Handling Sparse and Missing Data in Functional Data Analysis:

A Functional Mixed-Effects Model Approach

by

Kimberly L. Ward

A Thesis Presented in Partial Fulfillment
of the Requirements for the Degree
Master of Arts

Approved November 2016 by the
Graduate Supervisory Committee:

Hye Won Suk, Chair
Kevin Grimm
Leona Aiken

ARIZONA STATE UNIVERSITY

December 2016

ABSTRACT

This paper investigates a relatively new analysis method for longitudinal data in the framework of functional data analysis. This approach treats longitudinal data as so-called sparse functional data. The first section of the paper introduces functional data and the general ideas of functional data analysis. The second section discusses the analysis of longitudinal data in the context of functional data analysis, while considering the unique characteristics of longitudinal data such, in particular sparseness and missing data. The third section introduces functional mixed-effects models that can handle these unique characteristics of sparseness and missingness. The next section discusses a preliminary simulation study conducted to examine the performance of a functional mixed-effects model under various conditions. An extended simulation study was carried out to evaluate the estimation accuracy of a functional mixed-effects model. Specifically, the accuracy of the estimated trajectories was examined under various conditions including different types of missing data and varying levels of sparseness.

ACKNOWLEDGMENTS

I thank my committee members Drs. Hye Won Suk, Leona Aiken, and Kevin Grimm for their help in developing this document. Their comments brought great insight into helping improve this project and future projects to come.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Figure Page

Figure Page

Figure Page

INTRODUCTION

Psychological research often involves repeated measurements on variables of interest, for example, changes in reading and mathematics abilities throughout early elementary school (Grimm, 2008). When considering behavioral outcomes that change in varying ways over time, the aim is to estimate the trajectories of individuals in order to answer different research questions. For instance, do boys' growth trajectories look different than girls'? Different literatures refer to measures of individual variables taken repeatedly by different names, among them longitudinal or functional data. Traditionally, the distinction between longitudinal data and function data rests upon the number of observed time points; longitudinal data are typically observed at a much sparser set of time points than are functional data. This project aimed to treat longitudinal data as *sparse functional data* (e.g., James, 2010; James, Sugar, & Hastie, 2000; Yao, Muller, & Wang, 2005).

The first section of the paper introduces functional data and the general ideas behind the functional data analysis framework. A discussion of how longitudinal data can be analyzed in the context of functional data analysis is given in the second section. This discussion involves considering the unique characteristics of longitudinal data such as sparseness and missing data. Functional data are typically collected over a number of time points, e.g., tens, hundreds, or even thousands. However, longitudinal data typically have much smaller number of time points (e.g., fewer than 10). Functional data are intensively measured over a relatively short period of time, and in most cases they do not have any missing values. However, longitudinal data are often collected over a long period of time and thus can have substantial amount of missing data. Therefore sparseness and missing data are a unique characteristic of longitudinal data that distinguish them from typical functional data. This implies that one needs to take into account these unique characteristics of longitudinal data when analyzing such data in the functional data analysis framework. There are an increasing number of studies aiming to develop functional data analysis methods for longitudinal data (e.g., Berk, 2012; James, 2010; James et al., 2000; Yao et al., 2005; Wu & Zhang, 2006). Missing data, to my knowledge, has not been discussed in the functional data analysis literature. An introduction to functional mixed-effects models is given in the third section; functional mixed-effects models have the capability to deal with the unique characteristics of longitudinal data (i.e., sparseness and missing data) in the framework of functional data analysis. The fourth section discusses a preliminary simulation study carried out to examine the performance of a functional mixed-effects model under various conditions. An extended simulation study was performed to evaluate functional mixed-effects models on the accuracy of the estimated trajectories under various conditions including different types of missing data and sparseness levels. Lastly, the results from the simulation are presented.

FUNCTIONAL DATA ANALYSIS

Functional data analysis (FDA) is a statistical approach for analyzing intensive streams of data, or so-called functional data (Ullah & Finch, 2013). This technique is gaining increased interest in various fields including biomedical science and psychology. Functional data refer to data arising from infinite-dimensional smooth curves or functions evaluated at a finite number of measurement occasions. Traditional examples of functional data are functional magnetic resonance imaging (fMRI), electroencephalography (EEG), and motion capture data. Functional data are often characterized by hundreds or thousands of measurement occasions. The vast number of time points help display the highly complex trajectories typically seen in functional data; these trajectories often cannot be easily characterized by simple parametric models (e.g., linear, quadratic). FDA was developed to capture and analyze these unique characteristics of functional data. By taking into account the continuity of curves or functions, FDA is capable of handling a large number of measurement occasions as well as unequally spaced and irregular time points varying across individuals. In addition, FDA is based on nonparametric approaches that make possible the characterization of complex nonlinear trajectories. Readers are referred to Ramsay and Silverman (2005) for a comprehensive overview of FDA.

The general concept behind FDA is that raw data reflect an underlying smooth process varying over a continuum such as time and space. Hereafter, it will be assumed that the continuum is *time*. While the underlying process varies smoothly over time, the observed raw data are perturbed by error and do not appear smooth. Therefore, the observed raw data can be modeled as:

$$y_{ij} = x_i\left(t_{ij}\right) + e_{ij}, \tag{1}$$

where y_{ij} denotes the observed response for the ith individual (i = 1, …, n) at the jth time point ($j = 1, \dots, N_i$), t_{ij} is the jth time point for individual i, $x_i(t_{ij})$ is the underlying smooth function for individual i evaluated at time t_{ij}, and e_{ij} is the measurement error. The first step in FDA is to estimate the underlying smooth curve, $x_i(t_{ij})$, from discrete raw data, y_{ij}. The estimated curves are then used for further analyses, setting aside the raw data (Levitin, Nuzzo, Vines, & Ramsay, 2007).

There are various ways of estimating the underlying smooth curves. Conventionally, low-degree polynomials are used to model the data. Figure 1 displays plots of polynomials of different degrees that have been fit to simulated data. The 10 data points (t_{ij}, y_{ij}), for one person (person *i*), were generated from the following model:

$$y_{ij} = sin(2\pi t_{ij}) + e_{ij}, \tag{2}$$

where t_{ij} $(j = 1, \dots, 10)$ were equally spaced time points ranging from 0 to 1 and e_{ij} represents random error. Specifically, e_{ij} were generated from a normal distribution with a mean of zero and a standard deviation of 0.3. Dots in Figure 1 represent the generated data, dashed lines indicate the true underlying curve ($sin(2\pi t)$), and solid lines are the fitted polynomials. A polynomial's degree indicates the number of parameters, which is the highest degree of its terms plus one. For example, a polynomial of degree 3 is a quadratic polynomial, while a polynomial of degree 4 is a cubic polynomial. Figure 1 depicts an example showing how increasing the degree of the polynomial yields better fits to the data. When a polynomial of degree 10 is used -- where the number of data points equals the number of parameters to estimate (i.e., a saturated model)-- the polynomial passes exactly though each data point. These higher degree polynomials tend to oscillate wildly at the boundaries of an interval, thus yielding a very poor representation of the true underlying curve. This is referred to as the Runge phenomenon (Dahlquist & Björk, 1974), and is shown in Figure 1 with the polynomials of degrees 9 and 10.

One way to address this issue of wild oscillation at the boundaries of higher degree polynomials is to use a low degree piecewise polynomial. Let $\tau_1, \tau_2, \ldots, \tau_K$ be *K* different time points that satisfy $a < \tau_1 < \tau_2 < \cdots < \tau_K < b$ where an interval [*a*, *b*] contains all the observed time points t_{ij}. These time points, $\tau_1, \tau_2, \ldots, \tau_K$, are called *knots* and divide the interval of interest [*a*, *b*] into (*K* + 1) subintervals. A piecewise polynomial is obtained by fitting a separate polynomial of the same degree within each subinterval. Piecewise polynomials avoid the Runge phenomenon making them useful for capturing local fluctuations. However, they might be inappropriate to estimate the underlying smooth curve because they can be discontinuous at the knots.

