HHS Public Access

Author manuscript

Biometrics. Author manuscript; available in PMC 2021 July 22.

Published in final edited form as:

Biometrics. 2021 June ; 77(2): 675–688. doi:10.1111/biom.13328.

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

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

[1]Department of Biomedical Informatics, The Ohio State University, Columbus, Ohio [2]Department of Health, Behavior and Society, Johns Hopkins University, Baltimore, Maryland [3]Department of Public Health Sciences, Medical University of South Carolina, Charleston, South Carolina

Abstract

In studies of infant growth, an important research goal is to identify latent clusters of infants with delayed motor development—a risk factor for adverse outcomes later in life. However, there are numerous statistical challenges in modeling motor development: the data are typically skewed, exhibit intermittent missingness, and are correlated across repeated measurements over time. Using data from the Nurture study, a cohort of approximately 600 mother-infant pairs, we develop a flexible Bayesian mixture model for the analysis of infant motor development. First, we model developmental trajectories using matrix skew-normal distributions with cluster-specific parameters to accommodate dependence and skewness in the data. Second, we model the cluster-membership probabilities using a Pólya-Gamma data-augmentation scheme, which improves predictions of the cluster-membership allocations. Lastly, we impute missing responses from conditional multivariate skew-normal distributions. Bayesian inference is achieved through straightforward Gibbs sampling. Through simulation studies, we show that the proposed model yields improved inferences over models that ignore skewness or adopt conventional imputation methods. We applied the model to the Nurture data and identified two distinct developmental clusters, as well as detrimental effects of food insecurity on motor development. These findings can aid investigators in targeting interventions during this critical early-life developmental window.

Keywords

conditional ignorability; food security; intermittent missing; matrix skew-normal; motor development; Pólya-Gamma distribution

Correspondence Dr. Brian Neelon, Department of Public Health Sciences, Medical University of South Carolina, Charleston, South Carolina. neelon@musc.edu.

SUPPORTING INFORMATION

Web Appendices, Tables, and Figures referenced in Sections 2-5 are available with this paper at the Biometrics website on Wiley Online Library. An R package BayesMSN for implementing these methods is available at https://github.com/carter-allen/BayesMSN and through the Biometrics website on Wiley Online Library.

1 | INTRODUCTION

Infant motor development is an important predictor of health later in life. Early motor development is associated with improved physical activity, cognitive function, and educational attainment (Taanila et al., 2005; Aaltonen et al., 2015), while delayed development is associated with increased sedentary time (Sánchez et al., 2017) and has been linked to adult cognitive disorders such as schizophrenia (Filatova et al., 2017). Thus, there is growing interest in identifying developmental patterns that may place infants at risk for long-term adverse health outcomes. One approach to tackling this problem is to identify underlying subgroups of infants with delayed motor development, and to isolate important predictors of subgroup membership. Our goal, therefore, is to introduce a flexible latent growth mixture model to detect high-risk developmental patterns and associated risk factors.

Our work is motivated by the Nurture study, a birth cohort of predominately black women and their infants residing in the southeast United States (Benjamin Neelon et al., 2017). The aim of the study was to examine how infant feeding, physical activity, motor development, sleep, and stress contribute to infant weight gain. The second aim was to identify infant subpopulations that exhibit unique motor development trajectories, and to examine cluster-specific associations between household food security and motor development.

The Nurture data pose several statistical challenges. First, the repeated outcomes are correlated across measurement occasions, and the pairwise correlations vary across timepoints, suggesting the need for a flexible error term covariance structure. Second, the development outcomes are skewed, with the direction of skewness varying over time. The Nurture data also feature intermittent missingness. Thus, we require a framework capable of addressing potentially nonignorable missing data. Finally, we seek to develop a model that incorporates covariate information into both the multivariate regression model of infant development trajectories and the clustering model.

To address these challenges, we propose a Bayesian multivariate mixture model for the analysis of longitudinal skewed infant motor development data with intermittent missing observations. Our approach builds on recent work on mixture models for skewed cross-sectional data. Frühwirth-Schnatter and Pyne (2010) proposed a multivariate skew-normal (MSN) model for high-dimensional flow cytometric data. However, their focus was on marginal inference (ie, density estimation) rather than cluster-specific inferences, as is our focus here. More recently, Lin et al. (2018) proposed a mixture of skew-*t* factor analyzers for settings in which cluster-specific inference is of primary interest (Lin et al., 2018). However, like Frühwirth-Schnatter and Pyne (2010), their approach excluded covariates in the cluster-membership model, a focal point in our study as we expect demographics to not only play a key role in predicting cluster membership, but also help characterize developmental trajectories within clusters. Additionally, their approach, while quite flexible, relied on a computationally elaborate expectation-conditional maximization algorithm that does not enjoy the inferential benefits of a Bayesian approach. Finally, the authors adopted a single-imputation scheme for ignorable missing data that does not readily account for the uncertainty in the imputation process without additional multiple imputation steps.

Our proposed model extends these prior studies in a number of ways. First, our model enables cluster-specific inferences for longitudinal growth trajectories, while accommodating skewness patterns that may vary over time and across clusters. Second, our model accommodates both time-dependent and time-invariant covariate designs. Third, we estimate parameters in a Bayesian framework that introduces covariates into the cluster-membership model using a novel application of Pólya-Gamma data augmentation (Polson et al., 2013). Fourth, we accommodate intermittent missingness of longitudinal responses under a "conditional ignorability" assumption, whereby the missing data mechanism is assumed to be ignorable conditional on cluster assignment. Marginally, we allow for dependence between the missing data mechanism and the missing responses, thus relaxing standard missing at random (MAR) assumptions. We develop a Markov chain Monte Carlo (MCMC) embedded imputation procedure in which missing observations are updated at each MCMC iteration conditional on cluster allocation. Finally, we propose a Bayesian modeling approach that makes use of convenient matrix skew-normal and skew-t representations. Our model is appropriate for settings where interest lies in identifying clusters in longitudinal data with complex features, such as skewness, heavy tails, and intermittent missing responses that are potentially missing not at random.

2 | NURTURE STUDY

The Nurture study is a birth cohort of predominately black women and their infants residing in the southeastern United States from 2013 and 2017 (Benjamin Neelon et al., 2017). The study followed mothers and infants for 12 months after birth and collected data on infant gross motor development and household food security, among other measures. Infant development was assessed quarterly at 3, 6, 9, and 12 months of age using the Bayley composite scale of motor development (Bayley, 2006), a standard measure of infant development ranging from 40 to 160, with higher scores indicating more advanced development compared to normally developing infants. Household food security was assessed using the 18-item US Household Food Security Survey Module restricted to the 10 items related to household food security measured during pregnancy (USDA, 2019). Following standard protocol, a final dichotomous food security exposure was defined as "food insecure" households and "food secure" households. The Institutional Review Board of Duke University Medical Center approved this study and protocol.

Of the 666 infants who were consented into the study, 106 were missing Bayley score measurements at all timepoints, 68 infants were missing Bayley scores at three timepoints, 72 infants were missing Bayley scores at two timepoints, 123 were missing Bayley scores at one time-point, and 297 were not missing any Bayley scores. We restricted our analytic sample to the 560 remaining infants who had at least one nonmissing Bayley score over the study period. Of the $560 \times 4 = 2240$ possible observations, 471 (21%) were missing, leaving an available-case sample size of 1769. Sample characteristics for the 560 participants are given in Web Table 1. In the sample, 68% of infants were black and 39% of households identified as food insecure during pregnancy. The Bayley motor development scores ranged from 49.0 to 145.0 across visits, with a mean of 102.4 and standard deviation (SD) of 13.5.

Figure 1 presents trajectory plots of the motor development scores for each infant in the available-case sample, with an overlay of the mean score at each visit. The plot indicates substantial heterogeneity in the trajectories. To quantify the mean trend, we fit a repeated-measures model of the form: $\mathbf{Y}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{e}_i$, where $\mathbf{X}_i$ includes an intercept, a linear time trend and effects for gender, race, and baseline food security status; and $\mathbf{e}_i$ is a multivariate normal error term with unstructured covariance pattern. The restricted maximum likelihood estimate of the linear trend coefficient was −1.16 (CI = [−1.29, −1.03]), suggesting an average decline in motor development over time in the Nurture cohort relative to normally developing infants. However, because most of the literature on infant motor development has focused on the average effect over time (Shoaibi et al., 2019), little is known about trends for specific subgroups of interest—for example, among infants who may be at high risk for delayed motor milestone achievement. Importantly, these subgroups may not be obvious from marginal trajectory plots such as Figure 1 and may only become evident through appropriate modeling of germane features of the data such as skewness, missingness, and explanatory covariates, among other factors. In this paper, we present methods for uncovering latent subgroups by modeling these important features of the data.

Figure 2 presents centered and scaled residual densities from the repeated-measures model used in Figure 1. The residuals were subset by visit to yield visit-specific residual density plots. As shown in Figure 2, the residuals are skewed at each visit, particularly at 3 and 6 months, with the direction of skewness varying over time. Shapiro-Wilk tests accounting for multiple testing rejected the null hypothesis of normality at 6 months, contravening standard assumptions. While there is a modest indication of skewness in the available-case sample, it is not clear how skewness patterns vary across latent subgroups of infants, or how missing observations impact skewness. We seek to answer these questions in subsequent analyses.

Additionally, the motor development scores are correlated over time, with pairwise correlations ranging in an unstructured pattern. As an illustration, we fit three repeated-measures models of the form used in Figure 1, but with varying correlation structures for the errors: AR1, compound symmetric and unstructured. The AIC values for these models were 27 599, 27 517, and 27 478, respectively, indicating best fit under the unstructured pattern among the patterns considered. We present the estimated correlation matrix from this model in Web Table 2. Finally, the Nurture data feature intermittent missing data, with approximately one third of the sample missing observations at any given visit (Web Table 1). While it may be reasonable to assume that the missing data are MAR, as we have no a priori reason to believe that the occurrence of missing observations is directly related to missing Bayley scores, we relax this assumption below by assuming ignorable missingness conditional on latent motor development cluster assignment.

