under a MVN model ($\boldsymbol{\alpha}_3 = \mathbf{0}$). 10000 iterations were run with a burn-in of 1000.

		Cluster 1			Cluster 2			Cluster 3		
Component	Param.	True	MSN Est. (95% CrI)	MVN Est. (95% CrI)	True	MSN Est. (95% CrI)	MVN Est. (95% CrI)	True	MSN Est. (95% CrI)	MVN Est. (95% CrI)
MVSN Regression	β_{k11}	110.00	110.20 (109.97, 110.41)	106.36 (105.97, 108.71)	90.00	90.17 (89.85, 90.44)	88.43 (88.05, 88.81)	100.00	100.02 (99.68, 100.70)	100.02 (99.82, 100.23)
	β_{k21}	115.00	115.13 (114.91, 115.33)	104.17 (103.93, 104.44)	85.00	85.31 (85.00, 85.58)	83.00 (82.57, 83.46)	100.00	100.25 (99.52, 100.73)	99.99 (99.79, 100.18)
	β_{k31}	120.00	120.08 (119.83, 120.49)	128.02 (128.57, 129.08)	80.00	80.23 (79.89, 80.54)	70.59 (70.08, 71.10)	100.00	100.13 (99.48, 100.77)	100.04 (99.83, 100.26)
	β_{k41}	125.00	125.15 (124.86, 125.49)	126.67 (126.31, 127.05)	75.00	74.94 (74.61, 75.26)	72.19 (71.64, 72.72)	100.00	99.81 (99.24, 100.40)	99.99 (99.78, 100.21)
	β_{k12}	1.00	0.97 (0.84, 1.11)	0.90 (0.74, 1.08)	-1.00	-1.08 (-1.25, -0.92)	-1.12 (-1.29, -0.93)	-1.00	-1.00 (-1.10, -0.89)	-1.00 (-1.10, -0.89)
	β_{k22}	1.50	1.51 (1.40, 1.62)	1.53 (1.41, 1.66)	-1.50	-1.51 (-1.73, -1.33)	-1.66 (-1.77, -1.52)	1.00	0.99 (0.89, 1.08)	0.98 (0.89, 1.08)
	β_{k32}	2.00	2.01 (1.89, 2.14)	2.20 (2.08, 2.33)	-2.00	-1.99 (-2.22, -1.78)	-2.44 (-2.67, -2.17)	-1.00	-0.92 (-1.01, -0.82)	-0.91 (-1.01, -0.81)
	β_{k42}	2.50	2.50 (2.35, 2.66)	2.46 (2.28, 2.64)	-2.50	-2.52 (-2.77, -2.29)	-2.68 (-2.92, -2.39)	1.00	1.04 (0.94, 1.15)	1.05 (0.95, 1.15)
	Σ_{k11}	1.00	0.96 (0.77, 1.14)	2.42 (2.06, 2.84)	1.00	0.96 (0.69, 1.48)	2.53 (2.17, 3.01)	1.00	0.98 (0.80, 1.14)	1.01 (0.88, 1.15)
	Σ_{k12}	0.50	0.47 (0.34, 0.61)	1.20 (0.99, 1.48)	0.50	0.50 (0.27, 0.99)	2.51 (2.14, 3.04)	0.50	0.51 (0.21, 0.71)	0.51 (0.41, 0.61)
	Σ_{k13}	0.25	0.25 (0.04, 0.40)	-0.54 (-0.75, -0.34)	0.25	0.26 (0.13, 0.72)	2.62 (2.20, 3.17)	0.25	0.25 (0.14, 0.37)	0.25 (0.15, 0.36)
	Σ_{k14}	0.12	0.11 (-0.02, 0.30)	-1.35 (-1.67, -1.06)	0.12	0.15 (-0.06, 0.67)	2.72 (2.24, 3.29)	0.12	0.09 (-0.12, 0.29)	0.09 (-0.01, 0.20)
	Σ_{k22}	1.00	0.99 (0.74, 1.19)	1.20 (0.99, 1.48)	1.00	1.03 (0.78, 1.54)	2.51 (2.14, 3.04)	1.00	0.99 (0.71, 1.24)	0.91 (0.80, 1.05)
