Statistics
in Medicine

Research Article

Received 10 May 2013 Accepted 16 March 2014 Published online 10 April 2014 in Wiley Online Library

(wileyonlinelibrary.com) DOI: 10.1002/sim.6166

A Bayesian approach to functional mixed-effects modeling for longitudinal data with binomial outcomes

Stephanie Kliethermes[a*†] **and Jacob Oleson**[b]

Longitudinal growth patterns are routinely seen in medical studies where individual growth and population growth are followed up over a period of time. Many current methods for modeling growth presuppose a parametric relationship between the outcome and time (e.g., linear and quadratic); however, these relationships may not accurately capture growth over time. Functional mixed-effects (FME) models provide flexibility in handling longitudinal data with nonparametric temporal trends. Although FME methods are well developed for continuous, normally distributed outcome measures, nonparametric methods for handling categorical outcomes are limited. We consider the situation with binomially distributed longitudinal outcomes. Although percent correct data can be modeled assuming normality, estimates outside the parameter space are possible, and thus, estimated curves can be unrealistic. We propose a binomial FME model using Bayesian methodology to account for growth curves with binomial (percentage) outcomes. The usefulness of our methods is demonstrated using a longitudinal study of speech perception outcomes from cochlear implant users where we successfully model both the population and individual growth trajectories. Simulation studies also advocate the usefulness of the binomial model particularly when outcomes occur near the boundary of the probability parameter space and in situations with a small number of trials. Copyright © 2014 John Wiley & Sons, Ltd.

Keywords: hierarchical Bayesian; cochlear implants; growth curves; speech perception; longitudinal

1. Introduction

A defining characteristic of many longitudinal studies, regardless of the distribution of the outcome, is that they contain multiple measurements on the same individuals over some period of time. In certain situations, it may be of interest to assume that the basic unit of measurement for these individuals is not an observation at a specific time point but rather an individual's curve created by the collection of observations. Current methods in functional data analysis (FDA) analyze these longitudinal applications with the assumption that the data are measurements from smooth, infinite-dimensional curves, and observations at each time point are noise along the underlying curve [1, 2]. Classical FDA assumes large numbers of observations per individual to fit the infinite-dimensional curve; yet, this is often not feasible for longitudinal studies as they can be expensive and time consuming. Many longitudinal studies result in a finite number of time points and likely include missing observations. Functional mixed-effects (FME) models are flexible enough to handle longitudinal data with limited outcomes and missing data [2], but the models generally assume that the outcome of interest is normally distributed or at least that it can be modeled as such. The assumption of normality often simplifies necessary statistical methods and eases interpretation of results. Percentage data are an example of one such outcome, which is often assumed to be normally distributed for ease of analysis. However, percentage data are more accurately described

[a]*Department of Medicine, Stritch School of Medicine, Loyola University Chicago, 2160 S. First Ave, Maywood, IL 60153, U.S.A.*
[b]*University of Iowa, 100 CPHB, Room 300, 145 N. Riverside Drive, Iowa City, IA 52242-2007, U.S.A.*
**Correspondence to: Stephanie Kliethermes, Department of Medicine, Stritch School of Medicine, Loyola University Chicago, 2160 S. First Ave, Maywood, IL 60153, U.S.A.*
[†]*E-mail: skliethermes@lumc.edu*

as a summation of independent Bernoulli outcomes at each longitudinal time point and thus would be more accurately considered binomially distributed data if the independent Bernoulli assumption can be justified.

When linear regression techniques, which assume normality, are used to model binomial data, various problems can arise. First, the homoscedasticity assumption of the error variance is violated because the probability of success differs at each time point. As a result, the variance of the parameter estimates may be biased, which can result in incorrect inferences. In addition, the error terms are not normally distributed, which can undermine the validity of hypothesis testing in studies with small sample sizes. Although the inherent heteroscedasticity can be fixed via weighted least squares, and normality can be asymptotically achieved with large enough sample sizes, a major problem with assuming normality for percentage data still remains. When using linear models to model percentage data, there is no restriction on the prediction falling within the possible range of probabilities. That is, predictions of probabilities greater than 100% and less than 0% are possible yet not interpretable for true binomial data. When this happens, investigators will often truncate the predicted values at 0% or 100%, potentially biasing the accuracy of results [3].

In this paper, we propose a Bayesian binomial FME model that extends the FME model introduced by Wensheng Guo [2] by specifically accounting for outcomes that are binomially distributed. At each time point, an individual undergoes a set number of 'N trials' resulting in x successes and $N - x$ failures. In the context of growth, one may reasonably assume that the number of successes for a given individual may improve for a period of time and then remain steady once maximum growth or learning has occurred. Thus, for binomial trials, he or she would obtain more successes at each ensuing time point until he or she reaches a point of maximum threshold where the number of successes no longer improves for all remaining time points. Taken together, the scores at each time point result in an observed growth curve of that individual's successes and can be regarded as noisy observations around the individual's true underlying curve. The goal is then to use these collected observations to estimate the true underlying curve for an individual.

Functional linear models have used smoothing splines, regression splines, B-splines, and kernel methods as a way to estimate functional (fixed) coefficients, but the FME model extends this framework to a longitudinal setting by adding functional random effects to allow for variability among individuals [2, 4]. Guo's FME model has the flexibility to model both subject-specific and population curves over time because of functional components for the fixed and random effects. This model, with a normal outcome, has direct parallels to the well-known linear mixed-effects (LME) models because of its use of smoothing splines. In fact, special cases of this functional model reduce to widely known models used in longitudinal data analysis such as the LME model, nonparametric longitudinal models, and functional extensions of analysis of variance [2]. A semiparametric modeling approach is also possible in this setting. If one has *a priori* knowledge of the parametric form of the curve, then either the random effects or fixed (population) parameters could incorporate a non-linear parametric function such as the Gompertz or logistic curves. These non-linear functions could also be used to impose constraints on the curves such as creating monotonically non-decreasing curves. In this paper, we prefer to be more general and thus model both the population and subject-specific curves using fully nonparametric techniques.

As previously mentioned, there are many ways to consider modeling percentage data—each with its own benefits and drawbacks. The number of successes can be modeled directly and assumed to be normally distributed. Although perhaps the easiest and most direct approach, this method allows for the possibility of predicting a negative number of successes or a number of successes greater than the number of trials. Thus, predictions may be unrealistic. As a result, we propose a hierarchical binomial FME model that both bounds the possible outcomes to realistic values and accounts for variability in the outcome score. In this paper, we will compare our proposed binomial approach to the normal FME model, which is equivalent to a nonparametric mixed-effects model but estimated from the Bayesian perspective; several strengths and weaknesses of each approach will be discussed.

In Section 2, we introduce FME models with Section 2.1 devoted to a brief introduction of normal FME models. In Section 2.2, we extend the normal FME model to account for binomially distributed data (e.g., percentage data). We apply our model to hearing data for cochlear implant (CI) patients in Section 3, and in Section 4, we present simulation studies to evaluate the performance of the approaches mentioned earlier with different population means. Finally, we conclude this paper with a discussion in Section 5.

2. Functional mixed-effects models

2.1. Normal functional mixed-effects model

The FME model is a nonparametric mixed-effects model which uses cubic smoothing splines to estimate both fixed and random effects [2]. This nonparametric model is a natural extension of parametric LME models where the fixed effects and random effects are now modeled as functions. Similar to LME models, FME models can handle complex designs and correlation structures, but their greatest appeal is in their ability to allow the data to determine the shape of the curve instead of a pre-specified function and the ease in which they can be estimated using Bayesian methodology. When group-average profiles and subject-specific deviations cannot be adequately modeled via parametric relationships between the outcome of interest and the time variable, nonparametric functions such as splines are often considered a viable alternative.

At the design points, the FME model can be formulated as an LME model, although it allows for nonparametric estimation of the fixed and random effects via cubic splines. Let t_{ij} be the design time points for the ith individual at the jth time point where $i = 1, 2, \ldots, n$ and $j = 1, 2, \ldots, m_i$ with m_i representing the number of time points for the ith individual. Then, y_{ij} denotes the response of subject i at time j. The model can be written, most generally [2] as follows:

$$y_{ij} = \boldsymbol{X}_{ij}\boldsymbol{\beta}\left(t_{ij}\right) + \boldsymbol{Z}_{ij}\alpha_i\left(t_{ij}\right) + e_{ij}, e_{ij} \sim N\left(0, \sigma_e^2\right) \tag{1}$$

Here, $\boldsymbol{X}_{ij} = (x_{ij1}, \ldots, x_{ijp})$ and $\boldsymbol{Z}_{ij} = (z_{ij1}, \ldots, z_{ijq})$ are known $1 \times p$ and $1 \times q$ design matrices, respectively. The vector $\boldsymbol{\beta}(t) = \{\beta_1(t), \ldots, \beta_p(t)\}^T$ is a $p \times 1$ vector of population functions of time t, $\boldsymbol{\alpha}_i(t) = \{\alpha_{i1}(t), \ldots, \alpha_{iq}(t)\}^T$ is a $q \times 1$ vector of subject-specific functions of time t, and e_{ij} is the measurement error. The functions in vectors $\boldsymbol{\beta}(t)$ and $\boldsymbol{\alpha}_i(t)$ are denoted as $\boldsymbol{\beta}\left(t_{ij}\right)$ and $\boldsymbol{\alpha}_i\left(t_{ij}\right)$, respectively, when evaluated at a specific design time point for a specific individual. The subject-specific functions are modeled as realizations of smooth processes—typically Gaussian. Finally, p and q are the number of population and subject-specific parameters modeled as functions, respectively.