A way to avoid discontinuities at the knots is to use a regression spline. A regression spline can be obtained by constraining a piecewise polynomial to join at each knot (Wu & Zhang, 2006, Chapter 5). In general, an degree-*M* spline is a piecewise polynomial of degree *M* that has continuous derivatives up to degree *M* - 2 (Hastie, Tibshirani, & Friedman, 2009, Chapter 5). For example, a spline of degree 4 with two interior knots $\{\tau_1, \tau_2\}$ can be represented as follows:

$$x_i(t) = \sum_{m=1}^{6} \phi_m(t) c_{im} \tag{3}$$

where $x_i(t)$ is the underlying curve for person *i* that is approximated by a spline of degree *M*, $\phi_m(t)$ is the *m*th truncated power basis function as defined in equation (4) evaluated at time *t*, and c_{im} is the coefficient for the *m*th basis function.

$$\phi_1(t) = 1;\ \ \phi_2(t) = t;\ \phi_3(t) = t^2;\ \phi_4(t) = t^3; \tag{4}$$

$$\phi_5(t) = [\max(0, t - \tau_1)]^3\ ; \phi_6(t) = [\max(0, t - \tau_2)]^3$$

From (3) and (4) we can see that in the first subinterval, where $a < t < \tau_1$, the underlying curve is represented by the following cubic polynomial:

$$x_i(t) = c_{i1} + c_{i2}t + c_{i3}t^2 + c_{i4}t^3 + c_{i5}(t - \tau_1)^3. \tag{5}$$

In the third subinterval, where $\tau_2 \leq t < b$, the underlying curve is represented by the following cubic polynomial:

$$x_i(t) = c_{i1} + c_{i2}t + c_{i3}t^2 + c_{i4}t^3 + c_{i5}(t - \tau_1)^3 + c_{i6}(t - \tau_2)^3. \tag{6}$$

It can be easily verified that the three models (4), (5), and (6) are continuous at the knots, where $t = \tau_1$ and $t = \tau_2$, and their first and second derivatives are also continuous at the knots.

In general, a spline of degree *M* with *K* interior knots can be represented as a linear combination of *M* + *K* truncated power basis functions. In addition, a spline can be represented by using other basis functions such as B-spline basis functions (de Boor, 2001). Basis functions can be thought of as the functional extension of basis vectors. Borrowing concepts from linear algebra, any vector in a vector space can be expressed as a linear combination of a set of basis vectors that generate the vector space. A set of basis vectors of a vector space are a set of linearly independent vectors that can represent every vector in that space via linear combinations (Leon, 2010). There are many different sets of basis vectors that can generate the same vector space. In other words, a vector can be represented as a linear combination of each different set of basis vectors. Similarly, there are many different sets of basis functions that can generate a function space. A spline function in the function space can be represented as a linear combination of each different set of basis functions.

The coefficients for the basis functions can be estimated by minimizing the sum of squared residuals given in (7). The estimated spline function is called a regression spline.

$$SSR = \sum_{j=1}^{n_i}(y_{ij} - x(t_{ij}))^2 = \sum_{j=1}^{n_i}(y_{ij} - \sum_{m=1}^{M+K+1}\phi_m(t)c_{im})^2 \quad (7)$$

Figure 2 displays regression splines of degree 4 fitted to the simulated data given in Figure 1. The left panel in Figure 2 shows a regression spline of degree 4 fit to the data with three equally spaced knots. The right panel shows a regression spline of degree 4 fit to the data with five equally spaced knots. As we can see, the performance of regression splines strongly depends on the choice of number and locations of the knots. In other words, we need to select the number and locations of the knots very carefully to obtain a good representation of the underlying curve. This is not a trivial problem. Refer to Wu & Zhang (2006, Chapter 3.3.3) for popular methods used for knot selection.

To bypass the problem of selecting the number and location of knots, one can estimate a spline function using the penalized least squares approach. This approach minimizes the following penalized sum of squared residuals:

$$PSSR(\lambda) = \sum(y_{ij} - x_i(t_{ij}))^2 + \lambda\int_a^b \left(x''_i(t)\right)^2 dt \quad (8)$$

where $x''_i(t)$ indicates the second derivative of $x_i(t)$ and $[a, b]$ defines the range over which the function $x_i(t)$ is defined. The estimated spline function that minimizes this penalized least squares criterion is called a smoothing spline.

The penalized least squares approach aims to minimize the least squares criterion (i.e., sum of squared residuals) with the addition of a roughness penalty term. The first term in (8) is the sum of squared residuals as given in (7). The second term is a penalty term indicating the integrated squared second derivative of $x_i(t)$, which is multiplied by a non-negative smoothing parameter (λ). The squared second derivative of a function at time t indicates the curvature or roughness of that function at time t. If $x_i(t)$ is a straight line having no curvature, the second derivative ($x''_i(t)$) will be zero over the entire range of t. If $x_i(t)$ has a curvature at time t, the second derivative of this function at time *t* will deviate from 0. Therefore, the integrated squared second derivative of this function indicates the overall roughness of $x_i(t)$. The non-negative smoothing parameter λ controls the importance of the penalty term. When $\lambda = 0$, the penalty term will vanish and the spline function is estimated in such a way that it fits the data as closely as possible. When $\lambda = \infty$, the penalty term will dominate the criterion (8) and even a tiny amount of curvature in $x_i(t)$ will yield a huge value of the criterion. Hence, minimizing the criterion (8) will yield the linear least squares line. In other words, a larger smoothing parameter λ will yield a smoother smoothing spline. Refer to Wu & Zhang (2006, Chapter 3.7) for a comprehensive overview of the methods used for smoothing parameter selection.

Interestingly, it has been proven that there is a unique minimizer of the penalized least squares criterion (8), which is a natural cubic spline with knots at each time point (Green & Silverman, 1994, Chapter 2). A natural cubic spline is a cubic (fourth degree) spline with the constraint that the function be linear beyond the boundary knots. Smoothing splines avoid knot selection problems completely by placing a knot at each unique time point, while controlling the estimated function's smoothness via the smoothing parameter. Figure 3 shows the estimated cubic smoothing spline for the simulated data given in Figure 1.

An important assumption of FDA is that responses are measured at a dense grid of measurement occasions. When this assumption is met, each curve can be estimated by fitting a smoothing spline to the raw data from the particular individual. However, the assumption of dense sampling might not be met, especially when researchers collect longitudinal data. This could place a potential limitation on the applicability of FDA to non-densely measured series.

LONGITUDINAL DATA AS SPARSE FUNCTIONAL DATA

Longitudinal data can be regarded as functional data that arise from a smooth underlying process, though, they are distinguished from traditional functional data in several aspects. Longitudinal data are measured at a relatively sparse set of time points. Consequently, they are referred to as “sparse” functional data as compared to “dense” standard functional data (e.g., James, 2010; James, Sugar, & Hastie, 2000; Yao, Muller, & Wang, 2005). Several examples are provided that show the application of FDA methods to sparse data. James (2010) analyzed spinal bone mineral density data measured from 280 individuals taken repeatedly at various ages. Even though there were a total of 860 observations over the entire range of time, each individual only contributed two to four observations. Yao et al. (2005) analyzed 283 individuals’ CD4 percentages, a commonly used marker for the health status of HIV infected persons. Many individuals missed scheduled visits and the number of observations per person varied widely from 1 to 14.

Longitudinal studies are also subject to missing data. A participant’s reason for missing a measurement may range from completely random to systematic. The reasons for missingness are classified by three different missingness mechanisms: missing completely at random (MCAR), missing at random (MAR), and missing not at random (MNAR) (Rubin, 1976). When data are MCAR, the probability of missing data on a variable is unrelated to other observed variables and unrelated to the unobserved value of the variable itself. Data are MAR when the probability of missing data on a variable is related to some other observed variable(s) in the analysis model, but not related to the unobserved value of the variable itself. Lastly, MNAR means the probability of missing data on a variable is related to the unobserved value of that variable, even after controlling for other variables (Enders, 2010). An example of systematic missingness (e.g., MAR or MNAR) is when a patient in a study investigating life expectancy after cancer treatment missed an appointment because he or she died; this is a case of MNAR because the reason for missingness on the outcome variable (life expectancy) is related to unobserved value of the variable (life expectancy) itself. An example of MAR is if a patient had a fever and they were not allowed to do a physical fitness test for the study; the reason for missingness on the fitness test is related to another variable in the analysis model, in this case, presence of fever. An example of non-systematic (i.e., ignorable) missingness is a patient missing an appointment because he or she had car trouble; the reason for missing data is assumed to be unrelated to variables under investigation.

In classic ANOVA and multiple regression analysis, systematically missing data can severely bias results, and special care is required in the analysis of data with systematic missingness (e.g., Enders, 2010; Yang & Maxwell, 2014). If traditional approaches to handling missing data are used (e.g., listwise deletion, pairwise deletion, mean imputation, etc.), ANOVA or multiple regression can yield biased parameter estimates. If data are not MCAR, both listwise and pairwise deletion can lead to biased parameter estimates, and mean imputation can attenuate correlations and covariances (Ender, 2010). More sophisticated analysis methods that use maximum likelihood estimation, such as mixed-effects models, can yield unbiased parameter estimates if data are MCAR or MAR provided all covariates of missingness are accounted for in the model. However, if the covariate(s) that account for missingness are omitted from the model or are unobserved (i.e., MNAR), the analysis can lead to biased inferences (Laird, 1988). Most of the research on missing data has focused on the MCAR and MAR mechanisms. A much smaller amount of research exists for the MNAR mechanism (e.g., Yang & Maxwell, 2014).