3 | MODEL

In Section 3, we develop a model that accounts for the important features of the Nurture data described in Section 2. Section 3.1 begins with developing a finite mixture model and proposes a MSN regression framework for within-cluster inference. Section 3.2 proposes a multinomial regression model for cluster probabilities that utilizes Pólya-Gamma data augmentation for efficient Gibbs sampling. Section 3.3 discusses extensions to the

multivariate skew-t (MST) setting, and Section 3.4 proposes a missing data imputation scheme under the assumption of conditional ignorability.

3.1 | Multivariate skew-normal mixture model

We propose a finite mixture model that accommodates relevant features of the data, namely skewness, missing values, and dependence among the responses. While alternative mixture models (eg, Dirichlet process mixtures) provide flexibility for marginal inferences and density estimation, finite mixtures are appealing when the focus is on practical within-cluster inferences. In such cases, the primary goal is to identify a small number of clinically relevant clusters to help design targeted interventions to improve health outcomes. However, to avoid misspecifying the number of finite mixtures, it is imperative to properly model the within-cluster distributions by accounting for important features, such as skewness or heavy tails. With this goal in mind, we present a repeated-measures regression model based on a MSN distribution—and by extension, a MST distribution—in which the Bayley scores across the J measurement occasions represent correlated responses. Specifically, let $\mathbf{y}_i = (y_{i1}, \dots, y_{iJ})^T$ be a $J \times 1$ vector of standardized Bayley scores for subject i ($i = 1, \dots, n$). We propose a mixture model of the form

$$f(\mathbf{y}_i) = \sum_{k=1}^{K} \pi_{ki} f(\mathbf{y}_i \mid \theta_k), \tag{1}$$

where $\boldsymbol{\theta}_k$ is the set of parameters specific to cluster k ($k = 1, \dots, K$) and $\boldsymbol{\pi}_{ki}$ is a subject-specific mixing weight representing the probability that subject i belongs to cluster k. For now we assume that K is fixed; we discuss model-selection strategies for choosing the optimal value of K in Section 3.5.2.

For posterior inference, we introduce a latent cluster indicator variable z_i taking the value $k \in \{1, \dots, K\}$ with probability $\boldsymbol{\pi}_{ki}$. Given $z_i = k$, we assume $\mathbf{y}_i$ is distributed according to a J-dimensional MSN density (Azzalini and Valle, 1996)

$$\begin{aligned} &\mathbf{y}_i \mid (z_i = k) \sim \text{MSN}_J(\boldsymbol{\zeta}_{ki}, \boldsymbol{\alpha}_k, \boldsymbol{\Omega}_k), \text{ with density} \\ &f(\mathbf{y}_i \mid z_i = k) = 2\phi_J(\mathbf{y}_i; \boldsymbol{\zeta}_{ki}, \boldsymbol{\Omega}_k)\Phi\{\boldsymbol{\alpha}_k^T(\mathbf{y}_i - \boldsymbol{\zeta}_{ki})\}, \end{aligned} \tag{2}$$

where $\phi_J(\mathbf{y}_i; \boldsymbol{\zeta}_{ki}, \boldsymbol{\Omega}_k)$ denotes a J-dimensional normal density with mean $\boldsymbol{\zeta}_{ki}$ and covariance matrix $\boldsymbol{\Omega}_k$; $\Phi(\cdot)$ is the CDF of a scalar standard normal random variable; $\boldsymbol{\zeta}_{ki}$ is a $J \times 1$ vector of subject- and cluster-specific location parameters; $\boldsymbol{\alpha}_k$ is a $J \times 1$ vector of cluster-specific parameters that control the skewness of each outcome in cluster k; and $\boldsymbol{\Omega}_k$ is a $J \times J$ cluster-specific scale matrix that captures dependence among the J responses for subject i. When $\boldsymbol{\alpha}_k = \mathbf{0}$, the MSN distribution reduces to the multivariate normal (MVN) distribution $\text{N}_J(\boldsymbol{\zeta}_{ki}, \boldsymbol{\Omega}_k)$, where $\boldsymbol{\zeta}_{ki}$ represents a $J \times 1$ mean vector and $\boldsymbol{\Omega}_k$ is a $J \times J$ unstructured covariance matrix.

We complete model (2) by incorporating covariates into $\boldsymbol{\zeta}_{ki}$. We first discuss the general case in which the model includes both time-varying and time-invariant predictors; later, we present simplifications when only time-invariant covariates are included in the model. Here,

we adopt a convenient conditional representation of the MSN density (Azzalini and Valle, 1996; Frühwirth-Schnatter and Pyne, 2010):

$$\mathbf{y}_i \mid (z_i = k, t_i) = \mathbf{X}_i\boldsymbol{\beta}_k + t_i\boldsymbol{\psi}_k + \epsilon_i, \tag{3}$$

where $\mathbf{X}_i$ is a $J \times Jp$ design matrix that includes potential time-dependent covariates; $\boldsymbol{\beta}_k = (\beta_{k11}, \ldots, \beta_{k1p}, \ldots, \beta_{kJ1}, \ldots, \beta_{kJp})^T$ is a $Jp \times 1$ vector of cluster- and outcome-specific regression coefficients; $t_i \sim \mathrm{N}_{[0,\infty)}(0,1)$ is a subject-specific standard normal random variable truncated below by zero; $\boldsymbol{\psi}_k = (\psi_{k1}, \ldots, \psi_{kJ})^T$ is a $J \times 1$ vector of cluster-specific parameters that control skewness; and $\boldsymbol{\epsilon}_i \mid (z_i = k) \sim \mathrm{N}_J(\mathbf{0}, \sum_k)$ is a $J \times 1$ vector of correlated error terms. Thus, conditional on t_i and $z_i = k$, $\mathbf{y}_i$ is distributed as $\mathrm{N}_J(\mathbf{X}_i\boldsymbol{\beta}_k + t_i\boldsymbol{\psi}_k, \sum_k)$. Marginally (integrated over t_i), $\mathbf{y}_i \mid (z_i = k)$ is distributed $\mathrm{MSN}_J(\boldsymbol{\zeta}_{ki}, \boldsymbol{\alpha}_k, \boldsymbol{\Omega}_k)$, where through back-transformation the parameters $\boldsymbol{\zeta}_{ki}$, $\boldsymbol{\Omega}_k$, and $\boldsymbol{\alpha}_k$ can be obtained as described in Web Appendix B.

As detailed in Web Appendix B, conjugate full conditionals are available for all parameters in model (3), leading to straightforward Gibbs sampling when both time-varying and time-invariant covariates are included in the model. However, the Nurture analysis described in Section 5 involves no time-varying covariates, only time-varying covariate effects. In such cases, we can express the MSN density more compactly using a matrix skew-normal (MatSN) representation. Structuring the data in this way greatly facilitates posterior computation by permitting low-dimensional matrix updates for the regression coefficients. For cluster k, let $\mathbf{Y}_k$ be an $n_k \times J$ matrix with rows $\mathbf{y}_i^T$ for $i = 1, \ldots, n_k$, where n_k is the number of subjects in cluster k. From Equation (2), it follows that $\mathbf{Y}_k$ is distributed as

$$\mathbf{Y}_k \mid \mathbf{B}_k, \boldsymbol{\alpha}_k, \boldsymbol{\Omega}_k \sim \mathrm{MatSN}_{n_k \times J}(\mathbf{X}_k\mathbf{B}_k, \boldsymbol{\alpha}_k, \mathbf{I}_{n_k}, \boldsymbol{\Omega}_k), \tag{4}$$

where $\mathbf{I}_{n_k}$ is the $n_k \times n_k$ identity matrix, and $\mathbf{X}_k$ and $\mathbf{B}_k$ are, respectively, $n_k \times p$ and $p \times J$ matrices described in Web Appendix B.

If we set $x_{i1} = 1$ for all i, then the first row of $\mathbf{B}_k$, $(\beta_{k11}, \ldots, \beta_{k1J})$, represents time-specific intercepts that capture the time trend for the reference covariate group in cluster k. Adapting Equation (7) from Chen and Gupta (2005), the density function for $\mathbf{Y}_k$ is

$$\begin{aligned} &f(\mathbf{Y}_k \mid \mathbf{B}_k, \boldsymbol{\alpha}_k, \boldsymbol{\Omega}_k) \\ &\quad = 2^{n_k}\phi_{n_k \times J}(\mathbf{Y}_k; \mathbf{X}_k\mathbf{B}_k, \mathbf{I}_{n_k}, \boldsymbol{\Omega}_k)\Phi_{n_k}\{(\mathbf{Y}_k - \mathbf{X}_k\mathbf{B}_k)\boldsymbol{\alpha}_k\}, \end{aligned} \tag{5}$$

where $\phi_{n_k \times J}(\mathbf{Y}_k; \mathbf{X}_k\mathbf{B}_k, \mathbf{I}_{n_k}, \boldsymbol{\Omega}_k)$ is the density function for a matrix normal (MatNorm) random variable of dimension $n_k \times J$ with mean $\mathbf{X}_k\mathbf{B}_k$ and scale matrices $\mathbf{I}_{n_k}$ and $\boldsymbol{\Omega}_k$, and $\Phi_{n_k}(\cdot)$ denotes the CDF of an n_k-dimensional standard MVN random variable.