When only one covariate is considered, say time, Equation (1) can be written in a simpler form:

$$y_{ij} = \beta\left(t_{ij}\right) + \alpha_i\left(t_{ij}\right) + e_{ij}, e_{ij} \sim N\left(0, \sigma_e^2\right) \tag{2}$$

The function $\beta(t)$ is a population mean function relating time to the outcome and $\beta\left(t_{ij}\right)$ represents the function evaluated at the jth time point for the ith individual. Similarly, $\alpha_i(t)$ is the departure of the ith individual function from the population function, $\beta(t)$, and $\alpha_i\left(t_{ij}\right)$ represents the ith individual's deviation function evaluated at the jth time point. Finally, e_{ij} is the measurement error. For the purpose of this paper, time is the only covariate included in the models, and thus, model (2) will be used as the reference model. This model can be regarded as a generalization of an LME model where the population function (i.e., fixed effects) and subject-specific functions (i.e., random effects) are treated as smooth curves over time. We are primarily interested in estimating the population–average curve, $\beta(t)$, as well as the subject-specific deviation curves, $\alpha_i(t)$ for each individual. Then, we can fully specify the ith individual's curve by combining the population curve and the subject-specific deviation curve for that individual: $\beta(t) + \alpha_i(t)$.

Guo [2] proposed a Bayesian estimation procedure for these curves via cubic smoothing splines; however, if $\beta(t)$ and $\alpha_i(t)$ are modeled with parametric functions, then models (1) and (2) reduce to an LME model. His technique draws on the relationship between cubic smoothing splines and linear mixed models explained by Wahba [5] and results in both functions (random and fixed effects) being modeled as cubic smoothing splines. See [2] or [6] for details on the estimation procedure of this model. Although Guo uses cubic smoothing splines to estimate the functions, this is primarily for convenience and ease of estimation. Extensions to other splines (e.g., B-splines) and kernel bases such as local polynomials can be used to obtain the smoothing. We consider cubic splines in this paper for ease of implementation and as a natural extension to Guo's model, which focused specifically on this type of smoothing splines.

2.2. Binomial functional mixed-effects model

2.2.1. Data model. Our proposed binomial hierarchical model assumes the observed number of successes, y_{ij}, for $i = 1, \ldots, n$ individuals and $j = 1, \ldots, m_i$ time points are distributed binomially

and models a function of the probability (π_{ij}) of achieving the observed outcome as opposed to directly modeling the outcome (y_{ij}). That is, the resulting smooth growth curves are estimates of the probability of achieving an outcome at each time point as opposed to estimates of the observed outcomes. An additional benefit in assuming a binomial setting is that this proposed model has the ability to capture the variability (uncertainty) of an individual outcome at each time point by treating it as a random outcome from the estimated probability of obtaining the outcome (π_{ij}). By placing a prior distribution on the $logit(\pi_{ij})$ values and treating them as random, the posterior estimates will take into account the binomial nature of the original data. The proposed hierarchical model is as follows:

$$y_{ij}|\pi_{ij} \sim Bin(N, \pi_{ij}) \tag{3}$$

$$logit(\pi_{ij}) = \beta\left(t_{ij}\right) + \alpha_i\left(t_{ij}\right) + e_{ij}$$

where $i = 1, \ldots, n$ individuals and $j = 1, \ldots, m_i$ time points. The total number of trials for an individual at a given time point is denoted by N, and, for simplicity, N is assumed to be constant at each time point and the same for all individuals. As introduced in Section 2.1, $\beta(t)$ represents the population-averaged profile and $\beta\left(t_{ij}\right)$ is the evaluation of this function for the ith individual at the jth time point. The functions $\alpha_i(t)$ are subject-specific random effects specifying individual deviations from the population function, and $\alpha_i\left(t_{ij}\right)$ are evaluations at specific time points. The random functions are modeled as realizations of smooth processes—typically Gaussian. For the scope of this paper, only one population function is modeled, and each individual only has one subject-specific function because time is the only covariate included in the model.

The measurement error of $logit(\pi_{ij})$ is captured by e_{ij} and is assumed to be normally distributed, $e_{ij} \sim N(\sigma^2)$. When using the log-normal framework to model the outcomes, the linear model contains a residual effect that includes other sources of variability not explained by the linear model such as higher-order interactions or unknown sources of variability in π_{ij}. The Bayesian hierarchical modeling approach is flexible enough that the beta-binomial approach could also be used to model the probabilities, resulting in a conjugate prior. However, the log-linear model approach is used because the inclusion of the fixed and random effects into the model is straightforward.

2.2.2. Process models. Although model (3) can accommodate a different number of time points per individual, for simplicity, we assume each subject has an equal number of time points, and we drop the i subscript within t_{ij}, where each subject is measured at time t_j for $j = 1, \ldots, m$. Then we can write $\boldsymbol{t} = (t_1, t_2, \ldots, t_m)^T$ as a vector of design time points, which is the same for all individuals.

A cubic smoothing spline is a smooth function with a continuous second derivative, which is estimated by minimizing a specific penalized least squares criterion. Wahba [5] proved that minimizing the penalized least squares criterion for a given function in the cubic spline framework is equivalent to putting a Bayesian prior of the form $f(t) = B_0 + B_1 t + w(t)$ onto the $\beta(t)$ and $\alpha_i(t)$ functions and estimating the functions by their posterior means. In a function of this form, B_0 and B_1 can be regarded as intercept and slope parameters, and $\boldsymbol{t} = (t_1, t_2, \ldots, t_m)^T$ is an $m \times 1$ vector of the design time points. The function $w(t)$ is a smooth (Gaussian) process with mean and covariance structure, $\boldsymbol{R}$, which ensures the smoothness of the posterior mean [2]. The realization of $w(t)$ can be thought of as the smooth deviation of $f(t)$ from the straight line $B_0 + B_1 t$. At the design time points, $f(t)$ is mathematically equivalent to a mixed-effects model, where B_1 and B_2 are the fixed effects and $w(t)$ represents the random effects.

When modeled this way, the posterior estimate of the population mean function $\hat{\beta}(t)$ can be represented by $E[\beta(t)\|\boldsymbol{y}]$ and the estimates of the subject-specific deviations $\hat{\alpha}_i(t) = E[\alpha_i(t\|\boldsymbol{y_i})]$. The population functions are estimated using all the data, whereas the subject-specific deviations primarily rely on data for the ith subject, although they do borrow strength from other subjects to estimate the random effects. Because both $\hat{\beta}(t)$ and $\hat{\alpha}_i(t)$ are cubic smoothing splines, $\hat{\beta}(t_{ij}) + \hat{\alpha_i}(t_{ij})$ can be considered a cubic smoothing spline as well [2, 5]. At the design time points, $\beta(\boldsymbol{t})$ and $\alpha_i(\boldsymbol{t})$ of the data model (3) can be written as the following LME models:

$$\beta(\boldsymbol{t}) = \boldsymbol{T}\boldsymbol{d} + \boldsymbol{u} \tag{4}$$

$$\alpha_i(\boldsymbol{t}) = \boldsymbol{M}_i\boldsymbol{c}_i + \boldsymbol{v}_i \tag{5}$$

The matrices $\boldsymbol{T} = \{\mathbf{1}_m, \boldsymbol{t}\}$ and $\boldsymbol{M}_i = \{\mathbf{1}_m, \boldsymbol{t}\}$ are $m \times 2$ design matrices; $\boldsymbol{t}$ is as defined earlier, and $\mathbf{1}$ is an $m \times 1$ vector of ones, which serves as a place holder for the population and subject-specific intercepts. The 2×1 vector, $\boldsymbol{d} = (B_0, B_1)^T$ contains the slope and intercept for the fixed effects (population) function and modeled with a large, diffuse variance, $\boldsymbol{d} \sim N_2(\mathbf{0}, \xi * \boldsymbol{I}_2)$. Similarly, $\boldsymbol{c}_i = (a_{i0}, a_{i1})^T$ are 2×1 vectors of the slopes and intercepts for the random effects functions and modeled as $\boldsymbol{c}_i \sim N_2(\mathbf{0}, \boldsymbol{Q})$, and $\boldsymbol{u}$ and$\boldsymbol{v}_i$ are random effects that control for the smoothness of the curve by estimating the departure of (4) and (5) from a straight line. They are specified as $\boldsymbol{u} \sim N_m(\mathbf{0}, \tau_u \boldsymbol{R}_u)$ and $\boldsymbol{v}_i \sim N_m(\mathbf{0}, \tau_v \boldsymbol{R}_v)$. The smoothing parameters τ_u and τ_v can be thought of as variance components of the mixed models (4) and (5), respectively, as they control the amount of smoothing in the population and subject-specific curves. When $\tau_u \to 0$ and/or $\tau_v \to 0$, a large amount of smoothing occurs, and the spline tends towards a straight line; however, as $\tau_u \to \infty$ and/or $\tau_v \to \infty$, little to no smoothing takes place. These smoothing parameters can be fixed by the investigator or estimated as part of the MCMC sampling algorithm.