In the above examples on measures of spinal bone mineral density and CD4 percentages, fitting a smoothing spline to each individual's data may fail because the number of observations per curve is too small to fully reflect the underlying curve. This is attributable to both sparseness and missingness. To estimate each individual's curve, it would be necessary to borrow information from all curves. James (2010) and Yao et al. (2005) suggest a mixed-effects framework is suitable for achieving this goal. These papers have investigated methods for handling sparse functional data; however, there is no body of literature available on the effects of missing data in FDA. In the following section, functional mixed-effects models to handle sparse functional data are discussed.

FUNCTIONAL MIXED-EFFECTS MODELS

Mixed-effects models are statistical models that contain both fixed and random effects (Searle, Casella, & McCulloch, 1992). Other terms for mixed-effects models include hierarchical linear models (Raudenbush & Bryk, 2002) and random coefficient models (Swamy, 1971). In social sciences and biomedical sciences, they are often referred to as multilevel models (Snjiders & Bosker, 1999).

Mixed-effects models aim to analyze data with a nested structure. Thus, they are very useful for analyzing repeated measures, or longitudinal data, where multiple repeated measurements are nested within individuals. Linear mixed-effects models (Laird & Ware, 1982) are among the most widely used methods for analyzing longitudinal data (Snijders & Bosker, 1999). The model given in equation (9) shows a linear growth curve model,

$$Y_{ij} = \mu_0 + \mu_1 t_{ij} + v_{0i} + v_{1i} t_{ij} + e_{ij} \tag{9}$$

where Y_{ij} is the response or outcome variable for the *i*th individual ($i = 1, \dots, n$) at t_{ij}, which is the *j*th time point ($j = 1, \dots, N_i$) for the *i*th individual, μ_0 is the mean intercept, μ_1 is the mean slope for time, v_{0i} indicates how much the intercept for person *i* deviates from the mean intercept, and v_{1i} indicates how much the slope for person *i* deviates from the mean slope. In other words, $\mu_0 + v_{0i}$ is the person-specific intercept and $\mu_1 + v_{1i}$ is the person-specific slope. The mean intercept (μ_0) and mean slope (μ_1) are fixed parameters; the intercept deviation (v_{0i}) and slope deviation (v_{1i}) are person-level random effects. The residual, e_{ij}, indicates how much the observation at the *j*th time point for person *i* deviates from the linear trajectory of the person and is a measurement-level random effect.

The observations for each person can be arranged in a vector and the model given in equation (9) can be re-written as:

$$\mathbf{y}_i = \mathbf{X}_i \boldsymbol{\mu} + \mathbf{Z}_i \mathbf{v}_i + \boldsymbol{e}_i \tag{10}$$

where $\mathbf{y}_i = [y_{i1}, \dots, y_{iN_i}]^T$ is a vector of observed responses for person *i* at N_i time points, $\boldsymbol{\mu}$ is a vector of fixed effects, $\mathbf{v}_{\boldsymbol{i}}$ is a vector of person-level random effects, $\boldsymbol{e}_i$ is a vector of residuals or measurement-level random effects for person *i*, and $\mathbf{X}_i$ and $\mathbf{Z}_{\boldsymbol{i}}$ are design matrices for the fixed and random effects, respectively. For example, the linear growth model given in equation (9) can be written as:

$$\begin{bmatrix} Y_{i1} \\ Y_{i2} \\ Y_{i3} \end{bmatrix} = \begin{bmatrix} 1 & t_{i1} \\ 1 & t_{i2} \\ 1 & t_{i3} \end{bmatrix} \begin{bmatrix} \mu_0 \\ \mu_1 \end{bmatrix} + \begin{bmatrix} 1 & t_{i1} \\ 1 & t_{i2} \\ 1 & t_{i3} \end{bmatrix} \begin{bmatrix} v_{0i} \\ v_{1i} \end{bmatrix} + \begin{bmatrix} e_{i1} \\ e_{i2} \\ e_{i3} \end{bmatrix} \tag{11}$$

where $\mathbf{y}_i = [y_{i1}, y_{i2}, y_{i3}]^T$, $\mathbf{X}_i = \begin{bmatrix} 1, & 1, & 1 \\ t_{i1}, & t_{i2}, & t_{i3} \end{bmatrix}^T$, $\boldsymbol{\mu} = [\mu_0, \mu_1]^{\boldsymbol{T}}$, $\mathbf{Z}_i = \begin{bmatrix} 1, & 1, & 1 \\ t_{i1}, & t_{i2}, & t_{i3} \end{bmatrix}^T$, $\mathbf{v}_{\boldsymbol{i}} = [v_{0i}, v_{1i}]^{\boldsymbol{T}}$, and $\mathbf{e}_i = [e_{i1}, e_{i2}, e_{i3}]^T$. In general, the random effects are assumed to follow multivariate normal distributions, that is, $\mathbf{v}_{\boldsymbol{i}} \sim N(\mathbf{0}, \mathbf{D})$ and $\mathbf{e}_{\boldsymbol{i}} \sim N(\mathbf{0}, \mathbf{R}_{\boldsymbol{i}})$. The person-level random effects $\mathbf{v}_{\boldsymbol{i}}$ and measurement-level random effects $\mathbf{e}_{\boldsymbol{i}}$ are assumed to be uncorrelated. In addition, the covariance matrix $\mathbf{R}_{\boldsymbol{i}}$ is often assumed to be diagonal. These covariance matrices are left unstructured for the purpose of this project.

The parameters $\boldsymbol{\mu}$, $\mathbf{D}$, and $\mathbf{R}_{\boldsymbol{i}}$, can be estimated by minimizing the twice negative generalized log likelihood (Wu & Zhang, 2006, Chapter 2), which is given by:

$$\text{GLL} = \sum_{i=1}^{n} \{[\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\mu} - \mathbf{Z}_i\mathbf{v}_i]^T \mathbf{R}_i^{-1} [\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\mu} - \mathbf{Z}_i\mathbf{v}_i] + \mathbf{v}_i^T \mathbf{D}^{-1} \mathbf{v}_i + \log|\mathbf{D}| + \log|\mathbf{R}_{\boldsymbol{i}}|\}. \tag{12}$$

The best unbiased predictor of the random effects $\mathbf{v}_i$ is obtained by:

$$\hat{\mathbf{v}}_i = \widehat{\mathbf{D}}\mathbf{Z}_i^T \left(\mathbf{Z}_i \widehat{\mathbf{D}} \mathbf{Z}_i^T + \widehat{\mathbf{R}}_i\right)^{-1} (\mathbf{y}_i - \mathbf{X}_i \hat{\boldsymbol{\mu}}) \tag{13}$$

where $\hat{\boldsymbol{\mu}}, \widehat{\mathbf{D}}$, and $\widehat{\mathbf{R}}_i$ are the estimates of $\boldsymbol{\mu}$, $\mathbf{D}$, and $\mathbf{R}_{\boldsymbol{i}}$, respectively. Refer to Wolfinger, Tobias, & Sall (1994) and McCulloch, Searle, & Neuhaus (2008, Chapter 6) for more detailed estimation algorithms.

Mixed-effects models do not require each person to be observed at the same number of time points. In other words, the length of the vector $\mathbf{y}_i$ in (10) can vary from person to person. Therefore mixed-effects models can intrinsically deal with missing data. In addition, the parameters of mixed-effects models are estimated simultaneously by minimizing a single objective function as given in (12). This indicates that each person's random effects are obtained by taking into account that individual's data and data from all individuals (i.e., shrinkage estimate). Therefore mixed-effects models are suitable for analyzing sparse data that have only a few time points for each person but a lot of time points across all individuals.

Functional mixed-effects models (FMEMs) are the functional extension of mixed-effects models. Functional mixed-effects models take on the general form:

$$Y_{ij} = \mu(t_{ij}) + v_i(t_{ij}) + e_{ij} \tag{14}$$

where $\mu(t_{ij})$ is the grand-mean function evaluated at time t_{ij}, $v_i(t_{ij})$ is a person-specific effect function evaluated at time t_{ij}. This indicates how much the person-specific trajectory of person *i* deviates from the grand mean trajectory at time t_{ij}; e_{ij} is the residual that indicates how much the observed response at time t_{ij} deviates from the person's trajectory.