	Σ_{k23}	0.50	0.49 (0.26, 0.66)	1.24 (1.06, 1.46)	0.50	0.59 (0.38, 1.03)	3.69 (3.18, 4.35)	0.50	0.56 (0.37, 0.71)	0.44 (0.34, 0.54)
	Σ_{k24}	0.25	0.24 (0.10, 0.43)	0.08 (-0.06, 0.21)	0.25	0.28 (-0.01, 0.61)	3.65 (3.09, 4.31)	0.25	0.25 (0.14, 0.37)	0.27 (0.17, 0.37)
	Σ_{k33}	1.00	0.99 (0.77, 1.09)	1.24 (1.06, 1.46)	1.00	1.09 (0.88, 1.59)	3.65 (3.03, 4.32)	1.00	0.95 (0.80, 1.10)	1.00 (0.87, 1.15)
	Σ_{k34}	0.50	0.47 (0.22, 0.65)	1.15 (0.93, 1.40)	0.50	0.54 (0.25, 0.99)	2.62 (2.20, 3.17)	0.50	0.57 (0.39, 0.73)	0.56 (0.45, 0.70)
	Σ_{k44}	1.00	1.01 (0.63, 1.23)	2.48 (2.15, 2.91)	1.00	1.02 (0.64, 1.60)	3.65 (3.09, 4.31)	1.00	1.06 (0.81, 1.60)	1.06 (0.94, 1.23)
	α_{k1}	-2.00	-2.05 (-2.28, -1.66)	/	-2.00	-2.19 (-2.50, -1.77)	/	0.00	-0.23 (-0.80, 0.42)	/
	α_{k2}	-1.00	-1.01 (-1.30, -0.75)	/	-2.50	-2.52 (-2.82, -2.10)	/	0.00	-0.33 (-0.94, 0.57)	/
	α_{k3}	1.00	0.97 (0.65, 1.28)	/	-3.00	-3.34 (-3.67, -2.90)	/	0.00	-0.12 (-0.93, 0.68)	/
	α_{k4}	2.00	1.97 (1.67, 2.28)	/	-3.50	-3.49 (-3.84, -3.00)	/	0.00	0.23 (-0.51, 0.94)	/
Multinomial Logit[†]	δ_{k1}	-0.27	-0.23 (-0.47, -0.09)	-0.14 (-0.35, 0.08)	0.14	0.12 (0.01, 0.21)	0.20 (0.00, 0.42)	Ref.	Ref.	Ref.
	δ_{k2}	0.07	0.07 (-0.26, 0.39)	0.08 (-0.24, 0.38)	0.17	0.16 (0.01, 0.38)	0.02 (-0.28, 0.30)	Ref.	Ref.	Ref.
Missing Data	γ_{k1}	-0.82	-0.84 (-0.96, -0.73)	-1.08 (-1.19, -0.99)	-0.93	-0.93 (-1.05, -0.81)	-1.02 (-1.15, -0.93)	-1.19	-1.22 (-1.39, -1.10)	-0.78 (-0.91, -0.67)
	γ_{k2}	-1.08	-1.01 (-1.20, -0.91)	-1.80 (-1.96, -1.64)	-1.14	-1.11 (-1.25, -1.00)	-0.72 (-0.80, -0.59)	-0.97	-0.93 (-1.10, -0.79)	-1.09 (-1.22, -0.98)
	γ_{k3}	-1.12	-1.08 (-1.20, -1.00)	-0.90 (-1.00, -0.80)	-0.98	-0.99 (-1.12, -0.85)	-1.04 (-1.16, -0.90)	-0.87	-0.88 (-1.02, -0.76)	-0.97 (-1.09, -0.86)
	σ_k^2	1.00	1.07 (0.92, 1.28)	0.89 (0.76, 1.07)	1.00	0.96 (0.83, 1.11)	1.21 (1.05, 1.41)	1.00	1.11 (0.96, 1.30)	0.91 (0.80, 1.05)
Clustering[‡]	π_l	0.32	0.32 (0.31, 0.33)	0.32 (0.30, 0.34)	0.29	0.29 (0.28, 0.30)	0.29 (0.28, 0.30)	0.39	0.39 (0.39, 0.39)	0.39 (0.39, 0.39)

† Multinomial logit parameters comparing clusters 1 and 2 to cluster 3 (reference cluster).