Further, let $\mathbf{t}_k = (t_1, \ldots, t_{n_k})^T$ denote the $n_k \times 1$ vector of latent variables for cluster k. By extending Equation (3), it follows that the conditional distribution of $\mathbf{Y}_k$ given $\mathbf{t}_k$ is

$$\mathbf{Y}_k \mid \mathbf{t}_k \sim \mathrm{MatNorm}_{n_k \times J}(\mathbf{X}_k^*\mathbf{B}_k^*, \mathbf{I}_{n_k}, \boldsymbol{\Sigma}_k), \tag{6}$$

where $\mathbf{X}_k^*$ is an $n_k \times (p+1)$ augmented design matrix formed by right column-binding $\mathbf{t}_k$ to $\mathbf{X}_k$, $\mathbf{B}_k^*$ is a $(p+1) \times J$ matrix of regression coefficients formed by lower row-binding $\boldsymbol{\psi}_k = (\psi_1, \ldots, \psi_J)^T$ to $\mathbf{B}_k$, and $\sum_k$ is the $J \times J$ covariance of $\boldsymbol{\epsilon}_i$ in Equation (3). Updating both $\boldsymbol{\psi}_k$ and $\mathbf{B}_k$ simultaneously using the augmented matrix $\mathbf{B}_k^*$ simplifies the MCMC sampler and is equivalent to separate updates of $\boldsymbol{\psi}_k$ and $\mathbf{B}_k$ when $\boldsymbol{\psi}_k$ and $\mathbf{B}_k$ are uncorrelated. This matrix normal representation admits conditionally conjugate prior distributions, which in turn leads to efficient Gibbs sampling for posterior inference. We formalize this in the following proposition, which establishes the conditional conjugacy of $\mathbf{B}_k^*$ and $\sum_k$.

Proposition 1. *Let* $\mathbf{B}_k^*$ *and* $\sum_k$ in Equation (6) *have a joint matrix normal-inverse Wishart (IW) prior, denoted* $MatNorm\text{-}IW_{(p+1)\times J}(\mathbf{B}_{0k}^*, \mathbf{L}_{0k}, \upsilon_{0k}, \mathbf{V}_{0k})$, *of the form*

$$\begin{aligned}\pi(\mathbf{B}_k^*, \boldsymbol{\Sigma}_k) = \pi(\mathbf{B}_k^* \mid \boldsymbol{\Sigma}_k)\pi(\boldsymbol{\Sigma}_k) \Rightarrow (\mathbf{B}_k^*, \boldsymbol{\Sigma}_k) \mid (\mathbf{B}_{0k}^*, \mathbf{L}_{0k}, \upsilon_{0k}, \mathbf{V}_{0k}) \\ \sim MatNorm_{(p+1)\times J}(\mathbf{B}_{0k}^*, \mathbf{L}_{0k}, \boldsymbol{\Sigma}_k) IW(\upsilon_{0k}, \mathbf{V}_{0k}),\end{aligned}$$

where $\mathbf{B}_{0k}^*$ *is a* $(p+1) \times J$ *prior location matrix,* $\mathbf{L}_{0k}$ *and* $\mathbf{V}_{0k}$ *are, respectively,* $(p+1) \times (p+1)$ *and* $J \times J$ *prior scale matrices, and* υ_{0k} *denotes the prior degrees of freedom. Then, the full conditional distribution of* $\mathbf{B}_k^*$ *is* $MatNorm_{(p+1)\times J}(\mathbb{B}_k^*, \mathbf{L}_k, \boldsymbol{\Sigma}_k)$, *where*

$$\begin{aligned}\mathbb{B}_k^* &= \mathbf{L}_k(\mathbf{L}_{0k}^{-1}\mathbf{B}_{0k}^* + \mathbf{X}_k^{*T}\mathbf{Y}_k) \\ \mathbf{L}_k &= (\mathbf{L}_{0k}^{-1} + \mathbf{X}_k^{*T}\mathbf{X}_k^*)^{-1},\end{aligned}$$

and $\mathbf{X}_k^*$ *is the augmented covariate matrix defined in* Equation (6). *Likewise, the full conditional distribution of* $\sum_k$ *is* $IW(\upsilon_k, \mathbf{V}_k)$, *where*

$$\begin{aligned}&\upsilon_k = \upsilon_0 + n_k + p + 1 \text{ and} \\ &\mathbf{V}_k = \mathbf{V}_{0k} + (\mathbf{B}_k^* - \mathbf{B}_{0k}^*)^T\mathbf{L}_{0k}^{-1}(\mathbf{B}_k^* - \mathbf{B}_{0k}^*) \\ &\qquad + (\mathbf{Y}_k - \mathbf{X}_k^*\mathbf{B}_k^*)^T(\mathbf{Y}_k - \mathbf{X}_k^*\mathbf{B}_k^*).\end{aligned}$$

The proof is provided in Web Appendix A.

3.2 | Pólya-Gamma multinomial regression for cluster probabilities

To accommodate heterogeneity in the cluster-membership probabilities, we model $\boldsymbol{\pi}_{ki}$ as a function of covariates using a multinomial logit model

$$\pi_{ki} = \Pr(z_i = k \mid \mathbf{w}_i) = \frac{e^{\mathbf{w}_i^T\boldsymbol{\delta}_k}}{\sum_{h=1}^K e^{\mathbf{w}_i^T\boldsymbol{\delta}_h}},\ k = 1, \ldots, K, \tag{7}$$

where $\mathbf{w}_i$ is an $r \times 1$ vector of subject-level covariates, $\boldsymbol{\delta}_k$ is an $r \times 1$ vector of cluster-specific regression parameters. For identifiability, we choose category K as reference and set $\boldsymbol{\delta}_K = \mathbf{0}$. To facilitate sampling, we adopt the efficient data-augmentation approach introduced by

Polson et al. (2013), which expresses the inverse-logit function as a scale-normal mixture of Pólya-Gamma densities. A random variable w is said to follow a Pólya-Gamma distribution with parameters $b > 0$ and $c \in \mathbb{R}$ if

$$f(w \mid b, c) = \frac{1}{2\pi^2} \sum_{s=1}^{\infty} \frac{g_s}{(s - 1/2)^2 + c^2/(4\pi^2)}, \tag{8}$$

where $g_s \overset{\text{iid}}{\sim} \text{Ga}(b, 1)$ for $s = 1, \ldots, \infty$. Polson et al. (2013) establish that, for a logistic regression model, the likelihood can be written as a scale-mixture of normal densities with Pólya-Gamma precision terms w, resulting in closed-form MVN full conditional distributions for logistic regression parameters. To extend the augmentation approach to the multinomial setting, we first introduce the binary indicators $U_{ki} = \mathbb{1}_{(z_i = k)}$, where $\mathbb{1}_{(z_i = k)}$ is the indicator function equal to 1 if $(z_i = k)$ and 0 otherwise. The conditional distribution of $\boldsymbol{\delta}_k$, given $\mathbf{U}_k = (U_{k1}, \ldots, U_{kn})^T$ and the remaining regression coefficients $\boldsymbol{\delta}_{h \neq k}$, is

$$\begin{aligned} &p(\boldsymbol{\delta}_k \mid \mathbf{z}, \boldsymbol{\delta}_{h \neq k}) \\ &\quad = p(\boldsymbol{\delta}_k \mid \mathbf{U}_k, \boldsymbol{\delta}_{h \neq k}) \propto p(\boldsymbol{\delta}_k) \prod_{i=1}^{n} \pi_{ki}^{U_{ki}} (1 - \pi_{ki})^{1 - U_{ki}}, \end{aligned} \tag{9}$$

where $p(\boldsymbol{\delta}_k)$ is the prior distribution of $\boldsymbol{\delta}_k$. We rewrite $\boldsymbol{\pi}_{ki}$ as

$$\pi_{ki} = \Pr(U_{ki} = 1) = \frac{e^{\mathbf{w}_i^T \delta_k}}{\sum_{h=1}^{K} e^{\mathbf{w}_i^T \delta_h}} = \frac{e^{\mathbf{w}_i^T \delta_k}}{\sum_{h \neq k}^{K} e^{\mathbf{w}_i^T \delta_h} + e^{\mathbf{w}_i^T \delta_k}},$$

where dividing throughout by $\sum_{h \neq k}^{K} e^{\mathbf{w}_i^T \boldsymbol{\delta}_h}$ yields

$$\pi_{ki} = \frac{e^{\mathbf{w}_i^T \delta_k - c_{ki}}}{1 + e^{\mathbf{w}_i^T \delta_k - c_{ki}}} = \frac{e^{\eta_{ki}}}{1 + e^{\eta_{ki}}},$$

with $c_{ki} = \log \sum_{h \neq k} e^{\mathbf{w}_i^T \boldsymbol{\delta}_h}$ and $\eta_{ki} = \mathbf{w}_i^T \boldsymbol{\delta}_k - c_{ki}$. We use c_{ki} and η_{ki} to reexpress Equation (9) as

$$\begin{aligned} p(\boldsymbol{\delta}_k \mid \mathbf{z}, \boldsymbol{\delta}_{h \neq k}) &\propto p(\boldsymbol{\delta}_k) \prod_{i=1}^{n} \left(\frac{e^{\eta_{ki}}}{1 + e^{\eta_{ki}}} \right)^{U_{ki}} \left(\frac{1}{1 + e^{\eta_{ki}}} \right)^{1 - U_{ki}} \\ &= p(\boldsymbol{\delta}_k) \prod_{i=1}^{n} \frac{(e^{\eta_{ki}})^{U_{ki}}}{1 + e^{\eta_{ki}}}, \end{aligned} \tag{10}$$

where the product term denotes the likelihood from a logistic regression model. We can therefore apply the Pólya-Gamma sampler for logistic regression to update each $\boldsymbol{\delta}_k$ one at a time based on the binary indicators U_{ki}. First, we define for $k = 1, \ldots, K$, the $n \times 1$ vector $\mathbf{U}_k^* = \left(\frac{U_{k1} - 1/2}{w_{k1}} + c_{k1}, \ldots, \frac{U_{kn} - 1/2}{w_{kn}} + c_{kn} \right)^T$. As shown in Web Appendix B, the conditional

distribution of $\mathbf{U}_k^*$ given $\boldsymbol{w} = (w_{k1}, \ldots, w_{kn})^T$ is $\mathrm{N}_n(\mathbf{W}\boldsymbol{\delta}_k, \mathbf{O}_k^{-1})$, where $\mathbf{O}_k = \mathrm{Diag}(w_{k1}, \ldots, w_{kn})$ and $\mathbf{W}$ is an $n \times r$ design matrix with rows $\mathbf{w}_i^T$ for $i = 1, \ldots, n$. Thus, the full conditional distribution of $\boldsymbol{\delta}_k$ is given by

$$p(\boldsymbol{\delta}_k \mid \mathbf{z}, \mathbf{O}_k, \boldsymbol{\delta}_{h \neq k}) \propto p(\boldsymbol{\delta}_k)\exp\left\{-\frac{1}{2}(\mathbf{U}_k^* - \mathbf{W}\boldsymbol{\delta}_k)^T\mathbf{O}_k(\mathbf{U}_k^* - \mathbf{W}\boldsymbol{\delta}_k)\right\}. \tag{11}$$

Assuming a $\mathrm{N}_t(\mathbf{d}_{0k}, \mathbf{S}_{0k})$ prior for $\boldsymbol{\delta}_k$ allows for Gibbs sampling for the clustering model as detailed in Web Appendix B.