Let us assume that there are a total of *K* distinct time points in the data, $\{\tau_1, \tau_2, \dots, \tau_K\}$. For example, if some individuals are measured at their age of 8, 9, 10, and 11 and other individuals are measured at their age of 10, 11, and 12, there are a total of *K* = 5 distinct time points in the data, $\{\tau_1 = 8, \tau_2 = 9, \tau_3 = 10, \tau_4 = 11, \tau_5 = 12\}$. Using vector-matrix notations, the model can be re-written as:

$$\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\mu} + \mathbf{Z}_i\mathbf{v}_i + \boldsymbol{e}_i \tag{15}$$

where $\mathbf{y}_i$ and $\boldsymbol{e}_i$ are defined as in equation (10), $\boldsymbol{\mu}$ is the vector of the values of the grand-mean function $\mu(t)$ evaluated at all distinct time points in the data set, and $\mathbf{v}_i$ is the vector of the values of the person-specific effect function $v_i(t)$ evaluated at all distinct time points. In this model, $\mathbf{Z}_i$ is identical to $\mathbf{X}_i$. $\mathbf{X}_i$ is an N_i by K incidence matrix whose *j*th row of this matrix has a 1 by *K* indicator vector indicating which value of all distinct time points equals t_{ij}. That is, the *k*th element of the indicator vector is 1 if the *j*th time point, t_{ij}, equals the *k*th distinct time point, i.e., $t_{ij} = \tau_k$, and all the other elements are 0. In this model, $\boldsymbol{\mu}$ is fixed and $\mathbf{v}_i$ and $\boldsymbol{e}_i$ are random. In general, the random effects are assumed to follow multivariate normal distributions, that is, $\mathbf{v}_{\boldsymbol{i}} \sim N(\mathbf{0}, \mathbf{D})$ and $\mathbf{e}_{\boldsymbol{i}} \sim N(\mathbf{0}, \mathbf{R}_{\boldsymbol{i}})$.

For example, suppose that there are a total of 5 distinct time points in the data, 1, 2, 3, 4, and 5, and person *i* is measured at the time points 1, 2, and 4. If so, then the model (16) for this person can be written as:

$$\begin{bmatrix} Y_{i1} \\ Y_{i2} \\ Y_{i3} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} \mu(\tau_1) \\ \mu(\tau_2) \\ \mu(\tau_3) \\ \mu(\tau_4) \\ \mu(\tau_5) \end{bmatrix} + \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} v_i(\tau_1) \\ v_i(\tau_2) \\ v_i(\tau_3) \\ v_i(\tau_4) \\ v_i(\tau_5) \end{bmatrix} + \begin{bmatrix} e_{i1} \\ e_{i2} \\ e_{i3} \end{bmatrix} \qquad (16)$$

where $\mathbf{y}_i = [y_{i1}, y_{i2}, y_{i3}]^T$, $\mathbf{X}_i = \mathbf{Z}_i = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix}$,

$\boldsymbol{\mu} = [\mu(\tau_1), \mu(\tau_2), \mu(\tau_3), \mu(\tau_4), \mu(\tau_5)]^{\boldsymbol{T}}$, $\mathbf{v}_{\boldsymbol{i}} = [v_i(\tau_1), v_i(\tau_2), v_i(\tau_3), v_i(\tau_4), v_i(\tau_5)]^{\boldsymbol{T}}$,

and $\mathbf{e}_i = [e_{i1}, e_{i2}, e_{i3}]^T$.

The parameters of the model (15) can be estimated by minimizing the following twice negative penalized generalized log likelihood criterion (PGLL) (Wu & Zhang, 2006, p.165):

$$\text{PGLL} = \sum_{i=1}^{n} \{[\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\mu} - \mathbf{X}_i\mathbf{v}_i]^T \mathbf{R}_i^{-1}[\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\mu} - \mathbf{X}_i\mathbf{v}_i] + \mathbf{v}_i^T\mathbf{D}^{-1}\mathbf{v}_i + \log|\mathbf{D}| + \log|\mathbf{R}_{\boldsymbol{i}}|\} + \lambda_\mu \int [\mu''(t)]^2 dt + \lambda_v \sum_{i=1}^{n} \left\{\int [v_i''(t)]^2 dt\right\}. \tag{17}$$

In the above expression, the first term is the twice negative generalized log likelihood as given in equation (12); the second term is the roughness of the fixed-effect function $\mu(t)$ multiplied by its smoothing parameter λ_μ; the third term is the sum of the roughness of the random-effect functions $v_i(t)$ multiplied by a common smoothing parameter λ_v. The smoothing parameter, λ_μ, controls the tradeoff between the goodness of fit and the roughness of $\mu(t)$; the smoothing parameter, λ_v, controls the tradeoff between the goodness of fit and the aggregated roughness of $v_i(t)$.

Using the roughness matrix $\mathbf{G}$ as defined in Wu & Zhang (2006, p.55), the PGLL criterion can be re-written as (Wu & Zhang, 2006, p.166):

$$\text{PGLL} = \sum_{i=1}^{n} \{[\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\mu} - \mathbf{X}_i\mathbf{v}_i]^T \mathbf{R}_i^{-1}[\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\mu} - \mathbf{X}_i\mathbf{v}_i] + \mathbf{v}_i^T\mathbf{D}^{-1}\mathbf{v}_i + \log|\mathbf{D}| + \log|\mathbf{R}_{\boldsymbol{i}}|\} + \lambda_\mu \boldsymbol{\mu}^T\mathbf{G}\boldsymbol{\mu} + \lambda_v \sum_{i=1}^{n} \{\mathbf{v}_i^T\mathbf{G}\mathbf{v}_{\boldsymbol{i}}\}. \tag{18}$$

The minimizers of the PGLL criterion, $\hat{\boldsymbol{\mu}}, \hat{\mathbf{v}}_i, \widehat{\mathbf{D}}$, and $\widehat{\mathbf{R}}_i$, are called cubic mixed-effects smoothing spline estimators (Wu & Zhang, 2006, Chapter 6.5.1). They can be obtained by using an EM algorithm (Wu & Zhang, 2006, p.170), which is implemented in the *sme* package (Berk, 2012) in R. After obtaining $\hat{\boldsymbol{\mu}}, \hat{\mathbf{v}}_i, \widehat{\mathbf{D}}$, and $\widehat{\mathbf{R}}_i$, the estimated fixed effect curve, $\hat{\mu}(t)$, and the predicted random effect curves, $\hat{v}_i(t)$, can be evaluated at any value of t using a simple formula given in Green & Silverman (1994, Chapter 2.4) or a simple interpolation.

PRELIMINARY SIMULATION STUDY

Various versions of FMEMs have been proposed by extending mixed-effects models to handle sparse functional data (e.g., James, 2010; James et al., 2000; Wu & Zhang, 2006; Yao et al., 2005). To date, however, there are no studies systematically examining how FMEMs perform under different degrees of sparseness and types of missingness. Therefore, I carried out a preliminary simulation study to investigate the effects of degree of sparseness and different types of missingness on the accuracy of the mean and individual trajectories estimated by FMEMs (Ward & Suk, 2015). More specifically, I examined the cubic smoothing spline mixed-effects model described in the previous section.

Four factors were examined in this study. First, sparseness level was manipulated. Sparseness was defined as the number of time points per curve (N_i). As N_i increases data become less sparse. Sparseness level varied at three levels: $N_i = 5, 10, 20$. These values were chosen to reflect typical number of time points found in various longitudinal studies (e.g. Hankin, Abramson, Moffitt, Silva, Mcgee, & Angell, 1998; Porter, Crampon, & Smith, 1976; Cinciripini, Lapitsky, Seay, Wallfisch, Kitchens, & Van Vunakis, 1995). Sample size was varied at three levels: *n* = 40,100,400. Again, these sample size values were chosen to reflect those typical in longitudinal studies (e.g., Adler, 2012; Cinciripini et al., 1995; Esser, Schmidt, & Woerner, 1990). Error variance level was manipulated at three levels: $\sigma_e^2 = .25, .5, 1$. These values were chosen to mimic previous simulation studies on functional data (e.g., Yao et al., 2005; Di, Crainiceanu, Caffo, & Punjabi, 2009). Lastly, three types of missingness were examined: no missing data, randomly missing, and time-dependent missing. In the no missing data condition, each curve was observed at every design time point. Both the random and time-dependent missing conditions had a total of 30% missing data, but in different ways. Typical rates of missing data in psychological studies are approximately 15-20% (Peugh & Enders, 2004). The value of 30% was chosen to be more extreme than this typical rate. In the randomly missing condition, data were missing completely randomly. That is, the probability of an observation being missing was 0.30 for all individuals across all time points. This condition was designed to mimic the missingness mechanism MCAR. In the time-dependent missing data condition, more data were missing at later time points; the probability of an observation being missing depended on time. This condition was designed to mimic the missingness mechanism MAR. More specifically, to generate time-dependent missing data, all observations (for all individuals and for all time points) were split into four quarters according to the time points at which they were measured. This procedure is done on the entire set of data points; this means, for example, 5 time points multiplied by 100 curves gives 500 total time points. These 500 time points are broken up into quarters and a given percent of data is removed depending on the quarter. No observations were removed in the first quarter. Observations belonging to the second quarter were randomly removed with the probability of 20%. For those belonging to the third quarter, 40% were randomly removed. Lastly, those in the fourth quarter, 60% were randomly removed. Therefore, a total of 30% of the observations were removed.