When expressing this model via its functional components, Equations (4) and (5), and following the modeling procedure detailed earlier, we obtain a hierarchical data structure consisting of the overall data model, process models, and hyperpriors. Let $\boldsymbol{\pi}_i = (\pi_{i1}, \pi_{i2}, \ldots, \pi_{im})^T$ be an $m \times 1$ vector of probabilities for the ith individual, where $i = 1, \ldots, n$ individuals and $j = 1, \ldots, m$ time points. Similarly, let $\boldsymbol{y}_i = (y_{i1}, y_{i2}, \ldots, y_{im})^T$ be an $m \times 1$ vector containing the number of successes at each time point for the ith individual and $\boldsymbol{e}_i$ the measurement error for the ith individual. The hierarchical model can thus be written as follows:

$$\boldsymbol{y}_i | \boldsymbol{\pi}_i \sim Bin(N, \boldsymbol{\pi}_i)$$

$$logit(\boldsymbol{\pi}_i) = \boldsymbol{T}\boldsymbol{d} + \boldsymbol{u} + \boldsymbol{M}_i \boldsymbol{c}_i + \boldsymbol{v}_i + \boldsymbol{e}_i$$

Parameter models:

In a Bayesian hierarchical model, the remaining hyperparameters in the model must be specified. Let

$$\boldsymbol{Q} \sim IW(\upsilon, \boldsymbol{\Psi}) \text{ and}$$

$$\tau_u \tau_v \sim IG(\alpha, \beta)$$

Define $\boldsymbol{T}_j$ (1×2) and $\boldsymbol{M}_{ij}$ (1×2) as the jth rows of $\boldsymbol{T}$ and $\boldsymbol{M}_i$, respectively. As previously mentioned, $\boldsymbol{d} = (B_0, B_1)^T$ is a 2×1 vector containing the slope and intercept for the population function, $\boldsymbol{d} \sim N_2(\mathbf{0}, \xi * \boldsymbol{I}_2)$ with $\xi \to \infty$. Further, $\boldsymbol{c}_i = (a_{i0}, a_{i1})^T$ are 2×1 vectors of the slopes and intercepts for the individual level random effects functions, with $\boldsymbol{c} = \left(\boldsymbol{c}_1^T, \boldsymbol{c}_2^T, \ldots, \boldsymbol{c}_n^T\right)^T$ and $\boldsymbol{c}_i \sim N_2(\mathbf{0}, \boldsymbol{Q})$. The random effects are modeled via $\boldsymbol{u}_i = (u_1, u_2, \ldots, u_m)^T$ and $\boldsymbol{v}_i = (v_{i1}, v_{i2}, \ldots, v_{im})^T$, which control for the smoothness of the curves by expressing the departure of the population and subject-specific curves, respectively, from a straight line at the design time points. Let $\boldsymbol{u} = \left(\boldsymbol{u}_1^T, \boldsymbol{u}_2^T, \ldots, \boldsymbol{u}_n^T\right)^T$, where $\boldsymbol{u}_1 = \boldsymbol{u}_2 = \ldots = \boldsymbol{u}_n$ and $\boldsymbol{u}_i \sim N_m(\mathbf{0}, \tau_u \boldsymbol{R}_u)$, and let $\boldsymbol{v} = (\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_n)^T$, where each $\boldsymbol{v}_i \sim N_m(\mathbf{0}, \tau_v \boldsymbol{R}_v)$. The $m \times m$ covariance structures, $\boldsymbol{R}_u$, and $\boldsymbol{R}_v$ represent the variance–covariance matrix of an integrated Wiener process at the design points [2,7,8] for the population and subject-specific random effects, respectively. These matrices share the same structures and values because they are not determined by the data. The matrix $\boldsymbol{R}_u^* = \begin{bmatrix} \boldsymbol{R}_u & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \ddots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \boldsymbol{R}_u \end{bmatrix}$ is an $nm \times nm$ block diagonal matrix of the $\boldsymbol{R}_u$ variance–covariance. The hyperprior for $\mathbf{Q}$, the covariance structure for the subject-specific slopes and intercepts, is an inverse Wishart and the prior distributions for the smoothing parameters, τ_u and τ_v, both follow inverse gamma distributions. The hyperparameters $(\xi, \upsilon, \boldsymbol{\Psi}, \alpha, \beta)$ for these priors can be specified based on prior knowledge (or lack of existing knowledge) of the parameters.

The estimation procedure for the normal FME model is similar to the binomial FME estimation; however, the outcomes, y_{ij}, are directly modeled as opposed to the probabilities of the outcomes. Given that the Bayesian hierarchical model is largely composed of conjugate priors, the MCMC sampling is primarily via the Gibbs updates. The full conditional distributions for parameters in this model can be found in Appendix A. Note that the full conditionals for the normal model mimic those of the binomial model, although a full conditional distribution is not specified on the y_{ij} in the normal model as it is in the binomial hierarchical model.

2.2.3. Capturing uncertainty with posterior predictive estimates. Oftentimes, observed outcomes are highly variable estimates in and of themselves, and it may be of further interest to model this uncertainty in the values of the observed outcomes. That is, in the binomial setting, there may be a measurement error if the same subject was to repeat the same trials at the exact same time. In this setting, one may reasonably expect that the resulting outcome would be different than the first attempt. Our proposed Bayesian binomial FME model allows us to straightforwardly account for this uncertainty by calculating the posterior predictive distributions of the outcomes at each time point. When using the original point estimates of the scores and treating them as fixed in the Bayesian paradigm, we ignore our uncertainty of the scores themselves. The posterior predictive distribution allows us to account for our uncertainty in these scores by predicting new scores based upon the probability of an individual achieving a specific score at a specific time point. The posterior predictive distribution of the data $\boldsymbol{y}$ can be represented as follows:

$$p\left(\boldsymbol{y}_{(new)}|\boldsymbol{y}\right) = \int p\left(\boldsymbol{y}_{(new)}|\boldsymbol{\pi}, \boldsymbol{y}\right) p\left(\boldsymbol{\pi}|\boldsymbol{y}\right) d\,\boldsymbol{\pi}$$

where $\boldsymbol{y}_{(new)}$ represents the newly predicted data, $\boldsymbol{y}$ is the original observed data, and $\boldsymbol{\pi}$ are the corresponding probabilities of observing $\boldsymbol{y}$. Although we cannot typically derive the posterior predictive distribution directly, we can first draw estimates of $\boldsymbol{\pi}$ from its posterior distribution and then simulate $\boldsymbol{y}_{(new)}$ from the sampling distribution of $\boldsymbol{\pi}$ [9]. By sampling from the posterior distribution of $\boldsymbol{\pi}$, we can easily estimate $\boldsymbol{y}_{(new)}$, which is independent of the observed data, $\boldsymbol{y}$. In the binomial FME model,

$$\left[\boldsymbol{y}_{(new)}|\boldsymbol{y}\right] \sim Bin(N, \boldsymbol{\pi})$$

where $logit(\boldsymbol{\pi})$, and thus $\boldsymbol{\pi}$, is being estimated at each iteration of the MCMC routine. Therefore, with every updated estimate of $logit\,(\boldsymbol{\pi})$, we can obtain an estimate of $\boldsymbol{y}_{(new)}$ by drawing from this distribution, resulting in a posterior predictive distribution for the new outcomes.

Although the posterior predictive distribution is often used in Bayesian analysis as a goodness-of-fit measurement to assess the accuracy of models, we use it here to construct credible intervals representing the uncertainty in our original data. Typically, if the data we actually observed reflects the pattern of our posterior predictive distribution, we have evidence that we chose our model and prior distributions well [9]. Yet by using the posterior predictive distribution to come up with a credible interval for the true underlying curve, we can express our uncertainty in an estimate of a curve that is subject to noise. This approach parallels the use of confidence and prediction bands in frequentist analyses. With slight sampling error, the mean of each $y_{i,j(new)}$ posterior distribution will be the same as the mean of the posterior distribution of π_{ij}. But the credible interval of the posterior predictive outcomes will account for the variability in the observed outcome score itself. This results in a 'prediction' credible interval that is wider than the credible interval for the actual probabilities that are being modeled in the basic binomial FME model.

3. Application

3.1. Background

The motivation for the proposed binomial FME model arose out of a desire from a group of researchers at the University of Iowa to more accurately predict growth curves in speech and hearing applications where outcome data are often not normally distributed. Parametric models often do not fit speech and hearing outcome data well because the data tend to asymptote at the upper end of the growth trajectory. Higher-ordered polynomials can sometimes accurately fit the data; however, interpretation of parameter estimates is often lost in the process, and these models can be difficult to fit. The original FME model is a viable approach to modeling these data because it relaxes the parametric assumptions of both the fixed and random effects and can be estimated via hierarchical Bayesian methods, which ease the interpretation and estimation of parameters of interest. Yet, the normality assumption of the FME model still does not accurately describe some outcome data. In this section, we apply the proposed binomial FME model to actual percentage data and compare the results to those obtained via the original FME model, which assumes normality.

The Iowa/Nucleus Hybrid FDA data originated from a clinical trial led by Dr Bruce Gantz at The University of Iowa between the years 1999 and 2008. Patients with severe to profound hearing loss were implanted with the Iowa/Nucleus 10-mm Hybrid CI, and their improvement in speech processing was

measured over time. The Hybrid dataset consists of 87 patients who were initially enrolled in the clinical trial [10, 11]. Before the initial implantation, each subject was given a baseline consonant–nucleus–consonant (CNC) word score test. This test consisted of two lists of 50 monosyllabic words, which were digitally recorded and played to the patient. The patient was then asked to repeat the word back to an investigator under various settings involving hearing aids and the CI. The number of words the patient repeated correctly (out of 100 possible words) was recorded and serves as the primary outcome for our proposed binomial FME model. It is assumed that each word is a Bernoulli trial, and the trials are independent. Follow-up CNC tests were given at approximately 3-month intervals in the first year with the length in-between testing increasing to 6 months in the second year. Because of an excessively large amount of missing data in the CNC scores at 9 months post-operation, this month was removed from consideration in all analyses, resulting in measurements at the following time points: baseline and 3, 6, 12, 18, and 24 months post-implantation.