‡ Estimated proportion in each cluster. True proportions are 0.32, 0.29 and 0.39, respectively.

Web Table 4: Results for clusters 2 and 3 from Simulation 2. Posterior means (95% CrIs) are presented for conditional ignorability and marginal ignorability as described in Section 4.2 of the manuscript. 10000 iterations were run with a burn-in of 1000.

Model Component	**Parameter**	**True Value**	**Conditional Ignorability**	**Marginal Ignorability**
$\boldsymbol{k=2}$	β_{k11}	0.23	0.09 (-0.36, 0.56)	0.41 (-0.87, 2.03)
	β_{k21}	-0.52	-0.57 (-0.80, -0.30)	-0.77 (-1.88, 0.67)
MSN	β_{k31}	-0.33	-0.29 (-0.76, 0.23)	-0.54 (-1.58, -0.45)
Regression	β_{k41}	-1.54	-1.51 (-1.73, -1.25)	-2.13 (-2.44, -1.83)
	β_{k12}	-0.15	-0.19 (-0.52, -0.04)	-0.29 (-0.71, 0.89)
	β_{k22}	-0.22	-0.37 (-0.57, -0.15)	-0.85 (-1.91, 0.74)
	β_{k32}	0.64	0.55 (0.03, 1.08)	1.82 (1.33, 2.29)
	β_{k42}	0.86	0.77 (0.55, 1.01)	0.85 (0.63, 1.09)
	α_{k1}	2.00	2.07 (1.51, 2.57)	1.82 (1.14, 2.54)
	α_{k2}	2.00	1.90 (1.27, 2.42)	2.14 (1.88, 2.31)
	α_{k3}	2.00	1.92 (1.21, 2.71)	2.21 (1.03, 3.48)
	α_{k4}	2.00	2.04 (1.39, 2.62)	1.50 (0.89, 2.17)
Multinomial	δ_{k1}	-0.12	-0.14 (-0.33, -0.06)	-0.13 (-0.34, 0.06)
Logit†	δ_{k2}	-0.02	-0.06 (-0.30, 0.26)	-0.02 (-0.31, 0.26)
Missing Data	γ_{k1}	-0.99	-1.08 (-1.51, -0.67)	/
	γ_{k2}	-0.86	-0.80 (-1.15, -0.46)	/
	γ_{k3}	-1.09	-1.18 (-1.55, -0.86)	/
	σ_k^2	1.00	1.03 (0.90, 1.17)	/
$\boldsymbol{k=3}$	β_{k11}	3.22	3.46 (3.09, 3.76)	1.11 (0.58, 1.63)
	β_{k21}	4.51	4.37 (4.18, 4.55)	4.33 (4.09, 4.57)
MSN	β_{k31}	3.21	3.36 (2.99, 3.67)	1.19 (0.69, 1.69)
Regression	β_{k41}	4.14	3.96 (3.77, 4.13)	3.91 (3.68, 4.14)
	β_{k12}	3.14	3.44 (3.08, 3.75)	1.15 (0.63, 1.66)
	β_{k22}	3.37	3.24 (3.05, 3.41)	3.16 (2.93, 3.39)
	β_{k32}	3.02	3.34 (2.97, 3.66)	0.94 (0.40, 1.48)
	β_{k42}	3.55	3.41 (3.21, 3.59)	3.34 (3.09, 3.59)
	α_{k1}	-3.00	-3.26 (-3.68, -2.79)	-3.34 (-4.12, 1.97)
	α_{k2}	-3.00	-3.02 (-3.48, -2.56)	-3.11 (-4.63,-1.91)
	α_{k3}	-3.00	-3.22 (-3.65, -2.72)	-3.71 (-3.01, -4.20)
	α_{k4}	-3.00	-3.26 (-3.69, -2.83)	-3.23 (-4.01, -2.76)
Multinomial	δ_{k1}	Ref.	Ref.	Ref.
Logit	δ_{k2}	Ref.	Ref.	Ref.
Missing Data	γ_{k1}	-1.17	-1.20 (-1.57, -0.85)	/
	γ_{k2}	-0.89	-0.89 (-1.18, -0.64)	/
	γ_{k3}	-0.65	-0.65 (-0.95, -0.39)	/
	σ_k^2	1.00	1.15 (0.97, 1.38)	/