3.3 | Extensions to multivariate skew-*t* distributions

To accommodate outliers and heavy tails, we extend Equation (1) by assuming, conditional on $z_i = k$, that $\mathbf{y}_i$ follows a MST distribution (Gupta, 2003):

$$\begin{aligned}\mathbf{y}_i \mid (z_i = k) &\overset{ind}{\sim} \mathrm{MST}_J(\boldsymbol{\zeta}_{ki}, \boldsymbol{\alpha}_k, \boldsymbol{\Omega}_k, \upsilon_k), \text{ with density} \\ f(\mathbf{y}_i \mid z_i = k) &= 2f_{t_J}(\mathbf{y}_i; \boldsymbol{\zeta}_{ki}, \boldsymbol{\Omega}_k, \upsilon_k)T_{\upsilon_k + J} \\ &\quad \times \left\{\boldsymbol{\alpha}_k^T(\mathbf{y}_i - \boldsymbol{\zeta}_{ki})\sqrt{\frac{\upsilon_k + J}{\upsilon_k + Q_{y_i}}}\right\},\end{aligned} \tag{12}$$

where $f_{t_J}(\mathbf{y}_i; \boldsymbol{\zeta}_{ki}, \boldsymbol{\Omega}_k, \upsilon_k)$ denotes the CDF of a J-dimensional t distribution with location $\boldsymbol{\zeta}_{ki}$, covariance $\boldsymbol{\Omega}_k$, and fixed degrees of freedom υ_k that may vary across clusters; $T_{\upsilon_k + J}$ denotes the distribution function of the scalar standard t distribution with $\upsilon_k + J$ degrees of freedom; and $Q_{y_i} = (\mathbf{y}_i - \boldsymbol{\zeta}_{ki})^T\boldsymbol{\Omega}_k^{-1}(\mathbf{y}_i - \boldsymbol{\zeta}_{ki})$. As before, we adopt a conditional representation for $\mathbf{y}_i$ to facilitate Gibbs sampling (Frühwirth-Schnatter and Pyne, 2010). Specifically, we augment the MSN conditional representation in Equation (3) by introducing subject-specific scale terms, d_i, yielding an MST regression conditional on z_i, t_i, and d_i of the form:

$$\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\beta}_k + \frac{t_i}{\sqrt{d_i}}\boldsymbol{\psi}_k + \frac{1}{\sqrt{d_i}}\boldsymbol{\epsilon}_i, \tag{13}$$

where $d_i \sim \mathrm{Gamma}\left(\frac{\xi}{2}, \frac{\xi}{2}\right)$, with ξ being a prespecified known degrees of freedom parameter common to all clusters, and t_i and $\boldsymbol{\epsilon}_i$ are defined as in Equation (3). In principle, ξ may be prespecified but vary across clusters (becoming ξ_k), though here we a constant value across clusters for simplicity. For details on posterior inference, see Web Appendix B.

3.4 | Cluster-specific imputation under conditional ignorability

To accommodate intermittent missing data, we propose a convenient MCMC-embedded imputation algorithm in which we assume that the missingness mechanism is conditionally ignorable given the cluster indicators z_i, extending recent work on latent class pattern mixture models for informative dropout (Roy, 2007). We use the term “MCMC-embedded” to denote the fact that each missing value is imputed once per MCMC iteration using current cluster-specific parameter values, allowing for convenient multiple imputation as part of the

MCMC algorithm. Ensuring subjects have complete response vectors also enables us to update the regression parameters in a compact manner, as described in Web Appendix B. Here, z_i functions as a discrete shared parameter that induces unobserved association between the missingness process and the missing data. Suppose $\mathbf{y}_i$ has $q_i \in (1, \ldots, J)$ observed values, denoted $\mathbf{y}_i^{obs}$, and $J - q_i$ intermittent missing values, denoted $\mathbf{y}_i^{miss}$. Let $\mathbf{R}_i = (R_{i1}, \ldots, R_{iJ})^T$ be a $J \times 1$ vector of binary response indicators, such that $R_{ij} = 1$ if infant i has a Bayley measurement at visit j. Under conditional ignorability, the conditional distribution of $\mathbf{R}_i$ given $(z_i, \mathbf{y}_i^{obs}, \mathbf{y}_i^{miss})$ is

$$\begin{aligned} f(\mathbf{R}_i \mid z_i = k, \mathbf{y}_i^{obs}, \mathbf{y}_i^{miss}, \mathbf{X}_i, \boldsymbol{\gamma}_k) \\ = f(\mathbf{R}_i \mid z_i = k, \mathbf{y}_i^{obs}, \mathbf{X}_i, \boldsymbol{\gamma}_k), \end{aligned} \tag{14}$$

where, in this context, $\mathbf{X}_i$ is a $J \times m$ design matrix and $\boldsymbol{\gamma}_k$ is an $m \times 1$ vector of cluster-specific parameters related to the missing data mechanism. As detailed in Step 4 of Web Appendix B, z_i serves as a latent shared parameter that induces marginal correlation between Y_i^{miss} and $\mathbf{R}_i$.

Under conditional ignorability, conditioning on z_i ensures that $\mathbf{R}_i$ does not depend on the missing observations $\mathbf{y}_i^{miss}$. We can therefore impute $\mathbf{y}_i^{miss}$ from its conditional MVN distribution given $(z_i, t_i, \mathbf{y}_i^{obs})$ as described in Web Appendix B. While the complete data vector $\mathbf{y}_i = \{\mathbf{y}_i^{obs}, \mathbf{y}_i^{miss}\}$ follows a MVN distribution conditional on t_i, after marginalizing over t_i, $\mathbf{y}_i$ follows a joint MSN distribution. Thus, the proposed conditional imputation procedure provides a convenient way of imputing missing MSN responses using samples from more standard densities.

Finally, given $z_i = k$, we independently model the J response indicators for infant i as

$$\begin{aligned} R_{ij} \mid (z_i = k, \boldsymbol{\gamma}_k, b_{ki}) \sim \text{Bern}(\phi_{ijk}),\ j = 1, \ldots, J \\ \text{logit}(\phi_{ijk}) = \mathbf{x}_{ij}^T \boldsymbol{\gamma}_k + b_{ki}, \end{aligned} \tag{15}$$

where $\mathbf{x}_{ij}$ is an $m \times 1$ vector of covariates, and $\boldsymbol{\gamma}_k$ is the $m \times 1$ vector of cluster-specific regression parameters from Equation (14). We note that while the missing data regression parameters may in principle be shared across clusters, cluster-specific parameters allow investigators to identify different missing data patterns across clusters. Further, correctly modeling cluster-specific missingness mechanisms is necessary to obtain appropriate inference for cluster-specific parameters. Because the response indicators may be correlated over time, we also include a subject-level random intercept b_{ki} conditionally distributed as $N(0, \sigma_k^2)$ given $z_i = k$. Although we assume conditional ignorability of $\mathbf{R}_i$ and $\mathbf{y}_i^{miss}$ given z_i, because the $\boldsymbol{\phi}_{ijk}$ terms from model (15) appear in the full conditional update for z_i (Web Appendix B), $\mathbf{R}_i$ and $\mathbf{y}_i^{miss}$ are marginally correlated, resulting in a marginal missing not at random (MNAR) mechanism.

3.5 | Bayesian inference

3.5.1 | Prior specification—We adopt a Bayesian approach and assign prior distributions to all model parameters. For designs not involving time-dependent covariates, we assign a joint MatNorm-IW$_{(p+1)\times J}(\mathbf{B}^*_{0k}, \mathbf{L}_{0k}, \upsilon_{0k}, \mathbf{V}_{0k})$ to $(\mathbf{B}^*_k, \sum_k)$ as described in Proposition 1. For time-varying designs, we assign independent MVN priors to $\boldsymbol{\beta}_k$ and $\boldsymbol{\psi}_k$ from Equation (3); details are provided in Step 5(b) of Web Appendix B. For the multinomial logit model, the regression parameters $\boldsymbol{\delta}_k = (\delta_{k1}, \ldots, \delta_{kr})^T$ are assigned a $N_r(\mathbf{d}_{0k}, \mathbf{S}_{0k})$ prior for $k = 1, \ldots, K-1$, which is conditionally conjugate under the Pólya-Gamma sampling scheme described in Section 3.2. Finally, from Equation (15), we assume a $N_m(\mathbf{g}_{0k}, \mathbf{G}_{0k})$ prior for $\boldsymbol{\gamma}_k$ and an inverse-gamma IG$(\lambda_{1k}, \lambda_{2k})$ prior for σ^2_k, where λ_{2k} is a scale parameter. In general, hyperparameters can vary across clusters, though they may be shared across clusters in practice. For the skew-t model, we assume $d_i \sim$ Gamma $(\frac{\xi}{2}, \frac{\xi}{2})$, where ξ is a prespecified value.