It is assumed that individuals implanted with the device will generally improve on the CNC test over time, and it is of interest to predict both the individual and population growth curves associated with these measures. Standard parametric mixed models are too restrictive in measuring this type of data given the shape of the growth curves. And current nonparametric methods fall short in accurately modeling the binomial outcomes for both subject-specific and population curves, or they are very difficult to implement. Thus, the FME model is a viable approach to modeling this set of data. Because the normality assumption of the FME model may not accurately capture the data, we compare results from the normal model to results obtained from our binomial FME model and the predictive binomial FME model. Figure 1 displays the raw data for the 87 patients in the dataset with the solid black line representing the population-averaged growth profile.

3.2. Missing data

As with many longitudinal studies involving human subjects, this dataset is not immune to missing observations. Overall, 33% of the data are missing, and after 12 months (1 year follow-up), the amount of missing data drastically increases. No missing data occur at baseline, but the percent missing increases to 4.9% at 3 months, 6.0% at 6 months, and 9.6% after 12 months. At 18 months, the amount of missing data more than doubles to 24.3% before reaching 33% after 24 months. The missing data at the latter time points are because this was a multi-center trial, and only individuals at one site (The University of Iowa) were followed up for the full 24 months; therefore, we assume the missingness mechanism to be missing at random [12]. With the Bayesian approach to modeling, all missing data can be treated as unknown latent parameters in the model, and we can estimate each missing outcome at the same time we estimate all other model parameters.

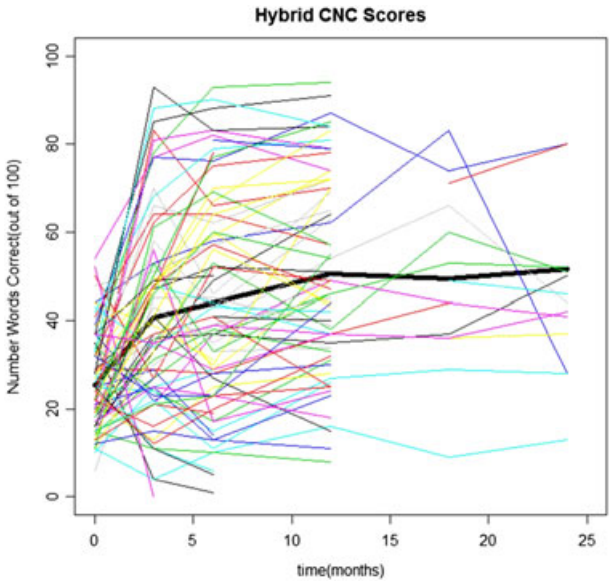

Figure 1. Observed CNC outcome curves for all individuals in CNC dataset.

The missing values for an individual, $\boldsymbol{y}_{i,mis}$, are estimated based on the observed values for that individual using functional multiple imputation techniques [6]. Suppose $\boldsymbol{y}_i = (\boldsymbol{y}_{i,mis}, \boldsymbol{y}_{i,obs})$, and $n_{i,mis}$ is the number of missing observations for an individual. The vector $\boldsymbol{y}_{i,mis}$ is an $n_{i,mis} \times 1$ vector of missing values to be estimated. For the normal case, He *et al.* [6] proposed a method for estimating the posterior distribution of each missing value via the Gibbs sampling as follows:

$$\left[\boldsymbol{y}_{i,mis} \| \boldsymbol{y}_{i,obs}, parameters\right] \sim N\left(\boldsymbol{w}_{i,mis}, \sigma_{\varepsilon^2} \boldsymbol{I}_{\boldsymbol{n_i}_{,mis}}\right)$$

where $\boldsymbol{w}_{i,mis}$ is the sub-vector of $\boldsymbol{w}_i = (\boldsymbol{T}\boldsymbol{d} + \boldsymbol{u}_i) + (\boldsymbol{M}_i \boldsymbol{c}_i + \boldsymbol{v}_i)$ corresponding to the ith subject's missing values. The $\boldsymbol{w}_i$ vectors utilize information from both the individual data and the population data to estimate the missing $\boldsymbol{y}_{i,mis}$. Starting values for this Gibbs sampler are given by implementing the last observation moved forward imputation method, although other reasonable approaches can be used instead to obtain starting values. For the proposed binomial model, instead of sampling from a normal distribution, we sample from a binomial distribution:

$$\left[\boldsymbol{y}_{i,mis} \| \boldsymbol{y}_{i,obs}, parameters\right] \sim Bin(N, \boldsymbol{\pi}_{i,mis})$$

where $\boldsymbol{\pi}_{i,mis}$ is the sub-vector of $\boldsymbol{\pi}_i = exp\{\boldsymbol{w}_i\} / (1 + exp\{\boldsymbol{w}_i\})$ and corresponds to the ith subject's missing values. Each iteration of the MCMC algorithm involves the estimation of model parameters, variance components, and missing data.

3.3. Model computing and MCMC setup

All analyses and estimation routines were implemented using R Version 2.15.1 [13] and a Windows 7 operating system. In addition to the base package of R, the following packages were used in our MCMC algorithms and diagnostic routines: mvtnorm [14], MCMCpack [15], bindata [16], Matrix [17], zoo [18], timeSeries [19], boa [20], and coda [21]. The Gibbs sampling routines were used for parameters with known full conditional distributions, and slice sampling was invoked for parameters with full conditional distributions of unknown form. These routines were chosen for convenience, and other MCMC sampling routines could be considered to improve computing efficiency and generalizability. The MCMC algorithm runtime is approximately 3.5 h for this dataset with 5000 iterations.

We use vague hyperpriors for all parameters in the model to reflect lack of pre-existing information on the parameters. In all models (normal, binomial, predictive binomial), the same priors are used. A $N_2(\boldsymbol{0}, 1000\boldsymbol{I_2})$ prior is used for all fixed effects parameters (population slopes and intercepts). The covariance matrix, $\mathbf{Q}$, has an independent $IW(0.01, 0.001\boldsymbol{I_2})$ prior, and both the smoothing parameters and error variances are given $IG(0.01, 0.01)$ priors. Sensitivity to our choice of priors was assessed within a reasonable range of noninformative hyperparameters, and similar results were achieved. The chains were run for 5000 iterations, and all parameters achieved burn-in after 500 iterations. Convergence was assessed and achieved by examining trace plots from various reasonable starting values and using the Geweke diagnostic criterion with $\alpha = 0.05$ [22, 23].

The estimated curves are compared using credible interval widths and a root mean square error (RMSE) measure. The average RMSE is a longitudinal version of the traditional RMSE and calculates the average RMSE of the estimated smooth curve from the actual smooth curve:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}\sum_{j=1}^{m}\left(\pi_{ij}^{*} - p_{ij}\right)^2}{n * m - 2}}$$

where π_{ij}^{*} is the estimated value of the smooth curve at time j for individual i, and $p_{ij} = y_{ij}/n$ is the observed proportion of success for the same individual at the same time point.

3.4. Results

The estimated population curves (Figure 2) for each model are relatively similar with the normal model providing the lowest *RMSE* of 1.63, followed by the binomial (and predictive binomial) with an *RMSE* of 1.80. Despite the differences in RMSE, the estimated means at each design time point are very similar, although the credible intervals for each model do differ. The credible interval widens at the latter time points in the normal model as a result of the large amount of missing data at these time points and

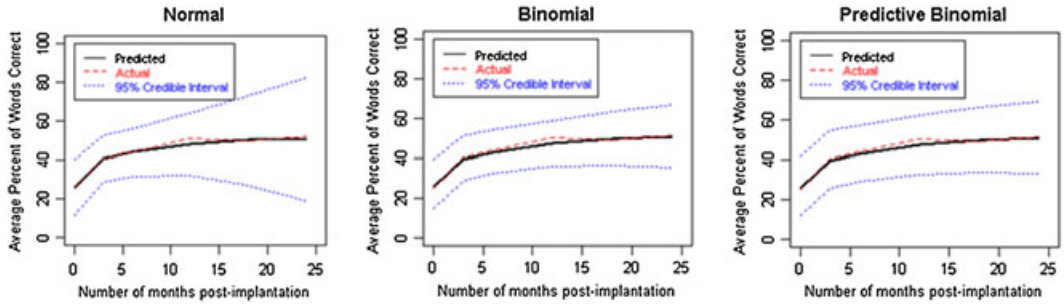

Figure 2. Estimated population curves with 95% credible intervals for normal, binomial, and predictive binomial FME models.

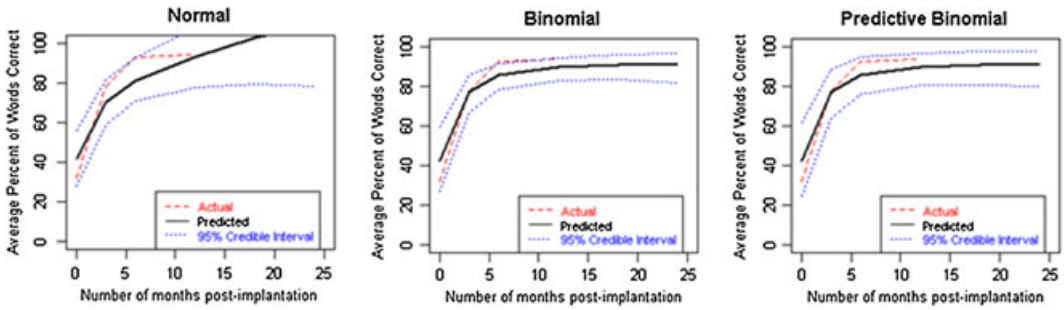

Figure 3. Estimated growth curves with 95% credible intervals for normal, binomial, and predictive binomial FME models for select individual with missing data.