† Multinomial logit parameters comparing cluster 2 to cluster 3 (reference cluster).

Web Table 5: Simulation 3 WAIC values for MSN models fit with $K = 2, 3, 4, 5$ clusters to data simulated from MSN models with $K = 2, 3, 4, 5$. Bold indicates best-fitting model. (*) Model did not converge due to empty or singleton clusters occurring within 10000 iterations of the MCMC algorithm.

		Fitted			
		$K = 2$	$K = 3$	$K = 4$	$K = 5$
	$K = 2$	**11624**	11963	*	*
Truth	$K = 3$	15193	**12390**	12811	*
	$K = 4$	15777	14359	**12412**	14237
	$K = 5$	15012	14359	14323	**13436**

Web Table 6: Estimated covariances (95% CrI), $\mathbf{\Sigma}_1, \mathbf{\Sigma}_2$, from the 2-cluster MSN model fit to the Nurture data as described in Section 5.

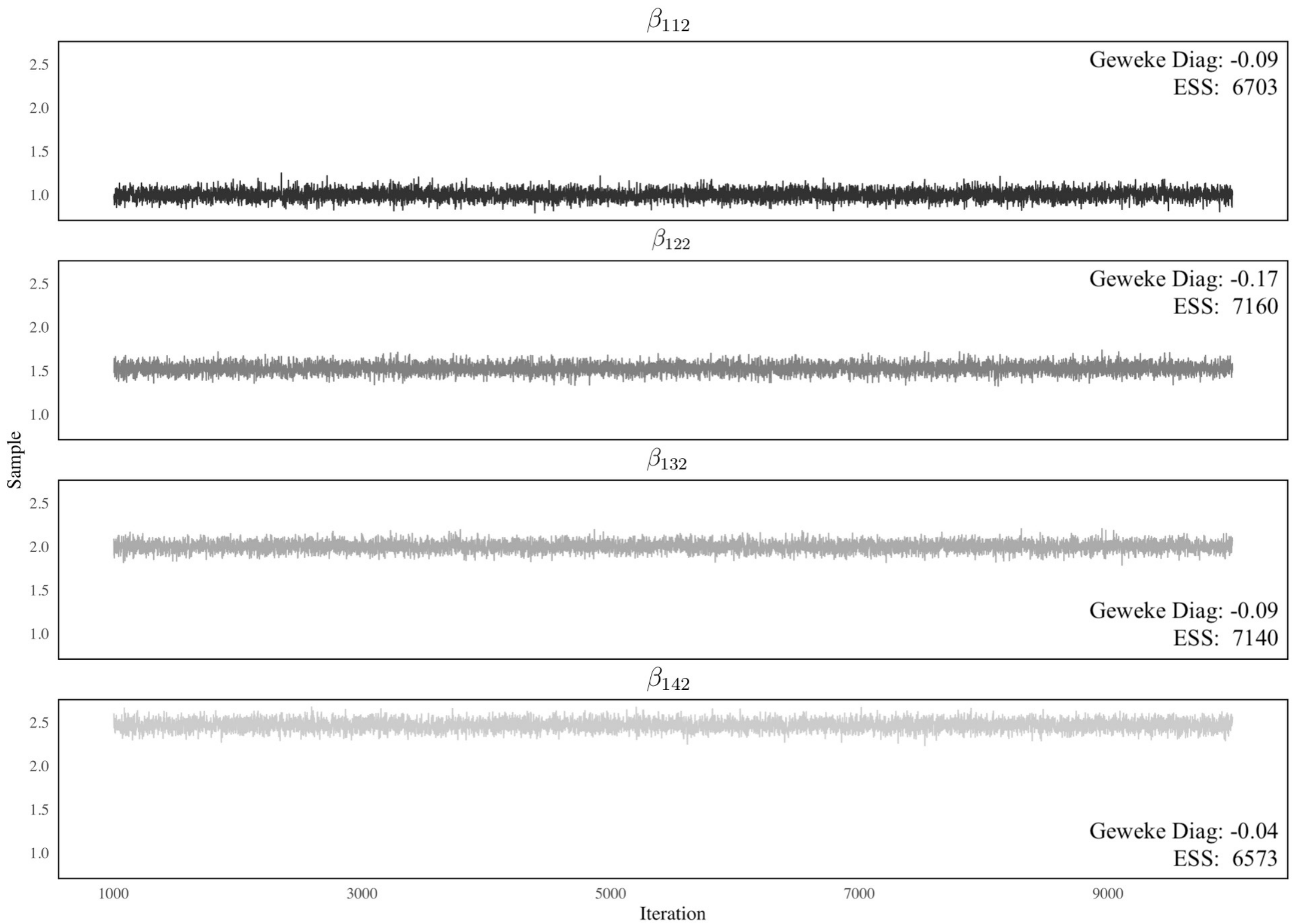
k = 1	3 mo.	6 mo.	9 mo.	12 mo.
3 mo.	0.41 (0.36, 0.49)			
6 mo.	0.38 (0.32, 0.44)	0.46 (0.40, 0.54)		
9 mo.	0.38 (0.32, 0.44)	0.36 (0.31, 0.43)	0.43 (0.37, 0.50)	
12 mo.	0.34 (0.29, 0.40)	0.35 (0.30, 0.40)	0.32 (0.12, 0.54)	0.52 (0.44, 0.61)
k = 2	3 mo.	6 mo.	9 mo.	12 mo.
3 mo.	1.18 (1.01, 1.39)			
6 mo.	0.75 (0.56, 0.95)	1.26 (1.11, 1.44)		
9 mo.	0.94 (0.77, 1.11)	0.82 (0.68, 0.99)	1.33 (1.16, 1.53)	
12 mo.	0.67 (0.52, 0.83)	0.80 (0.67, 0.95)	0.88 (0.74, 1.04)	1.27 (1.12, 1.45)