In sum, this simulation had a total of 81 conditions (3 sparseness levels x 3 sample sizes x 3 error variances x 3 types of missingness); 100 data sets were generated per condition. Each observation for the lth data set (l =1, … , 100) under each condition was generated as in Yao et al. (2005) from the following model:

$$Y_{ijl} = x_{il}\left(t_{ijl}\right) + \ e_{ijl} \tag{19}$$

where Y_{ijl} is the observed response for the *i*th individual (*i* = 1, …,*n*) in the *l*th data set at the *j*th time point ($j = 1, \dots, N_i$), $x_{il}\left(t_{ijl}\right)$ is the true underlying person-specific curve for individual *i* in the *l*th data set evaluated at the *j*th time point for this person, t_{ijl}; e_{ijl} is the measurement error. This model is identical to the model given in (1). The person-specific curve, $x_{il}(t)$, is modeled by:

$$x_{il}(t) = \mu(t) + v_{il}(t), \tag{20}$$

where $\mu(t)$ is the grand mean function and $v_{il}(t)$ is the person-specific effect function for the *i*th individual in the *l*th data set. Inserting (20) into (19) yields the following model:

$$Y_{ijl} = \mu\left(t_{ijl}\right) + v_{il}\left(t_{ijl}\right) + \ e_{ijl}, \tag{21}$$

which is identical to the model given in (14).

The grand mean function, $\mu(t)$, was defined as:

$$\mu(t) = t + \sin(t), \tag{22}$$

which is the green curve shown in Figures 6, 7 and 8 below. The person-specific effect functions, $v_{il}(t)$, were generated from the following model:

$$v_{il}(t) = \textstyle\sum_{m=1}^{2} \phi_m(t) c_{iml}, \tag{23}$$

where the two basis functions, $\phi_1(t)$ and $\phi_2(t)$, were defined as:

$$\begin{aligned}\phi_1(t) &= -\cos(\tfrac{\pi t}{10})/\sqrt{5} \\ \phi_2(t) &= sin(\tfrac{\pi t}{10})/\sqrt{5}\end{aligned}, \tag{24}$$

and the two person-specific basis function coefficients, c_{i1l} and c_{i2l}, were randomly generated from the following bivariate normal distribution:

$$\begin{bmatrix} c_{i1l} \\ c_{i2l} \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 4 & 0 \\ 0 & 1 \end{bmatrix}\right). \tag{25}$$

The time points, t_{ijl}, were randomly generated from a uniform distribution over the range [0,10]. The measurement errors, e_{ijl}, were randomly generated from a normal distribution of a mean of 0 and a variance of σ_e^2. After the responses, Y_{ijl}, were generated as described above, 30% of them were removed in the random missing and time-dependent missing conditions.

Each data set under each condition was analyzed using a FMEM. The grand mean and person-specific effect functions were estimated by minimizing the PGLL criterion given in (17) using the *sme* package in R (Berk, 2012). Analysis results were evaluated in the following way. First, to evaluate the accuracy of the estimated grand mean function, $\mu(t)$, a mean square error (MSE) was calculated under each conditions. The MSE is defined as:

$$\mathrm{MSE}\big(\hat{\mu}(t)\big) = \frac{\sum_{l=1}^{100}\sum_{s=1}^{100}(\hat{\mu}_l(t_s) - \mu(t_s))^2}{100}, \tag{26}$$

where $\hat{\mu}_l(t_s)$ is the estimated grand mean function for the *l*th data set evaluated at time t_s, t_s is the *s*th time point out of the 100 equally spaced time points over the range of [0,10], and $\mu(t)$ is the true grand mean function given in (22). In the numerator, the term $\sum_{s=1}^{100}(\hat{\mu}_l(t_s) - \mu(t_s))^2$, measures how much the estimated grand mean function for the *l*th data set deviates from the true grand mean function across 100 equally spaced time points. Therefore the mean square error, $\text{MSE}\big(\hat{\mu}(t)\big)$, measures how much the estimated grand mean function deviates from the true grand mean function across 100 equally spaced time points, on average across 100 data sets. Second, the accuracy of the predicted person-specific trajectories was assessed by calculating the average mean square error (AMSE) defined as:

$$\text{AMSE}\big(\hat{x}(t)\big) = \frac{\sum_{l=1}^{100}\sum_{i=1}^{n}\sum_{s=1}^{100}\big(\hat{x}_{il}(t_s) - x_{il}(t_s)\big)^2}{(100)(n)} \tag{27}$$

where $\hat{x}_{il}(t_s)$ indicates the predicted curve for the *i*th individual in the *l*th data set evaluated at the *s*th time point, and $x_{il}(t_s)$ is the true curve for the *i*th individual in the *l*th data set evaluated at the *s*th time point. Similar to $\text{MSE}\big(\hat{\mu}(t)\big)$, a lower value of $\text{AMSE}\big(\hat{x}(t)\big)$ indicates that the estimated person-specific trajectories are closer to the true curves, indicating better recovery.

It was expected that the estimated mean and person-specific trajectories would become more accurate as data became less sparse, sample size increased, and error variance decreased. The time-dependent missing data condition was expected to produce the least accurate results, followed by the randomly missing condition. No missing data was expected to produce the most accurate results.

The results of the simulation study are presented in Figures 4 and 5. Figures 6, 7, and 8 illustrate the estimated mean and individual trajectories. The results revealed that model performance increased with increasing sample size and decreasing sparseness, which was as expected. However, two surprising results were observed. First, the model worked well even with extremely sparse data (i.e., only five measurement points), without a noteworthy gain in accuracy with increasing measurement points. This result contradicts the expectation that more measurement occasions would yield considerably more accurate estimates of the underlying trajectories. Second, the model's ability to recover the underlying curves was virtually unaffected by different types of missingness. Should these results generalize, the use of a FMEM may permit researchers to avoid biased outcomes in longitudinal analyses with limited number of observations and time-dependent missingness.

PROPOSED SIMULATION STUDY

The first simulation study was limited to data generation that used a single monotonically increasing mean trajectory, given in (22), and a single percentage of missing data (30%). The proposed simulation aimed to extend the first simulation study to more diverse conditions. That is, the proposed simulation was carried out to ensure the results found in the first simulation were not an artifact of the data generation process (i.e., using a specific mean trajectory), and to examine the impact of more extreme percentages of missingness on accuracy of estimation for the mean and individual trajectories.

The factors manipulated in the proposed study were the shape of the underlying mean trajectory, percentage of missing data, type of missingness, degree of sparseness, and irregularity of time points. More specifically, three different shapes of trajectories were examined: quadratic, asymptotic, and periodic. Figures 9, 10, and 11 display the three different shapes of functions. These were chosen to mimic various trajectories found in longitudinal studies (e.g., Cinciripini et al., 1995; Ramsay & Silverman, 2005). To test the ability of FMEMs to handle more extreme amounts of missing data, percent missingness was examined across four levels: 30%, 50%, 70%, and 90%. Similar to the previous simulation, three different types of missingness were examined: no missing data, randomly missing throughout the range of time (MCAR), and time-dependent missing (MAR). In the time-dependent missing condition, data were again generated such that later time points contained greater missingness; data points were deleted with higher probability for later time points. Table 1 displays the percentage of randomly picked and removed data at each quintile under each percentage of missingness used to generate time-dependent missing data. For instance, for 30% time-dependent missing data, the third quintile had 30% of the data point in that quintile randomly selected and removed. Three sparseness levels were examined with the number of time points: $N_i = 5$, 11, and 21. Lastly, the irregularity of time points was manipulated with two different conditions: fixed and random. The fixed condition simulates longitudinal studies where measurements occur at predefined intervals and thus time points are common to all individuals. For example, each subject is collected at exactly times [0, 5, 10, 15, 20]. The random condition simulates measurements obtained at random intervals (e.g., Cinciripini et al., 1995) where time points vary across individuals. For example, subject 1 is collected at times [.03, 3, 11.99, 17, 19.6], while subject 2 is collected at times [3, 4.02, 6.7, 6.9, 7.1]. Sample size was fixed at n =100, and error variance was fixed at $\sigma_e^2 = 30$. The value of 30 was chosen in order to maintain an approximate signal-to-noise ratio around 10; the goal was to have approximately 10 times the amount of signal variance in the data compared to noise variance. Sample size was chosen based on results in the previous simulation; the patterns of results were similar across all levels of this factor. Therefore, I selected the middle value for sample size.