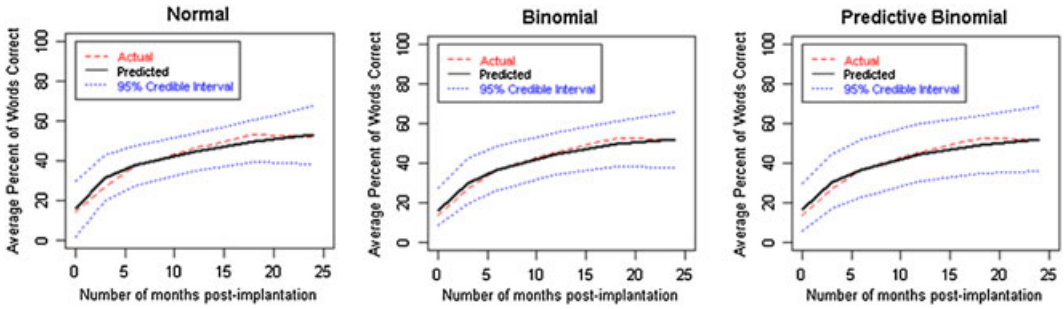

Figure 4. Estimated growth curves with 95% credible intervals for normal, binomial, and predictive binomial FME models for select individual with complete data.

the homoscedastic error assumption made with this distribution. The credible intervals of the binomial model are narrower, on average, than the normal model. Although the widths of the credible intervals still widen at the later time points, they remain much more stable than the widths associated with the normal model. The credible intervals for the predictive binomial model are the widest as would be expected.

Differences between the models, as well as limitations, are more apparent when considering the results for individual curves. The major limitation of the normal model is visible in Figure 3 as the model predicts an individual's curve beyond 100% correct at the last design time point. The average margin of error associated with the binomial curve is slightly smaller than that for the predictive binomial curve: 0.08 and 0.11, respectively. Both are smaller than the normal curve with an average width of 0.18. When considering the RMSE for each curve, the binomial and predictive binomial have the smallest average *RMSE* at 6.64 followed by an *RMSE* of 8.46 for the normal model.

For situations where individual curves are complete and fall in the middle of the parameter space, the predicted individual curves are more similar. In Figure 4, curves are estimated for an individual with complete data, and they are remarkably similar. The average *RMSE* for these curves are still smallest for the binomial curves (2.70 for binomial, and 2.77 for predicted binomial) and largest for the normal at 3.02. When comparing the average credible interval widths across models, the normal and binomial have

the smallest average width of 0.11, followed closely by the predicted binomial (0.14), which is expected to have the largest width.

Despite the larger width of the CIs for the predictive binomial curves, they are often within reasonable ranges of the binomial and normal models. This predictive binomial model will be especially useful in situations where an individual is newly implanted with CIs and the physician would like to predict where the patient's maximum growth will result without having to wait for a long period of time. Furthermore, the binomial and normal models provide very similar curves when the dataset is complete and with curves in the middle of the percent correct range. Depending on the situation, either curve could be used; however, the binomial model is foolproof from predicting curves outside of a feasible range of outcomes.

4. Simulation studies

Simulation studies are constructed to assess the benefit of the proposed binomial FME model over the traditional FME model that assumes normality of the original data. We also compare these models to our proposed predictive binomial model.

For each simulation study, 100 simulated datasets were randomly generated by pre-specifying all population and subject-specific parameters in the binomial FME model needed to create *logit* probability estimates for each individual at all time points. In constructing all datasets, an integrated Wiener covariance structure was used for the **R** matrix. The datasets consist of 20 individuals and have observations for both six and 20 time points within the same time frame to assess the impact that the number of time points has on the model. The *logit* probability estimates were then transformed to the probability scale and used to simulate outcomes from a binomial distribution with $N = 10, 20$, and 100 trials. The specific values used to generate the $logit(\pi_{ij})$ were chosen such that the resulting probability curves fall within the high end of the probability parameter space (70–100% correct) as well as in the middle of the probability parameter space (20–80% correct). To create datasets in the high end of the parameter space and in the middle of the parameter space, only the population slope and intercept parameters were altered. This procedure resulted in 12 combinations of 100 simulated datasets that were used for the simulation studies (Table I).

The same vague priors were used for all simulation studies. A $N(0, 1000)$ prior was used for the population slope and intercept, and the covariance matrix, **Q**, was given an independent $IW(0.01, 0.001\boldsymbol{I}_2)$ prior. Both smoothing parameters and the error variance were given $IG(0.01, 0.01)$ priors. Sensitivity to choice of prior was assessed for all simulations within a reasonable range of noninformative hyperparameters, and similar results were achieved. The chains were run for 5000 iterations, and all parameters achieved burn-in after 500 iterations. This was assessed by examining trace plots from various reasonable starting values and using the Geweke diagnostic criterion with $\alpha = 0.05$ [22, 23].

Table I. Twelve combinations of trials, time points, and parameter space location of growth curve for simulation studies.

Number of trials	Number of time points	Parameter space
10	6	High
		Middle
	20	High
		Middle
20	6	High
		Middle
	20	High
		Middle
100	6	High
		Middle
	20	High
		Middle

4.1. Goodness-of-fit measures

Because the purpose of these models is to estimate the true underlying curves, traditional Bayesian goodness-of-fit measures such as deviance information criterion (DIC) and posterior predictive p-values are inappropriate. These measures assess goodness-of-fit of the model to the observed data; however, in the functional framework, we treat the observed data as 'noise' around a true underlying smooth curve and wish to find the model that most accurately resembles the underlying curve, not the observed data. Furthermore, when used as a selection criterion, DIC (unlike its frequentist analog AIC), can only be used when the outcomes are the same because the pD is not invariant to transformations of the data [24, 25]. Similar to He *et al.* [6], we use a combination of measures to assess goodness-of-fit including average coverage probability, average credible interval width/2, average RMSE, and average relative bias.

The coverage probability calculation is averaged across all individual curves in each dataset and is the percentage of true probabilities along the underlying smooth curve that falls within the credible interval of the estimated smooth curve. The coverage probabilities in the high range are only calculated using observed values less than one because the binomial model cannot estimate a perfect score of one. The average credible interval width is the margin of error for all credible intervals averaged within individuals, and across individuals for each dataset. The average RMSE is the same as that introduced in Section 3, where π_{ij}^* is the estimated value of the smooth curve at time j for individual i; however, π_{ij} is now the true (simulated) underlying value of the smooth curve for the same individual at the same time point. The true π_{ij} are known in the simulation studies because they are used to simulate the 100 noisy datasets; however, these values are unknown in real data analyses, and so the noisy observations are used instead. Lastly, the average relative bias is used to evaluate performance in a manner similar to the RMSE. The 'average' relative bias measures the accuracy of the point estimator and is calculated as follows for one dataset:

$$RBias = \frac{\sum_{i=1}^{n}\sum_{j=1}^{m}\left(\left|(\pi_{ij}^* - \pi_{ij})\right|/\pi_{ij}\right)}{n*m} \times 100$$

In other words, the *RBias* takes the average of the absolute value of the bias divided by the real underlying probability ($RBias = avg(|bias/true|)$ for all individuals at all time points. These measures are calculated for each dataset in the simulation study, and the average across all 100 datasets is reported. From here on out, average will be dropped in discussion of the performance criteria, and the criteria should be understood to be averages within individual curves, across individuals, and among all datasets.

4.2. Comparison of different modeling approaches

The proposed binomial FME model is designed to more accurately model longitudinal growth outcomes that are binomially distributed than existing methods. The normal distribution is often a reliable approximation of binomial data when the number of trials is large enough, and so it is important to assess the value of the binomial FME model with both small and large numbers of trials. Further, outcomes that occur closer to the edge of the parameter space of π_{ij} (as opposed to the middle) will be handled differently depending on the assumed distribution. The normal distribution has the potential to predict probabilities outside of a valid parameter space, whereas the binomial model may better handle outcomes occurring near the edge of this parameter space (near 0 or 1).

When a binomial outcome occurs with only 10 trials, the asymptotic properties of the normal distribution as an approximation to the binomial may break down. The simulation results in Table II show the binomial model outperforming the normal model in RMSE and relative bias when the curve falls in the high end of the parameter space with a small number of time points ($N = 10, m = 6$). The binomial model has an average relative bias and *RMSE* of 2.42 ($SE = 0.27$) and 3.45 ($SE = 0.39$), respectively, whereas the normal model's average relative bias is 3.09 ($SE = 0.42$) and average *RMSE* is 4.16 ($SE = 0.70$). In terms of the coverage probability, all models appear to do sufficiently well (above 90% coverage) while estimating curves in the middle end of the parameter space; however, the normal model performs best among models in the high end of the parameter space. The average margin of error for the normal distribution is greater than the average margin of error for both the binomial models in the high end of the parameter space ($m = 6$ and $m = 20$); yet the normal RMSE is smaller than that of the binomial in the middle of the probability parameter space.

Table II. Simulation results using goodness-of-fit measures for estimated curves in all proposed models, where number of trials for each outcome measure equals 10 (100 simulations; 5000 iterations; burn-in = 500 iterations).