Web Table 7: Estimated correlations from the 2-cluster MSN model fit to the Nurture data as described in Section 5.

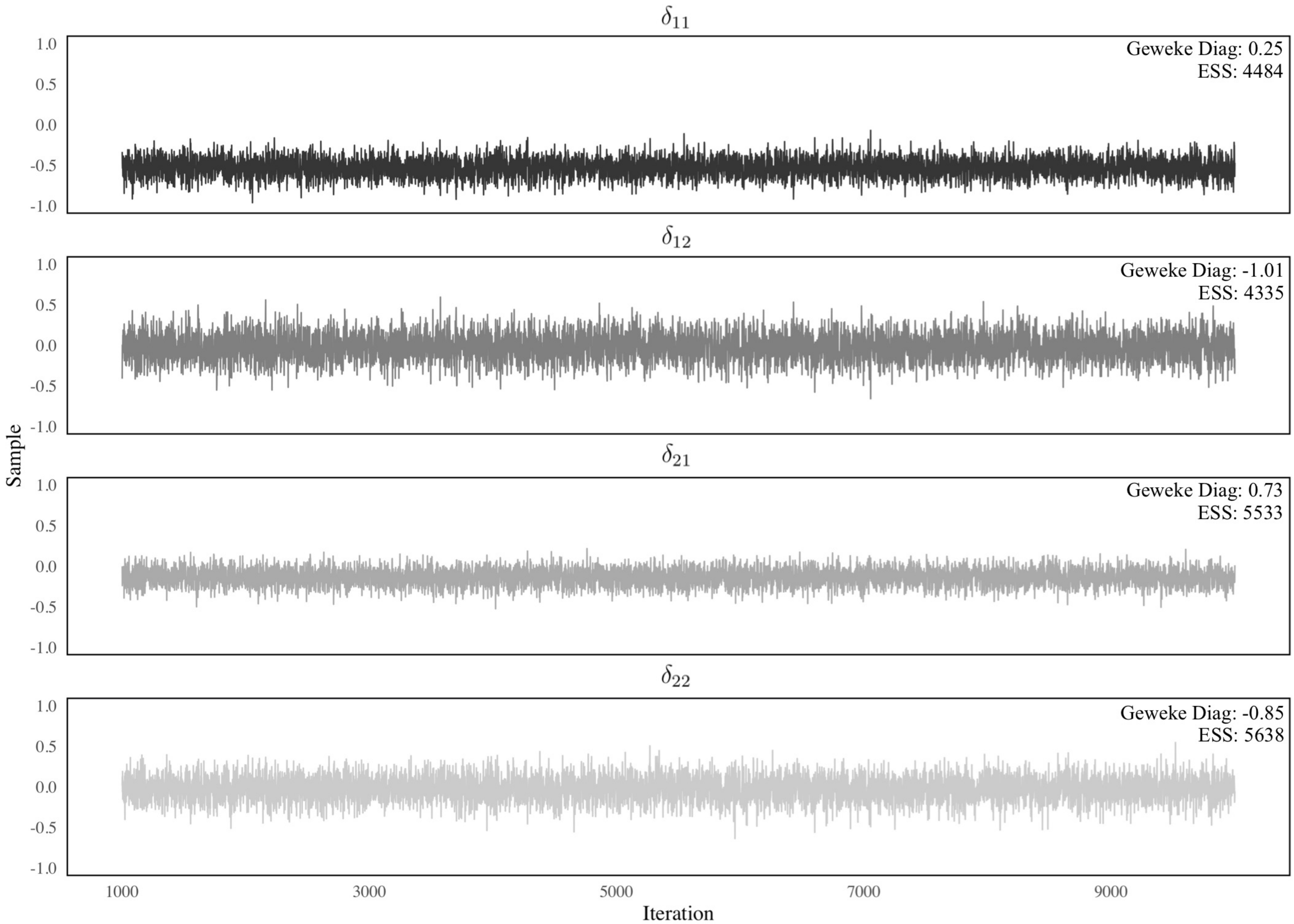
k = 1	3 mo.	6 mo.	9 mo.	12 mo.
3 mo.	1.00			
6 mo.	0.88	1.00		
9 mo.	0.90	0.81	1.00	
12 mo.	0.74	0.72	0.68	1.00
k = 2	3 mo.	6 mo.	9 mo.	12 mo.
3 mo.	1.00			
6 mo.	0.62	1.00		
9 mo.	0.75	0.63	1.00	
12 mo.	0.55	0.63	0.68	1.00