3.5.2 | Posterior computation, assessment of MCMC convergence, label switching, and model selection—The above prior specification induces closed-form full conditionals for all model parameters, which can be efficiently updated as part of the Gibbs sampler detailed in Web Appendix B. We monitor MCMC convergence through standard diagnostics, such as trace plots and effective sample sizes. To address label switching, a common issue for Bayesian mixture models, we implemented the iterative Equivalence Classes Representatives (ECR) relabeling algorithm included in the label.switching package in R (Papastamoulis, 2016). In our simulation studies and application, we observed immediate convergence of the ECR algorithm, indicating no evidence of label switching in our analyses. Because our primary objective is to identify a small number of clinically meaningful motor development clusters, we adopt the widely applicable information criterion (WAIC) to select the number of clusters K (Watanabe, 2010). In Section 4.3, we demonstrate that this measure accurately recovers the true number of clusters under realistic parameter settings.

4 | SIMULATION STUDIES

4.1 | Simulation to compare the MSN model to the MVN model

Our first simulation compared MSN and MVN mixture models to investigate whether ignoring skewness leads to poor inferences in a setting resembling the Nurture study. To emulate the Nurture study, we simulated $n = 1000$ subjects from the following model:

$$f(\mathbf{y}_i) = \sum_{k=1}^{3} \pi_{ki} f(\mathbf{y}_i \mid \theta_k), \tag{16}$$

where $\mathbf{y}_i = (y_{i1}, \ldots, y_{i4})^T$ to conform to the $J = 4$ measurement occasions in the Nurture study; $\boldsymbol{\theta}_k$ is the set of parameters specific to cluster k ($k = 1, 2, 3$), and $\mathbf{y}_i | \boldsymbol{\theta}_k \sim \text{MSN}_4(\boldsymbol{\zeta}_{ki}, \boldsymbol{\alpha}_k, \boldsymbol{\Omega}_k)$; $\boldsymbol{\zeta}_{ki} = (\zeta_{ki1}, \ldots, \zeta_{ki4})^T$, $\zeta_{ki1} = \beta_{kj1} + \beta_{kj2} x_i$, and x_i is a N(0,1) covariate whose effect varies across the J measurement occasions. We modeled the cluster probabilities in Equation (7) as a function of an intercept and one baseline covariate, w_{i1}, implying that $r = 2$. We did

not introduce missing data into this simulation, as we address missing data in the second simulation study. As a result, the total number of complete measurements was $N = n \times J = 4000$. The generated data included $n_1 = 318$ infants in cluster 1, $n_2 = 288$ in cluster 2, and $n_3 = 394$ in cluster 3.

Because the model included no time-varying covariates—only time-varying effects—we used the matrix normal formulation given in Proposition 1, yielding a $(p + 1) \times J = 3 \times 4$ matrix $\mathbf{B}_k^*$. We chose the matrix normal hyperparameters described in Section 3.5.1 to be homogeneous across the three clusters by setting, for $k = 1, 2, 3$, $\mathbf{B}_{0k}^* = \mathbf{0}_{3\times 4}$, $\mathbf{L}_{0k} = \mathbf{I}_3$, $\mathbf{V}_{0k} = \mathbf{I}_4$, and $v_{0k} = J + 2 = 6$, which gives $\mathrm{E}(\boldsymbol{\Sigma}_k) = \mathbf{I}_4$. Similarly, for the clustering model, we set $\mathbf{d}_{01} = \mathbf{d}_{02} = (0, 0)^T$ and $\mathbf{S}_{01} = \mathbf{S}_{02} = \mathbf{I}_2$, noting that $k = 3$ is the reference cluster. To investigate the effect of ignoring skewness, we allowed the vector of skewness parameters, $\boldsymbol{\alpha}_k$, to vary across clusters; for cluster 3, we assumed no skewness ($\boldsymbol{\alpha}_3 = \mathbf{0}$), implying MVN data for this cluster. We then fit both MSN and MVN mixture models to data generated from model (16). We ran the MCMC for 10 000 iterations with a burn-in of 1000. MCMC diagnostics indicated rapid convergence and excellent mixing (Web Figure 1).

The WAIC values for the MSN and MVN mixture models were 12 112 and 17 499, respectively, indicating better fit for the MSN model, as expected. Table 1 presents posterior mean estimates and 95% credible intervals (CrIs) for cluster 1 from the MSN and MVN models. Web Table 3 presents the results for the other two clusters. As expected, the MSN model provided accurate estimates throughout, whereas the MVN model consistently produced incorrect estimates with poor coverage when data were skewed, as in clusters 1 and 2. In particular, ignoring skewness inflated the variance estimates under the MVN model as a way to compensate for the skewness in the data. However, when data were not skewed, as in cluster 3, both models performed similarly (Web Table 3). Thus, the MSN model can be reliably used in place of the MVN model even when data are not overtly skewed.

4.2 | Simulation to compare imputation methods

Next, we evaluated the effect of failing to account for the missing data model in Equation (15). To do so, we generated $n = 1000$ observations from a 3-cluster ($K = 3$) MSN mixture model similar in design to Simulation 1. We then removed observations intermittently across the four measurement occasions according to model (15), which included two continuous covariates and an intercept, implying $m = 3$ from Equation (15). The model also included a random intercept with a common variance of $\sigma_k^2 = 1$ across clusters. After removing missing data, the number of available measurements in each cluster was $N_1 = 1463$, $N_2 = 819$, and $N_3 = 1209$. We ran each model for 10 000 iterations with a burn-in of 1000. MCMC diagnostics showed rapid convergence as shown in Web Figure 2.

We then fit two MSN mixture models to the simulated data, each with different missing data assumptions. The first method assumed conditional ignorability, as described in Section 3.4, where the missing responses and missing data pattern were assumed to be independent conditional on z_i, and a model of the missing data pattern was fit as in Equation (15). The second method assumed marginal ignorability, where the missing responses and missing data pattern were assumed to be independent marginally (ie, not conditional on z_i). Thus, the

marginal ignorability approach did not adopt a model of the missing data mechanism as in Equation (15). Both imputation methods utilized MCMC-embedded imputation, where missing values were updated from cluster-specific multivariate normal conditional distributions at each MCMC iteration using the current values of parameters in the sampler.

As shown in Table 2, the conditional ignorability imputation method more accurately recovered true parameter values when compared to marginal ignorability. This result suggests that even when all other components of the model are correctly specified, making the strict marginal ignorability assumption (and thus ignoring model (15) altogether) can lead to biased estimates.

4.3 | Simulation to validate choice of K

We conducted a final simulation to validate the use of WAIC for determining the number of clusters, K. We generated four MSN data sets; one data set for each value of $K = \{2, 3, 4, 5\}$. For each simulated data set, we fit the proposed Bayesian MSN model with $K = \{2, 3, 4, 5\}$ and computed WAIC in each case. For each scenario, we ran the MCMC algorithms for 10 000 iterations with a burn-in of 1000. MCMC diagnostics indicated rapid convergence for all models (Web Figure 3). As shown in Web Table 5, the WAIC measure recovered the true value of K in all cases. For some simulations (eg, true $K = 2$), we were unable to fit the MSN model when the fitted K was large due to the occurrence of vacant clusters during MCMC sampling. We have found that this generally occurs when the data do not support large values of K.

5 | APPLICATION TO NURTURE STUDY

We applied our proposed model to the Nurture data by fitting an MSN mixture model that included Bayley scores centered and scaled by timepoint as the response, indicators for the four study visits corresponding to timepoint-specific intercepts, and binary food security status as the exposure of interest. The model also included time-invariant birth weight for gestational age z-score, number of children in the household, and an indicator for breastfeeding, as these likely impact infant development within each cluster. We allowed the covariate effects to vary over time, resulting in a parameter dimension of $p = 20$ for this component of the model (Table 3). For the multinomial logit cluster-membership model, we included an intercept, birth weight for gestational age z-score, infant race, and infant gender as covariates, as these variables are believed to affect the placement of infants into latent development clusters. The 471 missing measurements were imputed using the MCMC-embedded MNAR imputation method described in Section 3.4. The missing data model (15) included a fixed intercept, birth weight for gestational age z-score, infant gender, infant race, and a random intercept. To select the number of clusters, we fit several MSN models with varying specifications for K and used WAIC to choose the best fitting model. The WAIC values were 9141, 10 088, 11 203, and 11 410 for $K = 2, 3, 4, 5$, respectively. We also fit 3-df MST models with two to five clusters; these yielded WAIC values of 13 228, 13 934, 14 002, and 14 356 respectively, suggesting that the 2-cluster MSN model provided best fit among all models considered. We ran each model for 10 000 MCMC iterations, with a burn-

in of 1000. We observed fast MCMC convergence in all cases with no evidence of label switching. MCMC diagnostics for the 2-cluster MSN model are presented in Web Figure 4.

Table 3 presents posterior means and 95% CrIs for the 2-cluster model. In cluster 1, we observed a significant detrimental effect of food insecurity at each timepoint. However, in cluster 2, we only observed a significant detrimental effect of food insecurity at months 9 and 12, though the effect sizes were more modest than in cluster 1. These trends are also displayed in Figure 3. We observed a significant positive effect of breastfeeding in cluster 1, but not in cluster 2, suggesting that breastfeeding may especially benefit infants exhibiting delayed motor development. We did not observe a significant effect of either birth weight for gestational age z-score or number of children in the household. From the Pólya-Gamma multinomial logit component, we found that female infants were more likely to belong to cluster 1. From the missing data model, the intercepts suggest that more missing observations occur for infants in cluster 1 compared to those in cluster 2 for the reference covariate group. Moreover, female infants in cluster 1 had significantly higher log-odds of missing a measurement compared to male infants in cluster 1, while black infants in cluster 2 had significantly lower log-odds of missing a measurement compared to other infants.