Time points	Parameter space	Model distribution	Coverage probability (*SE*)	CI width/2 (*SE*)	*RMSE* (*SE*)	Relative bias (*SE*)
6	High	Normal	93.1% (4.73%)	7.58 (0.83)	4.16 (0.70)	3.09 (0.42)
		Binomial	88.4% (7.43%)	4.47 (0.81)	3.45 (0.39)	2.42 (0.27)
		Predictive binomial	98.8% (1.65%)	10.88 (0.81)	3.45 (0.39)	2.42 (0.27)
	Middle	Normal	94.6% (3.37%)	13.76 (0.87)	6.95 (0.69)	13.76 (0.87)
		Binomial	90.4% (3.88%)	13.37 (0.51)	7.62 (0.68)	10.89 (1.08)
		Predictive binomial	99.9% (0.28%)	27.17 (0.59)	7.62 (0.68)	10.89 (1.08)
20	High	Normal	93.2% (2.80%)	5.47 (0.29)	2.95 (0.28)	2.29 (0.21)
		Binomial	90.6% (6.0%)	4.03 (0.42)	2.91 (0.31)	2.02 (0.20)
		Predictive binomial	99.4% (1.0%)	8.37 (0.39)	2.91 (0.31)	2.02 (0.20)
	Middle	Normal	95.2% (2.39%)	9.62 (0.32)	4.90 (0.41)	9.20 (0.94)
		Binomial	91.9% (2.73%)	10.10 (0.25)	5.24 (0.40)	10.82 (1.08)
		Predictive binomial	100% (0%)	27.54 (0.24)	5.25 (0.40)	10.82 (1.08)

Table III. Simulation results using goodness-of-fit measures for estimated curves in all proposed models, where number of trials for each outcome measure equals 20 (100 simulations; 5000 iterations; burn-in = 500 iterations).

Time points	Parameter space	Model distribution	Coverage probability (*SE*)	CI width/2 (*SE*)	*RMSE* (*SE*)	Relative bias (*SE*)
6	High	Normal	92.5% (4.08%)	5.57 (0.64)	3.20 (0.44)	2.38 (0.29)
		Binomial	90.6% (6.0%)	4.03 (0.42)	2.91 (0.31)	2.02 (0.20)
		Predictive binomial	99.4% (1.0%)	8.37 (0.39)	2.91 (0.31)	2.02 (0.20)
	Middle	Normal	94.3% (3.36%)	10.59 (0.82)	5.40 (0.51)	7.37 (0.77)
		Binomial	92.8% (3.65%)	10.78 (0.41)	5.85 (0.54)	8.31 (0.79)
		Predictive binomial	99.9% (0.23%)	20.27 (0.36)	5.85 (0.55)	8.31 (0.79)
20	High	Normal	93.7% (2.79%)	4.19 (0.22)	2.40 (0.27)	1.83 (0.19)
		Binomial	92.2% (3.39%)	3.54 (0.17)	2.42 (0.24)	1.78 (0.17)
		Predictive binomial	99.9% (0.05%)	9.28 (0.27)	2.42 (0.24)	1.78 (0.17)
	Middle	Normal	94.9% (2.41%)	7.35 (0.24)	3.85 (0.30)	7.13 (0.69)
		Binomial	92.5% (2.44%)	7.87 (0.15)	4.04 (0.31)	8.12 (0.73)
		Predictive binomial	100% (0%)	20.10 (0.14)	4.05 (0.31)	8.12 (0.73)

In some situations (e.g., small time points and small number of trials), where obtaining credible intervals that contain the true value is important, the predictive binomial appears to be the best choice. The RMSE and relative bias are the smallest, and one is nearly assured to capture the true underlying probability within the credible interval. The trade-off, of course, is a larger margin of error associated with the low end of the credible interval. Moreover, with 10 trials, the normal and binomial models both improve upon the RMSE, coverage probability, and relative bias when increasing the number of time points.

To assess the impact of the normality assumption on binomial outcomes, we chose to look at the situation where asymptotic normality is commonly assumed to be justified ($N = 20$). The results are presented in Table III. In this setting, the binomial model again appears to be the best choice with limited time points and outcomes near the high end of the parameter space (i.e., $m = 6$, parameter space=high). The coverage probability in the normal model is 92.5% compared to 90.6% in the binomial model. Both coverage probabilities increase as the number of time points increase. With $N = 20$ trials, the coverage probability in the middle of the parameter space is acceptable (greater than 90%) for all models with both $m = 6$ and $m = 20$ time points. In general, the goodness-of-fit measures improve with an increase in time points. Thus, when $N = 20$, the binomial model is still an appropriate model choice for growth curves with high success probabilities as it performs similarly to, and sometimes better than, the normal model.

Figure 5 provides a visual depiction of select simulation results in the high range of the parameter space. The actual smooth curve used to generate the observed data is represented by the black dotted

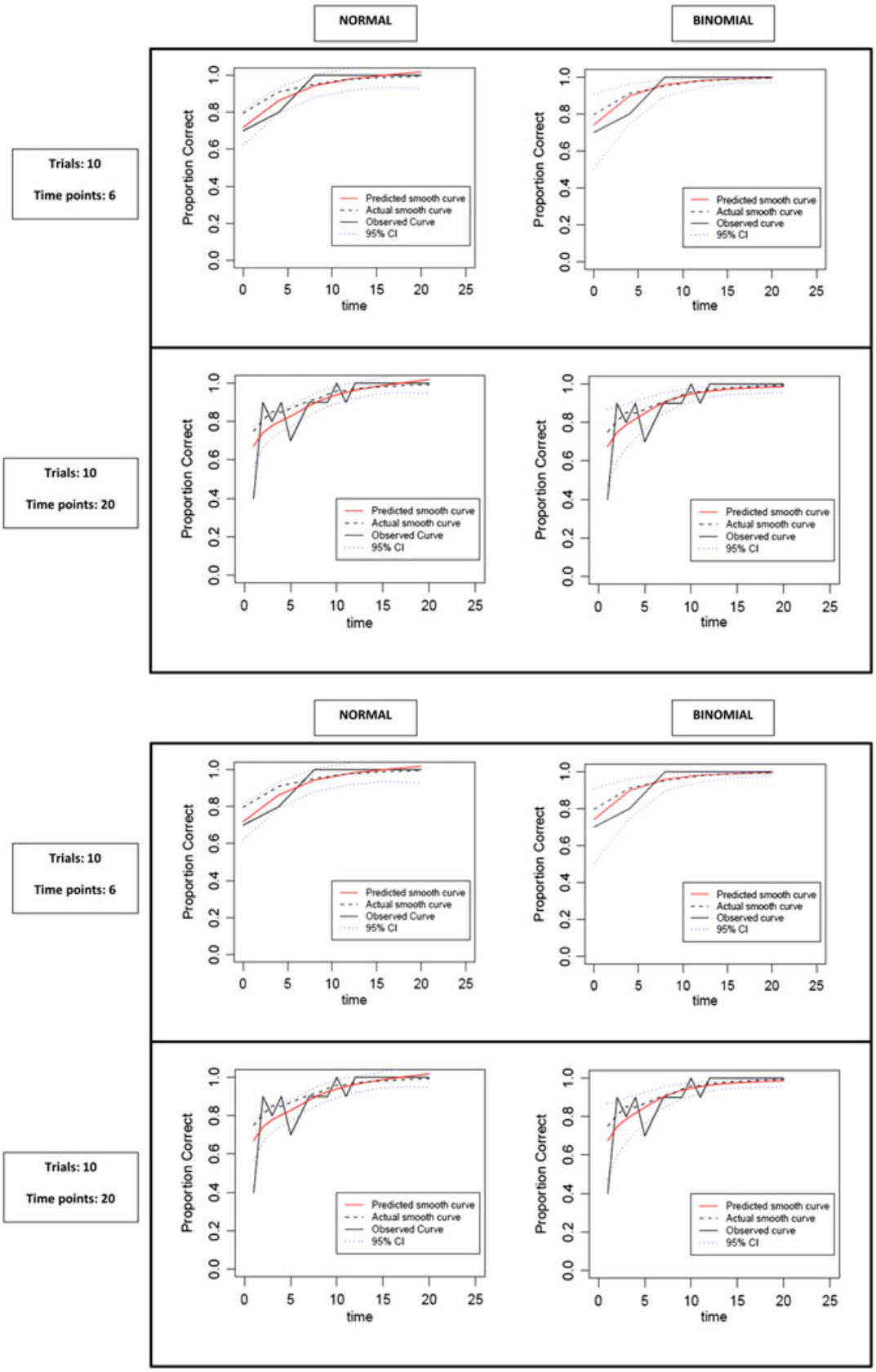

Figure 5. Estimated subject-specific growth curves with 95% credible intervals for simulated data in the high end of the parameter space and for differing combinations of trials and time points.

Table IV. Simulation results using goodness-of-fit measures for estimated curves in all proposed models, where number of trials for each outcome measure equals 100 (100 simulations; 5000 iterations; burn-in = 500 iterations).

Time points	Parameter space	Model distribution	Coverage probability (*SE*)	CI width/2 (*SE*)	*RMSE* (*SE*)	Relative bias (*SE*)
6	High	Normal	93.5% (2.82%)	3.18 (0.25)	1.71 (0.17)	1.26 (0.12)
		Binomial	92.8% (3.8%)	2.63 (0.19)	1.76 (0.19)	1.19 (0.12)
		Predictive binomial	99.6% (0.62%)	4.41 (0.13)	1.76 (0.19)	1.19 (0.12)
	Middle	Normal	94.6% (2.88%)	5.62 (0.42)	2.90 (0.25)	3.98 (0.36)
		Binomial	94.5% (2.58%)	6.02 (0.17)	3.05 (0.28)	4.28 (0.39)
		Predictive binomial	99.7% (0.46%)	9.91 (0.11)	3.05 (0.28)	4.28 (0.39)
20	High	Normal	94.1% (2.04%)	2.27 (0.10)	1.41 (0.10)	1.04 (0.07)
		Binomial	93.1% (2.89%)	2.07 (0.08)	1.42 (0.10)	1.02 (0.08)
		Predictive binomial	99.9% (0.07%)	4.67 (0.06)	1.42 (0.10)	1.02 (0.08)
	Middle	Normal	94.8% (2.26%)	3.92 (0.14)	2.11 (0.15)	3.95 (0.32)
		Binomial	94.3% (1.92%)	4.21 (0.08)	2.16 (0.15)	4.09 (0.34)
		Predictive binomial	99.9% (0.07%)	9.48 (0.05)	2.16 (0.15)	4.09 (0.34)

line, and the predicted smooth curve is represented by the red solid line. The blue dotted lines represent 95% credible intervals for the predicted smooth curve, and the black solid line is the simulated observed data. In all cases, the predicted smooth curve accurately smooths the observed data and follows a similar trajectory to the 'true' underlying smooth curve.