Web Appendix E: Web Figures

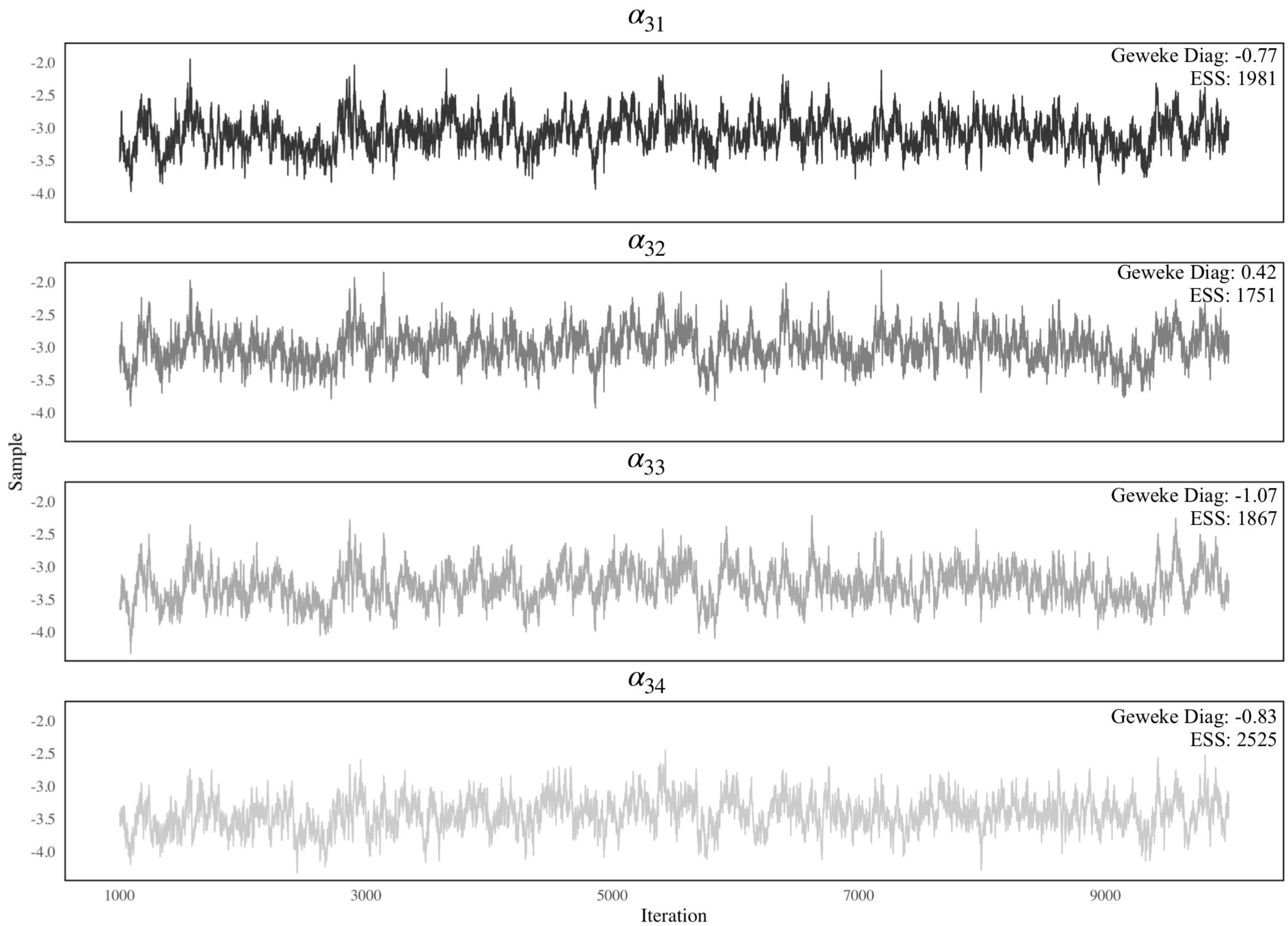
Web Figure 1: Trace plots of a selection of parameters from Simulation 1. Geweke diagnostics and effective sample sizes (ESS) are shown for each parameter. MCMC sampling was run for 10000 iterations with a burn-in of 1000. All parameters were initialized at 0 and prior parameters were chosen to be weakly informative.