As shown in Table 3, the skewness estimates for cluster 1 indicate little evidence of skewness, as all associated 95% CrIs contained zero. However, in cluster 2, the predicted Bayley scores were negatively skewed at 6 months, in agreement with the preliminary analysis presented in Section 2. This suggests that the skewness observed in the data was driven primarily by the healthy-developing class, highlighting the model's ability to discern different skewness patterns across clusters. Further, the clusters identified by the model were distinct from one another, as 510 (91%) of infants remained in the same cluster across the postburn-in MCMC iterations. Finally, the estimated covariance and correlation matrices (Web Table 6 and 7, respectively), indicated an unstructured pattern for both clusters, with greater variability in cluster 2.

6 | DISCUSSION

We have developed Bayesian MSN and MST for skewed longitudinal data that feature intermittent missingness. The model has many appealing features: it accounts for skewness in the infant development scores, associations among repeated measures, and allows for efficient inference of the cluster assignment probabilities. The model can be applied to skewed as well as symmetric data, since the symmetric version is contained as a special case. Additionally, the model handles missing data under a conditional ignorability assumption that relaxes standard MAR assumptions.

Through simulations, we showed that ignoring skewness in even moderately skewed data results in incorrect inference, whereas the MSN mixture model recovers the true parameter values when the data are skewed. Furthermore, we showed that failing to account for conditional ignorability results in biased estimates when the response mechanism depends on cluster assignment. Finally, we conducted simulations to validate the use of WAIC, supporting the use of this measure in practice.

We applied our method to the Nurture data to assess the effect of household food security during pregnancy on motor development scores and to investigate possible clustering of infant development trajectories. We identified two distinct clusters of infants: one with delayed motor development and significantly impaired by food insecurity, and a second that exhibited healthy motor development and was only modestly affected by food insecurity toward the end of infancy. This suggests that household food insecurity may compound the negative impacts of delayed motor development. On the other hand, we found that breastfeeding improved motor development among infants with delayed development. These results add to the growing body of literature on the effect of household food security on infant outcomes.

To extend this work, the model could accommodate dropout in addition to intermittent missingness using a cluster-specific discrete time-to-event model. Additionally, cluster-specific shared parameters could link the outcome and missing data models, relaxing the conditional ignorability assumption. More broadly, the method should prove useful in a range of settings involving multivariate skew data with informative missing responses. From a practical perspective, investigators looking to model clustered repeated-measures data can use the diagnostics described in Section 2 to determine whether the MSN model is appropriate. Given that the computational demand of the MSN and MST models is negligible compared to the MVN model, we recommend fitting the MSN or MST model first and using the estimated skewness parameters to determine whether simplifications to the MVN model can be made.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGMENT

This work was supported by NIH grants R21 LM012866 and R01DK094841.

Funding information

U.S. National Library of Medicine, Grant/Award Number: R21 LM012866; National Institute of Diabetes and Digestive and Kidney Diseases, Grant/Award Number: R01DK094841

DATA AVAILABILITY STATEMENT

The Nurture data were collected as part of grant R01DK094841 from the National Institutes of Health and are housed at Duke University Medical Center. The Nurture data are available upon request with appropriate permissions, agreements between institutions, and documentation of ethical approval.

REFERENCES

Aaltonen S, Latvala A, Rose RJ, Pulkkinen L, Kujala UM, Kaprio J, et al. (2015) Motor development and physical activity: a longitudinal discordant twin-pair study. Medicine and Science in Sports and Exercise, 47, 2111–2118. [PubMed: 26378945]

Azzalini A and Valle AD (1996) The multivariate skew-normal distribution. Biometrika, 83, 715–726.

Bayley N (2006). Bayley-III: Bayley Scales of Infant and Toddler Development. San Antonio, TX: Giunti OS.

Benjamin Neelon SE, Østbye T, Bennett GG, Kravitz RM, Clancy SM, Stroo M, et al. (2017) Cohort profile for the Nurture Observational Study examining associations of multiple caregivers on infant growth in the Southeastern USA. BMJ Open, 7, e013939.

Chen JT and Gupta AK (2005) Matrix variate skew normal distributions. Statistics, 39, 247–253.

Filatova S, Koivumaa-Honkanen H, Hirvonen N, Freeman A, Ivandic I, Hurtig T, et al. (2017) Early motor developmental milestones and schizophrenia: a systematic review and meta-analysis. Schizophrenia Research, 188, 13–20. [PubMed: 28131598]

Frühwirth-Schnatter S and Pyne S (2010) Bayesian inference for finite mixtures of univariate and multivariate skew-normal and skew-*t* distributions. Biostatistics, 11, 317–336. [PubMed: 20110247]

Gupta A (2003) Multivariate skew *t*-distribution. Statistics: A Journal of Theoretical and Applied Statistics, 37, 359–363.

Lin T-I, Wang W-L, McLachlan GJ and Lee SX (2018) Robust mixtures of factor analysis models using the restricted multivariate skew-*t* distribution. Statistical Modelling, 18, 50–72.

Papastamoulis P (2016) label.switching: an R package for dealing with the label switching problem in MCMC outputs. Journal of Statistical Software, 69, 1–24.

Polson NG, Scott JG and Windle J (2013) Bayesian inference for logistic models using Pólya–Gamma latent variables. Journal of the American Statistical Association, 108, 1339–1349.

Roy J (2007) Latent class models and their application to missing-data patterns in longitudinal studies. Statistical Methods in Medical Research, 16, 441–456. [PubMed: 17656451]

Sánchez GFL, Williams G, Aggio D, Vicinanza D, Stubbs B, Kerr C, et al. (2017) Prospective associations between measures of gross and fine motor coordination in infants and objectively measured physical activity and sedentary behavior in childhood. Medicine, 96, e8424. [PubMed: 29145249]

Shoaibi A, Neelon B, Østbye T and Benjamin-Neelon SE (2019) Longitudinal associations of gross motor development, motor milestone achievement and weight-for-length *z* score in a racially diverse cohort of US infants. BMJ Open, 9, e024440.

Taanila A, Murray GK, Jokelainen J, Isohanni M and Rantakallio P (2005) Infant developmental milestones: a 31-year follow-up. Developmental Medicine and Child Neurology, 47, 581–586. [PubMed: 16138663]

USDA (2019) Food security in the US: measurement. Available at: https://www.ers.usda.gov/topics/food-nutrition-assistance/food-security-in-the-us/measurement.aspx [Accessed 11 January 2020].

Watanabe S (2010) Asymptotic equivalence of Bayes cross-validation and widely applicable information criterion in singular learning theory. Journal of Machine Learning Research, 11, 3571–3594.

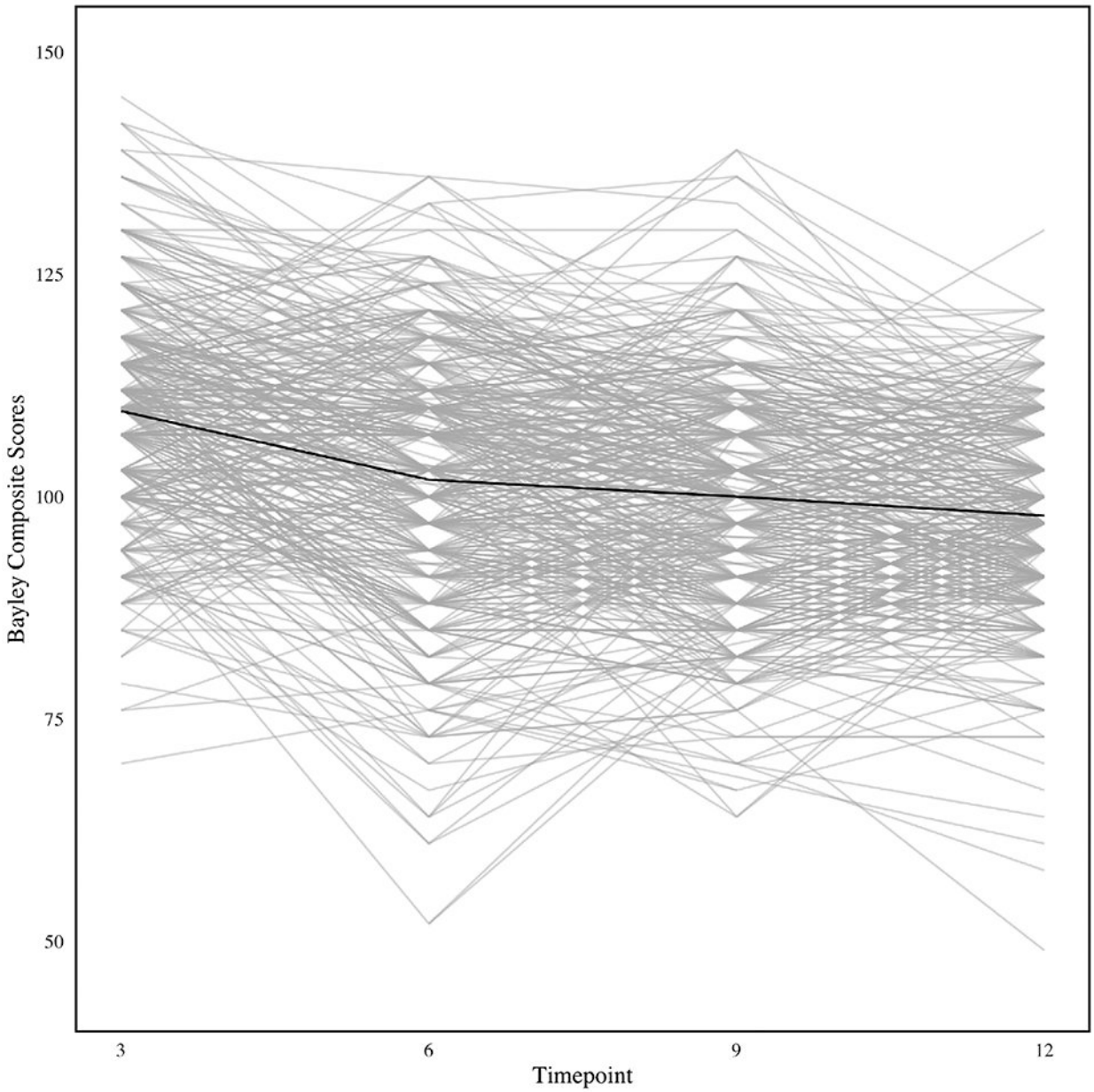

FIGURE 1.