Lastly, we considered the situation where the number of trials was large ($N = 100$), and asymptotic normality could safely be assumed when estimating the growth curves under the normal model. In this setting (Table IV), both the normal and binomial models work remarkably similarly in terms of RMSE and relative bias. Although the normal model appears to be the best model via the chosen criteria, the binomial is not far behind. In the high range simulations with only six time points, the binomial model actually outperforms the normal model in all goodness-of-fit measures except for coverage probabilities. One argument for using the binomial model in this situation is that the parameter estimates are guaranteed to be within the parameter space, whereas the normal model still predicts probabilities greater than 100%. In the simulations involving 100 trials, the coverage probabilities perform well in both models and generally improve with more time points. Furthermore, the average margin of error of the credible intervals are smallest when $N = 100$. Interestingly, when the growth curves fall within the middle of the parameter space for probabilities, the average margin of error for the normal model is smaller than that of the binomial model. Certainly, when the number of trials is large, all models prove to be a viable choice for modeling depending on the goals and needs of the researcher.

5. Discussion

By implementing a Bayesian approach, and capitalizing on the hierarchical nature of the FME model, we are able to offer an alternative method for analyzing longitudinal growth curve data for binomial outcomes. The binomial FME model offers a new way to model percentage data, without assuming normality, by adding a hierarchical level to the estimation of Guo's FME model. In doing so, the predicted curves (and outcomes) are guaranteed to be within the parameter space appropriate for percentage data. Furthermore, we can account for the variability in the observed outcomes by using the predictive binomial model to estimate the outcomes given the MCMC estimated probabilities. This method results in a credible interval for the outcomes themselves rather than the credible interval for the probabilities in the binomial model and provides a more conservative approach for predicting differences in growth trajectories.

Our application to the speech perception data shows that all models perform similarly when modeling the population curves, although the normal model appears to be most affected by the large amount of missing data in the latter time points. Therefore, with the large number of trials at each time point, the normal distribution remains a reliable approximation to binomial data for estimation of population level curves. However, individual curves are more troublesome when modeled with the FME model assuming normality because this model can (and does) predict curves outside of the appropriate probability bounds. Given the large amount of missing data and variability among growth curves, the binomial

model (and posterior predictive extension) shows the most promise in accurately modeling growth for both individuals and populations. Functional models such as the one we propose are often used in larger datasets with a large number of time points; however, the speech data used in our application only has six time points. Although this model should be further tested on a dataset with a larger number of time points, we have shown that it can be useful in situations with fewer time points.

To begin addressing differences in model performance due to the number of time points, we have conducted simulation studies with only six time points and simulation studies with 20 time points. Simulation studies suggest that the binomial model is most appropriate in situations with binomial outcomes where number of trials and number of time points are small and the outcomes occur at the high end of the binomial parameter space (70–100%). In these situations, the asymptotic normality assumptions break down, and accurately specifying the true distribution of the binomial outcomes is necessary. When the number of trials is large, say 100, the normal model is a viable alternative even with a small number of time points in the high end of the parameter space. In general, the normal model proves to be a good approximation for percentage data when N is large (e.g., $N = 100$), and it even outperforms the binomial models in RMSE results. Yet, the binomial model often performs sufficiently well in these situations. One remaining limitation to the normal FME model is its prediction of outcomes/curves outside of the meaningful parameter space for percentage data. This results in modeling unrealistic growth curves; whereas the binomial model always ensures the predicted outcomes are between 0% and 100%, and thus, all curves can be accurately interpreted.

The binomial FME model shows promise in its ability to accurately model growth over time. Although the normal model appears to be a good approximation to the binomial model in certain cases, there are distinct advantages to choosing the binomial approach. Our next research goal is to extend this model such that covariates other than time are included in the model; this extension would allow us to compare growth curves between various groups of individuals in one run of the model and assess the role other variables may play in promoting or inhibiting growth.

Appendix A: Full conditionals

Let $\boldsymbol{w}_i$ be an $m\times 1$ mean vector such that $\boldsymbol{w}_i = (\boldsymbol{T}\boldsymbol{d} + \boldsymbol{u}_i)+(\boldsymbol{M}_i\boldsymbol{c}_i + \boldsymbol{v}_i)$ and $\boldsymbol{w} = \left(\boldsymbol{w}_1^T, \boldsymbol{w}_2^T, \ldots, \boldsymbol{w}_n^T\right)^T$. Given that the Bayesian hierarchical model is largely composed of conjugate priors, the Gibbs sampling is the primary MCMC sampling method used for both the population and subject-specific model parameters. However, the full conditional distribution of $logit\,(\boldsymbol{\pi}_i)$ does not follow any known distributional form, and so a slice sampler is used to estimate these parameters. Prior distributions for σ_ε^2,τ_u and τ_v are inverse gamma, $IG(\alpha,\beta)$, and the prior for $\mathbf{Q}$ is inverse Wishart, $IW(\boldsymbol{\Psi}, \upsilon)$. Hyperparameters can be chosen based on prior knowledge or kept noninformative to create diffuse priors that allow the data to drive the estimation. The full conditional distributions for all model parameters follow.

1. The conditional distribution of $\boldsymbol{z}_i = logit(\boldsymbol{\pi}_i)$ is

$$P\,(\boldsymbol{z}_i|\boldsymbol{w}_i, \boldsymbol{y}_i) \propto \mathrm{P}\,(\boldsymbol{y}_i|\boldsymbol{w}_i, \boldsymbol{z}_i)\,\mathrm{P}(\boldsymbol{z}_i)$$

$$[\boldsymbol{z}_i\,|\boldsymbol{w}_i] \propto \left(\prod_{j=1}^{m} \left(\pi_{ij}\right)^{y_{ij}} \left(1-\pi_{ij}\right)^{N-y_{ij}}\right) exp\left\{\frac{1}{2\sigma_\varepsilon^2}(\boldsymbol{z}_i - \boldsymbol{w}_i)'(\boldsymbol{z}_i - \boldsymbol{w}_i)\right\}$$

   Let $\boldsymbol{z} = \left(\boldsymbol{z}_1^T, \boldsymbol{z}_2^T, \ldots, \boldsymbol{z}_n^T\right)^T$ be a $nm \times 1$ vector of all estimated *logit* probabilities combined and $\boldsymbol{\pi}_i = (\pi_{i1}, \pi_{i2}, \ldots, \pi_{im})^T$. Because of autocorrelation among the population parameters, the population curve parameters were estimated simultaneously.
2. Thus, the conditional distribution of $\boldsymbol{du} = (\boldsymbol{d}^T, \boldsymbol{u}_1^T)^T$ given $\boldsymbol{\theta} = (\tau_u, \boldsymbol{c}, \boldsymbol{v}, \sigma_b^2, \sigma_\varepsilon^2)$ is

$$P\,(\boldsymbol{du}|\boldsymbol{\theta}, \boldsymbol{z}) \propto \mathrm{P}\,(\boldsymbol{z}\,|\,\boldsymbol{\theta}, \boldsymbol{du})\,\mathrm{P}(\boldsymbol{du})$$

$$[\boldsymbol{du}|\boldsymbol{\theta}] \sim N_{m+2}\left(\left(\frac{1}{\sigma_\varepsilon^2}\tilde{X}'\tilde{X} + \boldsymbol{P}\right)^{-1}\left(\frac{1}{\sigma_\varepsilon^2}\tilde{X}'(\boldsymbol{z} - \boldsymbol{Mc} - \boldsymbol{v}\right), \left(\frac{1}{\sigma_\varepsilon^2}\tilde{X}'\tilde{X} + \boldsymbol{P}\right)^{-1}\right)$$

   where $\boldsymbol{P} = diag\left[\frac{1}{\sigma_b^2}\boldsymbol{I}_2, (\tau_u \boldsymbol{R_u})^{-1}\right]$ and $\tilde{X} = [\boldsymbol{T}, \boldsymbol{I}_m] \otimes \mathbf{1}_{nm\times(2+m)}$.