In sum, there are a total of 162 conditions. More specifically, under no missing data, there are 18 conditions to examine: 3 trajectories x 3 sparseness levels x 2 irregularity levels. Under missing data, there are 144 conditions to examine: 3 trajectories x 4 percent missingness levels x 2 types of missingness (random vs time-dependent) x 3 sparseness levels x 2 irregularity levels. Each observation for the *l*th data set (*l* =1, … , 100) under each condition was generated from the following model:

$$Y_{ijl} = x_{il}\left(t_{ijl}\right) + \ e_{ijl} \tag{28}$$

where Y_{ijl} is the observed response for the *i*th individual (*i* = 1, …,*n*) in the *l*th data set at the *j*th time point ($j = 1, \dots, N_i$), $x_{il}\left(t_{ijl}\right)$ is the true underlying person-specific curve for individual *i* in the *l*th data set evaluated at the *j*th time point for this person, t_{ijl}; e_{ijl} is the measurement error. This model is identical to the model given in (1). The person-specific curve, $x_{il}(t)$, is modeled by:

$$x_{il}(t) = \mu(t) + v_{il}(t), \tag{29}$$

where $\mu(t)$ is the grand mean function and $v_{il}(t)$ is the person-specific effect function for the *i*th individual in the *l*th data set. Inserting (28) into (29) yields the following model:

$$Y_{ijl} = \mu\left(t_{ijl}\right) + v_{il}\left(t_{ijl}\right) + \ e_{ijl}, \tag{30}$$

which is identical to the model given in (14).

The three grand mean functions, $\mu(t)$, were defined as:

$$\mu_1(t) = \frac{1}{2}(t - 10)^2 + 10, \tag{31}$$

$$\mu_2(t) = 50(1 - e^{-.3t}) + 10, \tag{32}$$

$$\mu_3(t) = 20\sin(.7t) + 40. \tag{33}$$

The person-specific effect functions, $v_{il}(t)$, were generated from the following model:

$$v_{il}(t) = \sum_{m=1}^{3} \phi_m(t) c_{iml}, \tag{34}$$

where $\phi_1(t)$, $\phi_2(t)$, and $\phi_3(t)$ were defined as three b-spline basis functions of degree 3 with equally spaced knots over [0,20]; Figure 12 shows a graphical representation of these basis functions. The three person-specific basis function coefficients, c_{i1l}, c_{i2l}, and c_{i3l} were randomly generated from the following multivariate normal distribution:

$$\begin{bmatrix} c_{i1l} \\ c_{i2l} \\ c_{i3l} \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 2 & 0 & 0 \\ 0 & 5 & 0 \\ 0 & 0 & 10 \end{bmatrix} \right) \tag{35}$$

The time points, t_{ijl}, were randomly generated from a uniform distribution over the range [0,20]. The grand mean curves and person-specific curves generated for the three different shapes of trajectories are shown in Figure 9, 10, and 11. The measurement errors, e_{ijl}, were randomly generated from a normal distribution of a mean of 0 and a variance of σ_e^2. Again, the MSE (26) and AMSE (27) were calculated for evaluating the accuracy of the estimated mean and individual trajectories under each condition.

It was hypothesized that the model would perform best with the asymptotic trajectory, the most time points, no missing data, and irregular time points. The asymptotic trajectory is the least complex trajectory and thus should be the easiest to estimate, followed by the quadratic then periodic. No missing data should produce the most accurate estimation results. The model should produce increasingly inaccurate estimates as the percentage of missingness increases with a possibility that the model might fail to run with data as sparse as the 90% missing condition. Based on the missing data literature, the time-dependent missing data condition should produce the least accurate results, followed by the randomly missing condition. However, based on the previous simulation, this difference should be minimal. Finally, random time points should produce more accurate results compared to fixed time points because the model would incorporate more information throughout the entire range of the functions. Since the curves are observed at a more diverse set of time points, the model can use that diverse information for more accurate estimation. This concept is best illustrated by example in Figure 13. The first pane of Figure 13 depicts the true underlying curve of the process under study. The second pane of Figure 13 shows a representation of what could happen if each person was sample at a fixed point in time; in this case, the oscillations are completely missed and the process would appear to be linear when it is in fact highly non-linear. The third pane of Figure 13 shows a representation of sampling people at random time points; in this case, the oscillatory nature of the process can be captured because data are collected on the people at a variety of time points.

The *sme* package (Berk, 2013) in R was used for the analysis. As noted in (17), two smoothing parameters are involved in estimation, one for the fixed and one for random effects; these smoothing parameters are denoted λ_μ and λ_ν in (17). Choice of a smoothing parameter can be obtained in several ways including cross-validation and likelihood based approaches among them AIC and BIC (Wu & Zhang, 2006). However, as Ramsay and Silverman (2005) noted, the selection of smoothing parameters can be subjective and is often best decided by examining various options. The *sme* package provides various measures that can be used for smoothing parameter selection, including AIC, AICc (corrected AIC), BIC, and BICn (BIC for longitudinal models with a sample size correction) (Berk, 2013). When smoothing parameters were chosen using these methods in the first simulation, the penalty was too severe. This resulted in a straight line fit for all conditions. This same issue occurred in the proposed simulation study; AIC, BIC, AICc, and BICn all imposed too severe of penalties and resulted in straight line fits to the data. Therefore, various combinations of smoothing parameters were sampled and examined. The best overall result was obtained when $\lambda_{\mu} = 50$ and $\lambda_{\nu} = 50$; these were the smoothing parameter values used in the simulation.

RESULTS

To illustrate the estimated mean and individual trajectories with different trajectory shapes under different sparseness levels, I included Figures 14 to 22. Figures 14, 15, and 16 show the estimated mean and individual trajectories for the quadratic trajectory under the no missing data condition at 5, 11, and 21 fixed time points, respectively. Figures 17, 18, and 19 present the estimated mean and individual trajectories for the asymptotic trajectory under the no missing data condition at 5, 11, and 21 random time points, respectively. Figures 20, 21, and 22 present the estimated mean and individual trajectories for the periodic trajectory under the no missing data condition at 5, 11, and 21 random time points, respectively.

This simulation aimed to investigate the effects of five factors on both MSE and AMSE: (1) shape of the underlying mean trajectory, (2) the degree of sparseness, (3) the percentage of missing data, (4) the type of missingness, and (5) the irregularity of time points. The results of the proposed simulation study are presented in Tables 2 to 6. Additionally, plots from the tables are shown in Figures 28 to 33. Overall, results from the simulation revealed that estimation accuracy increased with decreasing sparseness and decreasing percent missing. Similar to the preliminary simulation, estimation of the mean curves was more accurate than estimation of the individual trajectories. This is evident in the lower MSE values compared to AMSE values as shown in Figures 23 to 27. The effects of each factor on the estimation accuracy are presented in the following paragraphs.

Trajectory Shape

The main effect of trajectory shape on MSE and AMSE are presented in Figure 23. On average, the FMEM yielded high MSE and AMSE values for the periodic trajectories. This is due to the highly complex nature of the periodic trajectories; the curves contain 2.25 cycles over the range of time. Estimation becomes increasingly difficult when the curves contain numerous cycles, as these curves do. More specifically, examining Figures 20, 21, and 22 reveal why the MSE and AMSE values tend to be so high for the periodic trajectories compared to the quadratic and asymptotic trajectories. As shown in Figure 20, measuring the periodic curves only at five time points is not enough to fully capture the 2.25 cycles (in green); the estimated mean curve (in red) contained less than 2 cycles. This in turn yielded a high MSE value reflecting a huge gap between the true mean curve and the estimated mean curve. This effect rapidly decreases as the number of time points increases from 5 to 11 and 21, as evident from the improvement in fits shown in Figures 20, 21, and 22. The effect of trajectory shape shows that the more complex the underlying trajectory is, the more sampling points should be used to estimate the trajectory accurately.