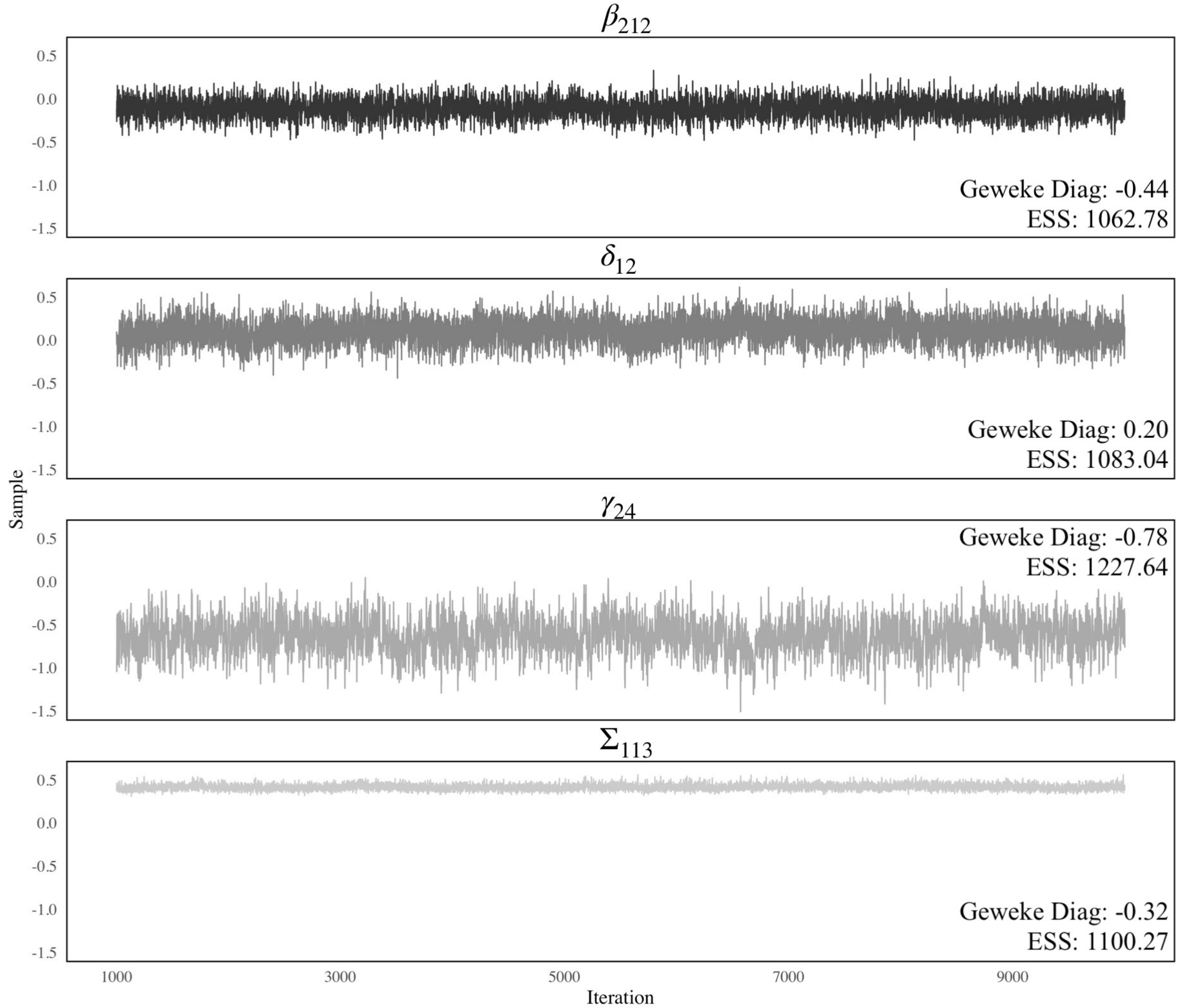
Web Figure 2: Trace plots of a selection of parameters from the MNAR imputation model in Simulation 2. Geweke diagnostics and effective sample sizes (ESS) are shown for each parameter. MCMC sampling was run for 10000 iterations with a burn-in of 1000. All parameters were initialized at 0 and prior parameters were chosen to be weakly informative.



Web Figure 3: Trace plots of a selection of parameters from the 3-class model in Simulation 3. Geweke diagnostics and effective sample sizes (ESS) are shown for each parameter. MCMC sampling was run for 10000 iterations with a burn-in of 1000. All parameters were initialized at 0 and prior parameters were chosen to be weakly informative.



Web Figure 4: Trace plots of a selection of parameters from the application to the Nurture Data. Geweke diagnostics and effective sample sizes (ESS) are shown for each parameter. MCMC sampling was run for 10000 iterations with a burn-in of 1000. All parameters were initialized at 0 and prior parameters were chosen to be weakly informative.



References:

1. Laurent, S. (2018). Matrixsampling: Simulations of Matrix Variate Distributions.
2. Geweke, J. (1992). Evaluating the accuracy of sampling-based approaches to calculating posterior moments. In *Bayesian Statistics 4 (ed JM Bernado, JO Berger, AP Dawid and AFM Smith)*. Clarendon Press, Oxford, UK.