Longitudinal profile plot of infant development trajectories, with mean Bayley motor development score shown in black *Note*. Plot is based on the $N = 1769$ available measurements for $n = 560$ infants

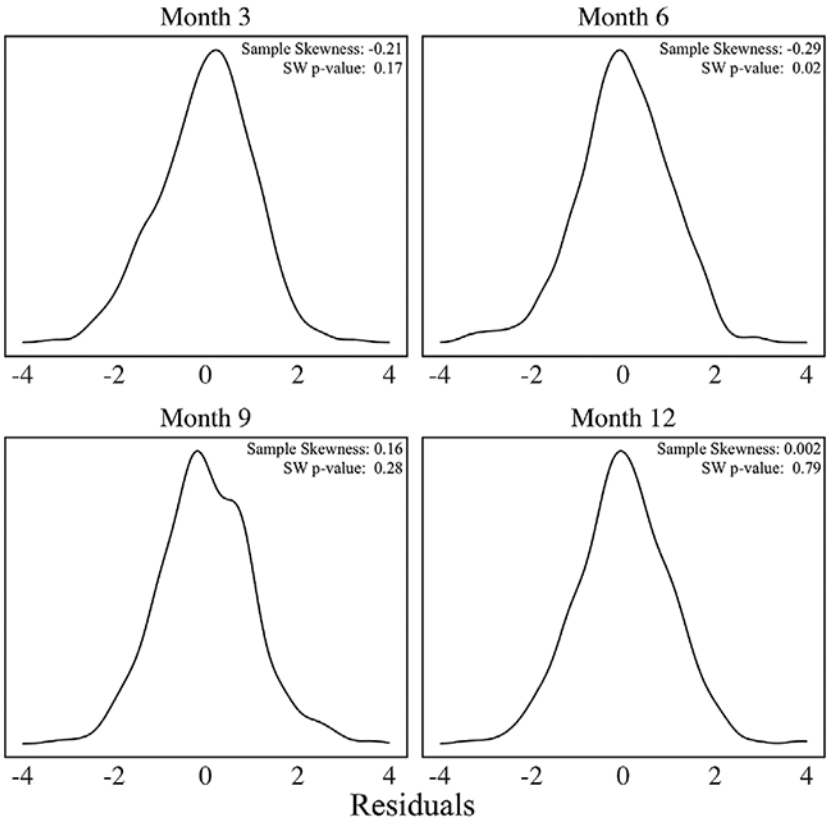

FIGURE 2.

Scaled residual plots at each visit based on a repeated-measures linear regression model with Bayley score as the outcome *Note*. Sample skewness statistics and *P*-values from Shapiro-Wilk (SW) tests are provided in the legends. Plots are based on the $N = 1769$ available measurements for $n = 560$ infants

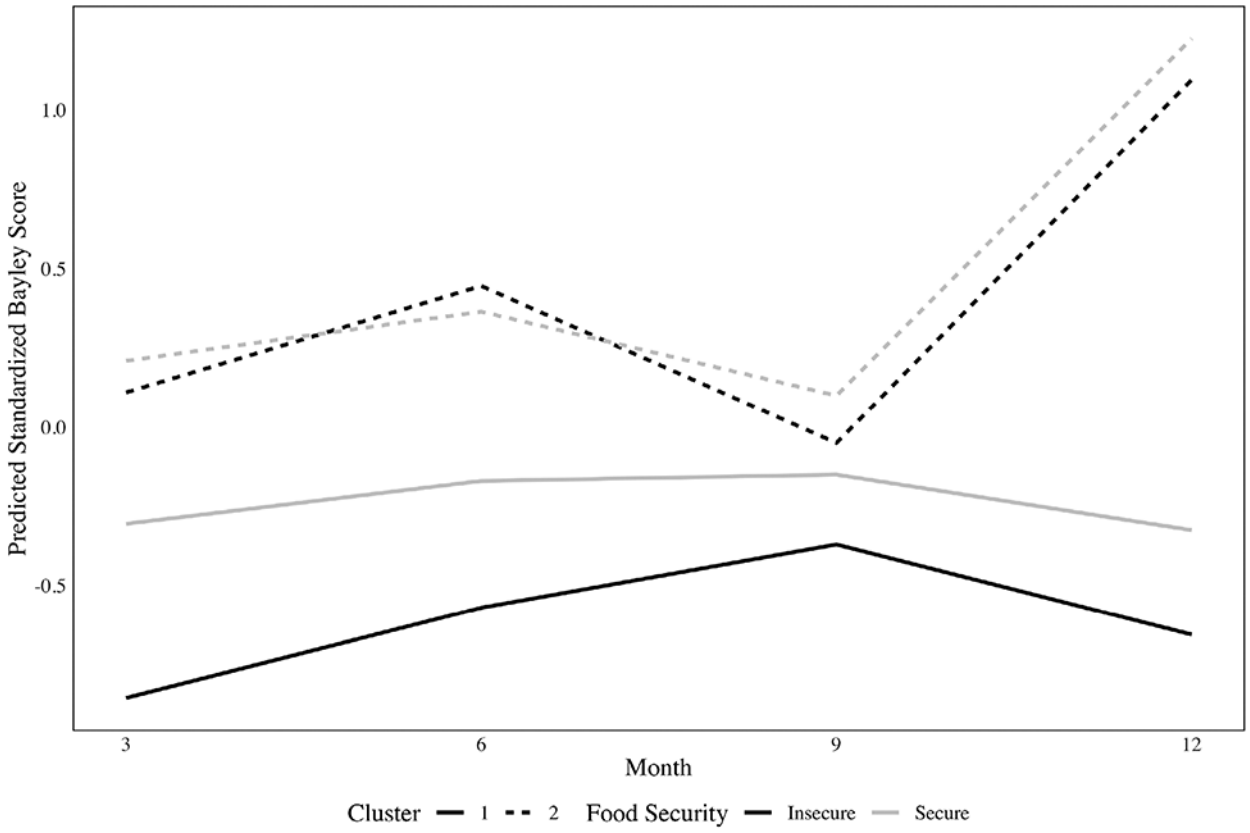

FIGURE 3.

Predicted motor development trajectories for each cluster and food security group in the application to the Nurture data *Note*. The model included timepoint-specific intercepts, time-invariant birth weight for gestational age *z*-score, the number of children in the household, and an indicator for breastfeeding. Estimated trajectories are given for a typical infant with a birth weight for gestational age *z*-score of 0, who was not breastfed, and who had 2.5 other children in the household. Solid lines indicate cluster 1 and dashed lines indicate cluster 2. Light shading represents food-secure infants, while dark shading represents food-insecure infants

TABLE 1

Results for cluster 1 from Simulation 1 with $n = 1000$, $J = 4$, $p = 2$, $K = 3$, $r = 2$

Model component ($k = 1$)	Parameter	True value	MSN Est. (95% CrI)	MVN Est. (95% CrI)
MSN regression	β_{111}	110.00	110.20 (109.97, 110.41)	106.36 (105.97, 108.71)
	β_{121}	115.00	115.13 (114.91, 115.33)	104.17 (103.93, 104.44)
	β_{131}	120.00	120.08 (119.83, 120.49)	128.02 (128.57, 129.08)
	β_{141}	125.00	125.15 (124.86, 125.49)	126.67 (126.31, 127.05)
	β_{112}	1.00	0.97 (0.84, 1.11)	0.90 (0.74, 1.08)
	β_{122}	1.50	1.51 (1.40, 1.62)	1.53 (1.41, 1.66)
	β_{132}	2.00	2.01 (1.89, 2.14)	2.20 (2.08, 2.33)
	β_{142}	2.50	2.50 (2.35, 2.66)	2.46 (2.28, 2.64)
	Σ_{111}	1.00	0.96 (0.77, 1.14)	2.42 (2.06, 2.84)
	Σ_{112}	0.50	0.47 (0.34, 0.61)	1.20 (0.99, 1.48)
	Σ_{113}	0.25	0.25 (0.04, 0.40)	−0.54 (−0.75, −0.34)
	Σ_{114}	0.12	0.11 (−0.02, 0.30)	−1.35 (−1.67, −1.06)
	Σ_{122}	1.00	0.99 (0.74, 1.19)	1.20 (0.99, 1.48)
	Σ_{123}	0.50	0.49 (0.26, 0.66)	1.24 (1.06, 1.46)
	Σ_{124}	0.25	0.24 (0.10, 0.43)	0.08 (−0.06, 0.21)
	Σ_{133}	1.00	0.99 (0.77, 1.09)	1.24 (1.06, 1.46)
	Σ_{134}	0.50	0.47 (0.22, 0.65)	1.15 (0.93, 1.40)
	Σ_{144}	1.00	1.01 (0.63, 1.23)	2.48 (2.15, 2.91)
	a_{11}	−2.00	−2.05 (−2.28, −1.66)	/
	a_{12}	−1.00	−1.01 (−1.30, −0.75)	/
	a_{13}	1.00	0.97 (0.65, 1.28)	/
	a_{14}	2.00	1.97 (1.67, 2.28)	/
Multinomial logit[a]	δ_{11}	−0.27	−0.23 (−0.47, −0.09)	−0.14 (−0.35, 0.08)
	δ_{12}	0.07	0.07 (−0.24, 0.37)	0.08 (−0.24, 0.38)
Missing Data	γ_{11}	−0.82	−0.84 (−0.96, −0.73)	−1.08 (−1.19, −0.99)
	γ_{12}	−1.08	−1.01 (−1.20, −0.91)	−1.80 (−1.96, −1.64)

Model component (k = 1)	Parameter	True value	MSN Est. (95% CrI)	MVN Est. (95% CrI)
	γ_{13}	−1.12	−1.08 (−1.20, −1.00)	−0.90 (−1.00, −0.80)
	σ_1^2	1.00	1.07 (0.92, 1.28)	0.89 (0.76, 1.07)
Estimated proportion[b]	π_1	0.32	0.32 (0.31, 0.33)	0.32 (0.30, 0.34)

Note. 10 000 iterations were run with a burn-in of 1000. Posterior means (95% CrIs) are presented for the multivariate skew-normal (MSN) and multivariate normal (MVN) mixtures. No missing data were introduced.