The FMEM yielded the lowest MSE and AMSE values for the asymptotic trajectories, on average. The asymptotic trajectories were smoother than the quadratic and periodic trajectory which explains why the asymptotic trajectories were the most accurately estimated. The results support the hypothesis of less accurate estimation of more complex trajectories.

Sparseness Level

The simple effect of sparseness level on MSE and AMSE under each shape of trajectory is presented in Figure 24. I did not aggregate MSE and AMSE values across the three different shapes of trajectory because the MSE and AMSE values for the periodic trajectories were so high that they were not comparable to those for the quadratic and asymptotic trajectories. In general, estimation accuracy increased with decreasing sparseness for both MSE and AMSE, as expected. The estimation accuracy increases when there are more data available to use, especially for the individual trajectories. This is an intuitive result. When data are sparse, each individual's data do not contain much information and the individual curve will be less accurately estimated compared to when each curve is densely observed. Estimating individual trajectories is more difficult than estimating the mean trajectory; this is evident by higher AMSE values compared to MSE values. This result is also intuitive. Estimating an individual's curve is reliant upon how much information is available in their curve and the mean curve. If an individual's curve does not contain much information (i.e., the data are sparse) then the estimate will be shrunk towards the mean curve, resulting in a higher AMSE value. Estimating the mean curve involves using data across all individual's curves and no shrinkage is involved, thus the MSE value is not affected by shrinkage.

Percent Missingness

The simple effect of percent missing on MSE and AMSE under each shape of underlying trajectory is presented in Figure 25. Under each shape of trajectory, both the

MSE and AMSE values increased with increasing percent missingness, as expected. As more data were removed from the model, less information was available to use for estimation of the curves, and the estimation accuracy decreased. For each trajectory, the MSE values were similar for missing data rates of 0%, 30%, 50%, and even 70% . This result was partly expected and mimicked the results from the preliminary simulation where a missing data rate of 30% was used. However, it was surprising that even with 70% missing data, the estimation accuracy was comparable to 0% missing data. Once rates reached greater than 70%, MSE values began substantially increasing; there was a large spike in MSE values when missing data rates transitioned from 70% to 90%. However, when missing data rates started deviating from 0%, the AMSE values began increasing more rapidly than the MSE values. This is because as percent missingness increased, only a small amount of information was available for each individual curve, making estimation of each curve less accurate.

Type of Missingness

The simple effect of type of missingness on MSE and AMSE under each shape of trajectory is presented in Figure 26. Results for all trajectories revealed that time-dependent missing data had higher MSE and AMSE values compared to randomly missing data. Interestingly for the asymptotic trajectory, the time-dependent missingness MSE value was only slightly higher than the randomly missing MSE value. Upon examination of Figure 30, in the $N_i = 5$ random time points with missing data conditions, randomly missing data had higher MSE values than time-dependent missing data. This effect is related to how much data are present at the boundaries. When data were randomly missing, compared to time-dependent missing, there were more data available at later time points around the boundary. Depending on whether there were more data above or below the true mean curve around the boundary, the estimated mean function curved either up or down. However, when the missingness mechanism was time-dependent, the estimated mean curves flatten out because the data were so sparse at later time points around the boundary that there was not enough information used to weight the curve in either direction. Figures 34 and 35 display this effect. The plots depict the estimated mean function, for the same data set, curved either up or down with randomly missing data, and contrast it to time-dependent missing data where the estimated mean function flattens out.

Irregularity of Time Point

The simple effect of the irregularity of time points under each shape of trajectory is presented in Figure 27. Results revealed estimation accuracy for the periodic trajectories were better in terms of both MSE and AMSE with random time points. This result was expected. However, for the quadratic and asymptotic trajectories results revealed estimation accuracy was better in terms of both MSE and AMSE with fixed time points. This result was unexpected as random time points incorporate a more diverse set of data to use in estimation and should at least yield the same estimation accuracy as fixed time points.

Upon investigation of this effect, the result is again likely due to lack of information at the boundaries. When data are collected at fixed time points, every individual is observed at the boundary time points, exactly 0 and exactly 20, which provides a good deal of information at the boundary locations. However, with random time points, individual curves were sampled around 0 and 20, but not necessarily exactly at those values. This causes the estimated mean curve to deviate from the true mean at the boundaries since there is less information present. Figures 36 and 37 depict the estimated mean function, for the same data set with either fixed or random time points, deviating from the true mean function with random time points due to lack of information at the boundaries.

Effect Size Measure

In order to get an estimate of how much variance in the observed data could be accounted for by the estimated underlying function, $\hat{x}_i(t_{ij})$, a pseudo-R^2 for the *k*th data set was calculated using the formula:

$$\text{pseudo-}R_k^2 = 1 - \frac{\sum_{i=1}^{n}\sum_{j=1}^{N_i}\left(\hat{y}_{ijk} - y_{ijk}\right)^2}{\sum_{i=1}^{n}\sum_{j=1}^{N_i}\left(y_{ijk} - \bar{y}_k\right)^2} \tag{36}$$

where y_{ijk} indicates the observed data point for person *i* in the *k*th data set at the *j*th time point, $\hat{y}_{ijk}$ is the predicted value on the curve, and $\bar{y}_k$ is the mean of all observed data in the *k*th data set. Then, an average pseudo-R^2 was calculated across the 100 data sets for each condition. High values of average pseudo-R^2 indicate a better fit to the data.

The results for average pseudo-R^2 are presented in Tables 7-9. Overall, fixed time points yielded higher average pseudo-R^2 than random time points. For quadratic and asymptotic trajectories, the average pseudo-R^2 decreases as the number of time points increases. This result makes sense because, in general, more data are less likely to yield overfitting. For periodic trajectories, the average pseudo-R^2 increased as the number of time points increased (this trend is much clearer for random time points) because a relatively large number of time points are required to capture the periodic trajectories accurately. A more thorough examination of the average pseudo-R^2 is given in the discussion section.

DISCUSSION

It is worthy to note that the FMEM occasionally failed for certain data sets with random time points in the 90% time-dependent missing data conditions when a curve was sampled at 11 and 21 time points. Estimation failed because knots were placed where data were too sparse, i.e., no data existed at the knot location. When this occurred, the number of knots used was decreased. For example, when a data set with $N_i = 11$ would fail to run with 11 knots, 5 equally spaced knots were used instead. Similarly, when a data set with $N_i = 21$ would fail to run with 21 knots, 11 equally spaced knots were used instead. For each trajectory type, approximately 6% of data sets in the 90% time-dependent missing data conditions with $N_i = 11$ failed to run with 11 knots, and were successfully re-run with 5 knots. For each trajectory type, approximately 11% of data sets in the 90% time-dependent missing data conditions with $N_i = 21$ failed to run with 21 knots, and were successfully re-run with 11 knots. Rather than placing knots at the fixed time point values, knots could have been placed at equally spaced percentiles of the random time points, which could have potentially avoided failed estimation due to sparseness at knot locations. Refer to Wu and Zhang (2006, Chapter 3.3.3) for a discussion on widely used methods for selecting knot locations.

Choosing an optimal set of smoothing parameters (λ_μ and λ_v) is important for estimating curves accurately. The *sme* package provides AIC, BIC, AICc, and BICn measures for selecting smoothing parameters. The *sme* package selects the best pair of smoothing parameters based on the model with the smallest AIC, AICc, BIC, or BICn (Berk, 2013). However, when I used these four methods, all of them yielded results that were over-smoothed to the data. Essentially a linear line was fit to each data set when these selection methods were used. For instance, the AIC yielded smoothing parameter values of $\lambda_\mu = 1{,}000{,}000$ and $\lambda_\nu = 1{,}000{,}000$ when it was used to fit one of the periodic trajectory data sets with 21 fixed time points and no missing data, resulting in a very poor fit to the data set. In order to avoid over-smoothing, I sampled a set of much smaller values. Specifically, I sampled values between 1 and 50 for each smoothing parameter and examined the fits of the models as the smoothing parameter values changed. There were only minor differences between results obtained under each combination of smoothing parameters, so values of $\lambda_\mu = 50$ and $\lambda_\nu = 50$ were selected as the final smoothing parameters. This is obviously a highly subjective approach and requires some knowledge of what the trajectories should look like. Without knowing what shape the true underlying trajectories should be, researchers could be at risk of fitting curves that are not accurate representations of the true underlying curves. Investigations as to why the suggested methods (e.g., AIC, BIC, etc.) performed poorly are currently being carried out. Clearly, a more robust approach to selecting smoothing parameters is needed.