3. The conditional distribution of $\boldsymbol{c_i}$ given $\boldsymbol{\theta} = (\boldsymbol{d}, \boldsymbol{u_i}, \boldsymbol{v_i}, \sigma_\varepsilon^2, \boldsymbol{Q})$ is

$$P(\boldsymbol{c}_i|\boldsymbol{\theta}\boldsymbol{z_i}) \propto \mathrm{P}(\boldsymbol{z_i}|\boldsymbol{\theta}, \boldsymbol{c}_i)\,\mathrm{P}(\boldsymbol{c}_i)$$

$$[\boldsymbol{c}_i|\boldsymbol{\theta}] \sim N_m\left(\left(\frac{1}{\sigma_\varepsilon^2}(\boldsymbol{M})'(\boldsymbol{M}) + (\boldsymbol{Q})^{-1}\right)^{-1}\left(\frac{1}{\sigma_\varepsilon^2}(\boldsymbol{M})'(\boldsymbol{z_i} - \boldsymbol{T}\boldsymbol{d} - \boldsymbol{u}_i - \boldsymbol{v}_i\right), \left(\frac{1}{\sigma_\varepsilon^2}(\boldsymbol{M})'(\boldsymbol{M}) + (\boldsymbol{Q})^{-1}\right)^{-1}\right)$$

4. The conditional distribution of $\boldsymbol{v}_i$ given $\boldsymbol{\theta} = (\tau_v, \boldsymbol{d}, \boldsymbol{u}_i, \boldsymbol{c}_i, \sigma_\varepsilon^2)$ is

$$P(\boldsymbol{v}_i|\boldsymbol{\theta}\boldsymbol{z}_i) \propto \mathrm{P}(\boldsymbol{z}_i|\boldsymbol{\theta}, \boldsymbol{v}_i)\,\mathrm{P}(\boldsymbol{v}_i)$$

$$[\boldsymbol{v}_i|\boldsymbol{\theta}] \sim N_m\left(\left(\frac{1}{\sigma_\varepsilon^2}\boldsymbol{I}_m + \tau_v \boldsymbol{R}_v^{-1}\right)^{-1}\left(\frac{1}{\sigma_\varepsilon^2}\boldsymbol{I}'_m(\boldsymbol{z_i} - \boldsymbol{T}\boldsymbol{d} - \boldsymbol{u}_i - \boldsymbol{M}\boldsymbol{c}_i\right), \left(\frac{1}{\sigma_\varepsilon^2}\boldsymbol{I}_m + \tau_v \boldsymbol{R}_v^{-1}\right)^{-1}\right)$$

And the conditional distributions for the variance components are as follows:

5. The full conditional distribution for the population smoother is

$$P(\tau_u|\boldsymbol{u}) \propto \mathrm{P}(\boldsymbol{u}|\tau_u)\,\mathrm{P}(\tau_u)$$

$$[\tau_u|\boldsymbol{u}] \propto |\tau_u \boldsymbol{R}_{\boldsymbol{u}}^*|^{-\frac{1}{2}} exp\left\{-\frac{1}{2}\boldsymbol{u}'(\tau_u \boldsymbol{R}_{\boldsymbol{u}}^*)^{-1}\boldsymbol{u}\right\} \tau_u^{-(\alpha_\tau+1)} exp\left\{-\frac{\beta_\tau}{\tau_u}\right\}$$

$$P(\tau_u|\boldsymbol{u}) \sim IG\left(\alpha_\tau + \frac{nm}{2}, \beta_\tau + \frac{\boldsymbol{u}'\boldsymbol{R}_{\boldsymbol{u}}^{*-1}\boldsymbol{u}}{2}\right)$$

$\boldsymbol{R}_{\boldsymbol{u}}^* = \begin{bmatrix} \boldsymbol{R}_u & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \ddots & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{R}_u \end{bmatrix}$ is an $nm \times nm$ block diagonal matrix of the $\boldsymbol{R}_u$ variance–covariance structure.

6. The full conditional distribution for the subject-specific smoother is written as

$$P(\tau_v|\boldsymbol{v}) \propto \mathrm{P}(\boldsymbol{v}|\tau_v)\,\mathrm{P}(\tau_v)$$

$$[\tau_v|\boldsymbol{v}] \propto |\tau_v \boldsymbol{R}_{\boldsymbol{v}}^*|^{-\frac{1}{2}} exp\left\{-\frac{1}{2}\boldsymbol{v}'(\tau_v \boldsymbol{R}_{\boldsymbol{v}}^*)^{-1}\boldsymbol{v}\right\} \tau_v^{-(\alpha_\tau+1)} exp\left\{-\frac{\beta_\tau}{\tau_v}\right\}$$

$$P(\tau_v|\boldsymbol{v}) \sim IG\left(\alpha_\tau + \frac{nm}{2}, \beta_\tau + \frac{\boldsymbol{v}'\boldsymbol{R}_{\boldsymbol{v}}^{*-1}\boldsymbol{v}}{2}\right)$$

$\boldsymbol{R}_{\boldsymbol{v}}^* = \begin{bmatrix} \boldsymbol{R}_v & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \ddots & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{R}_v \end{bmatrix}$ is an $nm \times nm$ block diagonal matrix of the $\boldsymbol{R}_v$ variance–covariance structure.

7. The full conditional distribution for the variance component of the subject-specific slopes and intercepts is

$$P(\boldsymbol{Q}|\boldsymbol{c}) \propto \mathrm{P}(\boldsymbol{c}|\boldsymbol{Q})\,\mathrm{P}(\boldsymbol{Q})$$

$$[\boldsymbol{Q}|\boldsymbol{c}] \sim IW(n + \alpha, \beta + \boldsymbol{c}'\boldsymbol{c})$$

8. And finally, the full conditional distribution of the model error can be specified as follows:

$$P(\sigma_\varepsilon^2|\boldsymbol{w}, z) \propto \mathrm{P}(\boldsymbol{w}|\boldsymbol{\theta})\,\mathrm{P}(\sigma_\varepsilon^2)$$

$$[\sigma_\varepsilon^2|\boldsymbol{w}, z] \sim IG\left(\frac{nm}{2} + 1 + \alpha, \beta + \frac{1}{2}(z - \boldsymbol{w})'(z - \boldsymbol{w})\right)$$

Acknowledgements

This research was supported in part by research grant 2 P50 DC00242 from the National Institutes on Deafness and Other Communication Disorders, National Institutes of Health; grant RR00059 from the General Clinical Research Centers Program, Division of Research Resources, National Institutes of Health; the Lions Clubs International Foundation; and the Iowa Lions Foundation. The authors would like to thank Dr Bruce Gantz for access to the speech perception data to test our methods and three anonymous referees for valuable feedback that helped improve the original version of this paper.

References

1. Ramsay J, Silverman B. *Functional Data Analysis*. Springer: United States of America, 2010.
2. Guo W. Functional mixed effects models. *Biometrics* 2002; **58**(1):121–8.
3. Zhao L, Chen Y, Schaffner D. Comparison of logistic regression and linear regression in modeling percentage data. *Applied and Environmental Microbiology* 2001; **67**(5):2129–35.
4. Wu H, Zhang J. *Nonparametric regression methods for longitudinal data analysis: mixed-effects modeling approaches*. Wiley-Interscience: Hoboken, NJ, 2006.
5. Wahba G. Improper priors, spline smoothing and the problem of guarding against model errors in regression. *Journal of the Royal Statistical Society. Series B (Methodological)* 1978; **40**(3):364–72.
6. He Y, Yucel R, Raghunathan T. A functional multiple imputation approach to incomplete longitudinal data. *Statistics in Medicine* 2011; **30**(10):1137–56.
7. Goldman M. On the first passage of the integrated Wiener process. *The Annals of Mathematical Statistics* 1971; **42**(6):2150–5.
8. Lefebvre M, Leonard É. On the first hitting place of the integrated Wiener process. *Advances in Applied Probability* 1989; **21**(4):945–8.
9. Gelman A, Carlin J, Stern H, Rubin D. *Bayesian data analysis*, Texts in Statistical Science Series. Chapman and Hall: Boca Raton, FL, 2004.
10. Gantz B, Turner C, Gfeller K. Acoustic plus electric speech processing: preliminary results of a multicenter clinical trial of the Iowa/Nucleus hybrid implant. *Audiology and Neurotology* 2006; **11**(1):63–8.
11. Gantz B, Hansen M, Turner C, Oleson J, Reiss L, Parkinson A. Hybrid 10 clinical trial. *Audiology and Neurotology* 2009; **14**(1):32–8.
12. Rubin D, Little R. *Statistical Analysis with Missing Data*. J Wiley & Sons: Hoboken, NJ, 2002.
13. R Core Team. R: A language and environment for statistical computing, 2012. 2.15.1.
14. Genz A, Bretz F, Miwa T, Mi X, Leisch F, Scheipl F, Hothorn T. Mvtnorm: multivariate normal and t distributions. R package version 0.9-9992, 2012. Available from: http://CRAN.R-project.org/package= mvtnorm [Accessed date March 7, 2013].
15. Martin A, Quinn K, Park J. MCMCpack: Markov chain Monte Carlo in R. *Journal of Statistical Software* 2011; **42**(9):1–2.
16. Leisch F, Weingessel A, Hornik K. Bindata: generation of artificial binary data. R package version 0.9-18, 2011. Available from: http://CRAN.R-project.org/package=bindata [Accessed date March 7, 2013].
17. Bates D, Maechler M. Matrix: Sparse and dense matrix classes and methods. R package version 1.0-6., 2012. Available from: http://CRAN.R-project.org/package=Matrix [Accessed date March 7, 2013].
18. Zeileis A, Grothendieck G. zoo: S3 infrastructure for regular and irregular time series. *Journal of Statistical Software* 2005; **14**(6):1–27.
19. Wuertz D, Chalabi Y. timeSeries: Rmetrics-financial time series objects. R package version 2160.95, 2012. (Available from: http://CRAN.R-project.org/package=timeSeries).
20. Smith B. Boa: an R package for MCMC output convergence assessment and posterior inference journal of statistical software. *Journal of Statistical Software* 2007; **21**(11):1–37.
21. Plummer M, Best N, Cowles K, Vines K. CODA: convergence diagnosis and output analysis for MCMC. *R News* 2006; **6**(1):7–11.
22. Geweke J. Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments. Federal Reserve Bank of Minneapolis, Research Department, 1991.
23. Cowles M, Carlin B. Markov chain Monte Carlo convergence diagnostics: a comparative review. *Journal of the American Statistical Association* 1996; **91**(434):883–904.
24. Statisticat &L. Bayesian inference [Internet]: CRAN. Available from: http://cran.r-project.org/web/packages/Laplaces Demon/index.html [Accessed date March 7, 2013].
25. Spiegelhalter D, Best N, Carlin B, Van Der Linde A. Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 2002; **64**(4):583–639.