[a]Multinomial logit parameters comparing cluster 1 to cluster 3 (reference cluster).

[b]Estimated proportion of infants in cluster 1.

Slashes (/) indicate that estimates are not applicable.

Biometrics. Author manuscript; available in PMC 2021 July 22.

TABLE 2

Results for cluster 1 from Simulation 2

Model component (k = 1)	Parameter	True value	Conditional ignorability	Marginal ignorability
MSN regression	β_{111}	−2.90	−3.03 (−3.70, −2.60)	−3.72 (−3.99, −3.45)
	β_{121}	−2.70	−2.82 (−2.96, −2.69)	−2.87 (−2.92, −2.64)
	β_{131}	−2.92	−2.79 (−3.69, −2.43)	−3.76 (−4.04, −3.48)
	β_{141}	−3.68	−3.87 (−4.01, −3.73)	−3.83 (−3.96, −3.69)
	β_{112}	−2.78	−2.67 (−3.42, −2.24)	−3.57 (−3.86, −3.29)
	β_{122}	−2.59	−2.81 (−2.94, −2.67)	−2.87 (−2.91, −2.73)
	β_{132}	−2.71	−2.43 (−3.11, −2.15)	−3.44 (−3.70, −3.17)
	β_{142}	−2.79	−2.98 (−3.11, −2.84)	−2.97 (−3.10, −2.83)
	Σ_{111}	1.00	1.25 (0.84, 1.82)	1.54 (1.31, 1.85)
	Σ_{112}	0.50	0.59 (0.19, 1.15)	1.12 (0.91, 1.39)
	Σ_{113}	0.25	0.24 (0.11, 0.38)	0.93 (0.73, 1.19)
	Σ_{114}	0.12	0.17 (0.08, 0.21)	0.85 (0.65, 1.10)
	Σ_{122}	1.00	0.95 (0.49, 1.51)	1.12 (0.91, 1.39)
	Σ_{123}	0.50	0.52 (0.14, 1.04)	1.66 (1.40, 1.97)
	Σ_{124}	0.25	0.31 (0.12, 0.41)	1.15 (0.92, 1.41)
	Σ_{133}	1.00	1.12 (0.81, 1.19)	0.93 (0.73, 1.18)
	Σ_{134}	0.50	0.53 (0.24, 0.89)	0.85 (0.65, 1.10)
	Σ_{144}	1.00	1.08 (0.61, 1.75)	0.93 (0.73, 1.18)
	a_{11}	−1.00	−0.81 (−1.36, −0.05)	−1.91 (−2.17, −1.74)
	a_{12}	−1.00	−1.18 (−1.63, −0.03)	−1.22 (−0.75, −1.66)
	a_{13}	−1.00	−1.10 (−1.66, −0.14)	−1.50 (−2.25, −0.64)
	a_{14}	−1.00	−1.29 (−1.62, −0.37)	−1.43 (−1.88, −1.01)
Multinomial logit[a]	δ_{11}	−0.54	−0.53 (−0.75, −0.31)	−0.64 (−0.85, −0.43)
	δ_{12}	−0.01	−0.02 (−0.33, 0.38)	−0.08 (−0.33, 0.28)
Missing data	γ_{11}	−1.10	−1.06 (−1.40, −0.75)	/
	γ_{12}	−1.27	−1.13 (−1.42, −0.86)	/

Model component (k = 1)	Parameter	True value	Conditional ignorability	Marginal ignorability
	γ_{13}	−1.07	−1.17 (−1.49, −0.87)	/
	σ_1^2	1.00	1.01 (0.86, 1.15)	/

Note. Posterior means (95% CrIs) are presented for conditional ignorability and marginal ignorability. 10 000 iterations were run with a burn-in of 1000.

[a]Multinomial logit parameters comparing cluster 1 to cluster 3 (reference cluster).

Slashes (/) indicate that estimates are not applicable.

TABLE 3

Results from the 2-cluster model applied to the Nurture data

Model component	Parameter	Variable	Cluster 1 (37.0%)[a] Est. (95% CrI)	Cluster 2 (63.0%)[a] Est. (95% CrI)
MSN regression	β_{k11}	3 mo	−0.33 (−0.48, −0.18)	0.26 (−0.04, 0.53)
	β_{k21}	6 mo	−0.22 (−0.37, −0.05)	0.54 (0.17, 0.86)
	β_{k31}	9 mo	−0.20 (−0.52, 0.11)	0.10 (−0.47, 0.56)
	β_{k41}	12 mo	−0.35 (−0.45, −0.27)	0.80 (0.37, 1.11)
	β_{k12}	FS (3 mo)	−0.55 (−0.68, −0.40)	−0.10 (−0.28, 0.12)
	β_{k22}	FS (6 mo)	−0.40 (−0.56, −0.23)	0.08 (−0.08, 0.32)
	β_{k32}	FS (9 mo)	−0.22 (−0.41, −0.03)	−0.15 (−0.27, −0.02)
	β_{k42}	FS (12 mo)	−0.33 (−0.50, −0.12)	−0.13 (−0.26, −0.06)
	β_{k13}	BW (3 mo)	−0.02 (−0.09, 0.06)	0.07 (−0.05, 0.16)
	β_{k23}	BW (6 mo)	−0.03 (−0.11, 0.04)	0.03 (−0.09, 0.11)
	β_{k33}	BW (9 mo)	−0.03 (−0.13, 0.06)	0.11 (−0.07, 0.29)
	β_{k43}	BW (12 mo)	−0.03 (−0.11, 0.04)	0.06 (−0.05, 0.14)
	β_{k14}	BF (3 mo)	0.41 (0.29, 0.51)	0.07 (−0.15, 0.22)
	β_{k24}	BF (6 mo)	0.46 (0.36, 0.55)	0.04 (−0.14, 0.20)
	β_{k34}	BF (9 mo)	0.62 (0.30, 0.91)	0.03 (−0.05, 0.12)
	β_{k44}	BF (12 mo)	0.17 (−0.21, 0.55)	0.04 (−0.12, 0.24)
	β_{k15}	TC (3 mo)	0.01 (−0.03, 0.06)	−0.02 (−0.09, 0.05)
	β_{k25}	TC (6 mo)	0.02 (−0.02, 0.06)	−0.07 (−0.13, −0.02)
	β_{k35}	TC (9 mo)	0.02 (−0.03, 0.07)	0.00 (−0.06, 0.06)
	β_{k45}	TC (12 mo)	0.01 (−0.03, 0.06)	0.17 (−0.01, 0.35)
	α_{k1}	Skewness (3 mo)	0.00 (−0.12, 0.11)	0.16 (−0.23, 0.41)
	α_{k2}	Skewness (6 mo)	−0.02 (−0.15, 0.1)	−0.53 (−0.80, −0.17)
	α_{k3}	Skewness (9 mo)	−0.02 (−0.16, 0.13)	0.05 (−0.32, 0.44)
	β_{k4}	Skewness (12 mo)	−0.03 (−0.16, 0.10)	−0.07 (−0.41, 0.28)
Multinomial logit[b]	δ_{k1}	Intercept	1.03 (0.79, 1.25)	Reference

Model component	Parameter	Variable	Cluster 1 (37.0%)[a] Est. (95% CrI)	Cluster 2 (63.0%)[a] Est. (95% CrI)
	δ_{k2}	BW	0.03 (−0.09, 0.15)	Reference
	δ_{k3}	Race (black)	−0.02 (−0.29, 0.27)	Reference
	δ_{k4}	Gender (female)	0.90 (0.65, 1.27)	Reference
Missing data	γ_{k1}	Intercept	0.37 (0.32, 0.41)	−0.16 (−0.19, −0.14)
	γ_{k2}	BW	0.05 (−0.51, 0.59)	0.03 (−0.14, 0.19)
	γ_{k3}	Gender (female)	0.80 (0.25, 1.57)	−0.04 (−0.41, 0.30)
	γ_{k4}	Race (black)	0.35 (−0.56, 1.37)	−0.60 (−1.02, −0.20)
	σ_k^2	Random intercept variance	1.34 (0.86, 1.74)	1.11 (0.79, 1.43)

Note. Posterior means (95% CrI) are presented in each cluster for the effects of time, food security during pregnancy (FS), birth weight for gestational age z-score (BW), an indicator for any breastfeeding throughout the study period (BF), and total number of children in the household (TC). The effects of time, FS, BW, BF, and TC were allowed to vary over time, yielding separate estimates for each 3-month visit. Posterior means (95% CrI) are also given for effects of birth weight for gestational age *z*-score, race, and gender in the multinomial logit clustering and missing data models.

[a]Posterior mean percent in each cluster.

[b]With only two clusters, this reduces to a conventional logistic model.