Typically, the R^2 value is used as a measure of how much variation in the data is accounted for by the effect of interest and is used as a goodness-of-fit measure. Cohen (p. 159; 1992) reports values of .0196, .1304, and .2592 as small, medium, and large effects for R^2 values. These benchmarks were developed with parametric relationships in mind, specifically a linear relationship between two variables. It is likely the case that these benchmarks will not be suitable for non-parametric relationships. For instance, it is very clear that Figure 20 is a poor fit to the data. However, the average pseudo-R^2 for that condition is .47. Cohen's benchmarks would suggest that this was a good model for the data, when clearly it is not. Since the researcher can control the goodness-of-fit of the FMEM by selecting smoothing parameter values that under-smooth the data, the R^2 values can become artificially increased without actually reflecting a proper model. Therefore, rather than using R^2 benchmarks like Cohen developed, it could be useful to compare R^2 values obtained from FMEMs to R^2 values obtained from non-linear parametric growth curve models as a way to gauge how accurately the model is fitting the data.

As previously mentioned, I considered the signal-to-noise ratio (SNR) when choosing the error variance. The SNR indicated the variance of the true signal (i.e., the data without being perturbed by error) to the variance of the error. More specifically, the error variance level was chosen to maintain an approximate signal-to-noise ratio (SNR) around 10. To confirm this was the case, the SNR was calculated for each trajectory, at each sparseness level ($N_i = 5,11,21$) for the fixed time points. The results for the SNR are presented in Table 10. The quadratic trajectory maintained the highest SNR over both asymptotic and periodic trajectory shapes. This indicates the quadratic trajectory has the highest signal variance and the periodic trajectory has the lowest signal variance. The periodic trajectory has the lowest SNR across each level of sparseness and interestingly also has the most stable SNR. The SNR ratio for the periodic trajectory ranges from 7.31 to 7.64, while the other trajectories range approximately from 10 to 16 and 7 to 13. In general, as the number of time points increased, the SNR decreased. When the curve is sampled at a smaller number of time points the signal variance is relatively large. However, as the curve is sampled at more time points the difference between the time points becomes smaller, decreasing the signal variance. Thus the SNR becomes smaller since the signal variance decreases, while the denominator remains constant. If the SNR were calculated for the random time points, a smaller signal variance is expected. Therefore, a smaller SNR is expected for the random time points. Interestingly, the SNR for the periodic trajectory slightly increases as the number of time points increases. This is likely due to the fact that 5 time points contains hardly any signal for the complex trajectory, but as the number of time points increases more signal is introduced.

The results of the simulation study support the notion that FMEMs can be useful tools for researchers who work with longitudinal data. Results revealed the estimation of the mean curve was accurate even when curves were observed at a few time points by design, data contained high amounts of missingness, which can be, time-dependent, and when data were generated from complex underlying trajectories. Though, caution should be taken when using a FMEM when data are generated from a complex underlying trajectory but are only observed at a few time points, and when missing data rates exceed 70%. Additionally, as the mean square error tended to be higher at the boundaries of the data, researchers should collect data at the boundaries of interest to reduce this error and obtain more accurate results.

These results can help researchers in designing their experiments to maximize the benefit of using a FMEM. Researchers studying phenomena with complex underlying trajectories should plan to collect a larger number of time points (e.g., at least 20). For instance, researchers seeking to measure phase synchrony in swinging pendulums (Fine, Likens, Amazeen, & Amazeen, 2015) should be sure to collect a large number of time points in order to accurately capture the oscillatory nature of the phenomena. These results also illustrate the importance of collecting data at the boundaries of interest for the researchers. For example, if a researcher is interested in studying growth in babies from birth to 3 months, it is advised to collect data from each baby at birth and at the end of the study (3 months) as much as possible in order to obtain an accurate estimate of the growth curve.

FMEMs hold great promise in exploratory data analysis. FMEMs offer the flexibility to view data without strict assumptions on the functional form. Thus, it makes a great tool for describing changes over time, especially at an early stage of investigation. When researchers are not certain about what parametric models should be used, starting analyses with a FMEM can guide researchers to select an appropriate parametric form. It can also inform researchers as to the appropriateness of a parametric model. In addition, the nonparametric nature of FMEMs enables researchers to capture the aspects of change that might be ignored by parametric models.

This study had some limitations. The foremost issue is the selection of smoothing parameters, as previously discussed. The other limitation of this study is that the MSE and AMSE values might not be easily interpretable. An MSE value of 15 is better than an MSE value of 50, but it is hard to tell whether 15 is an acceptable level or not. Therefore, the MSE and AMSE values do not have any objective criteria against which the performance of the FMEM can be evaluated. It would be useful to compare the MSE and AMSE values obtained from a FMEM to those obtained from parametric nonlinear growth curve models (Grimm, Ram, & Hamagami, 2011). This comparison will enable researchers to gauge how well the FMEM performs compared to plausible parametric models.

It would be largely beneficial to extend this work in two main directions: smoothing parameter selection and adding covariates to the model. As discussed, subjectively selecting smoothing parameters is a drawback if researchers do not have any a priori knowledge about the shape of the trajectories that should be observed. It would be beneficial to further investigate other software packages to see what techniques are used to select smoothing parameters. A comparison of likelihood based approaches (e.g., AIC, BIC) and generalized cross-validation (GCV; Craven & Wahba, 1979) would help identify the more robust method of selection. Additionally, the FMEMs considered in this thesis do not consider covariate when estimating individual and mean trajectories. However, most researchers are often interested in more than just describing the trajectories over time. How the shape of the trajectories is affected by covariates such as age, gender, and treatment condition is often of main interest to the researcher. Extending this work to incorporate covariates, and potentially allowing those covariates to account for missingness, would make FMEMs more useful (Green & Silverman, 1994, Chapter 4; Wu & Zhang, 2006, Chapter 8).

REFERENCES

Adler, J. M. (2012). Living into the story: Agency and coherence in a longitudinal study of narrative identity development and mental health over the course of psychotherapy. *Journal of Personality and Social Psychology, 102*(2), 367-389. doi:http://dx.doi.org/10.1037/a0025289.

Bennett, D. (2001). How can I deal with missing data in my study? *Australian and New Zealand Journal of Public Health*, 25(5), 464-469.

Berk, M (2013). sme: Smoothing-splines Mixed-effects Models. R package version 0.8. h ttp://CRAN.R-project.org/package=sme

Berk, M. (2012). Statistical Methods for Replicated, High-Dimensional Biological Time Series (Doctoral dissertation). Retrieved from http://wwwf.imperial.ac.uk/~mab201/MauriceBerkPhDThesisFinal.pdf.

Cinciripini, P., Lapitsky, L., Seay, S., Wallfisch, A., Kitchens, K., & Vunakis, H. (1995). The effects of smoking schedules on cessation outcome: Can we improve on common methods of gradual and abrupt nicotine withdrawal? *Journal of Consulting and Clinical Psychology,* 63(3), 388-399.

Cohen, J. (1992). A power primer. *Psychological Bulletin*, 112, 155-159. doi:10.1037/0033-2909.112.1.155

Craven, P., & Wahba, G. (1979). Smoothing noisy data with spline functions: Estimating the correct degree of smoothing. *Numer. Math. Numerische Mathematik*, 31(4), 377-403. doi:10.1007/bf01404567

Dahlquist, G., & Bjorck, A. (1974). *Numerical methods*. Englewood Cliffs, N.J.: Prentice-Hall.

De Boor, C. (2001). *A practical guide to splines (Revised Edition).* New York: Springer-Verlag.

Di, C., Crainiceanu, C., Caffo, B., & Punjabi, N. (2009). Multilevel functional principal component analysis. *Ann. Appl. Stat. The Annals of Applied Statistics*, 3(1), 458-488.

Enders, CK. (2010). *Applied missing data analysis*. New York: Guilford Press.

Esser, G., Schmidt, M. H. & Woerner, W. (1990). Epidemiology and Course of Psychiatric Disorders in School-Age Children—Results of a Longitudinal Study. *Journal of Child Psychology and Psychiatry*, 31: 243–263. doi: 10.1111/j.1469-7610.1990.tb01565.x.

Fine, J. M., Likens, A. D., Amazeen, E. L., & Amazeen, P. G. (2015). Emergent complexity matching in interpersonal coordination: Local dynamics and global variability. *Journal of experimental psychology: Human Perception and Performance*, *41*(3), 723

Green, P., & Silverman, B. (1994). *Nonparametric regression and generalized linear models: A roughness penalty approach* (1st ed.). London: Chapman & Hall.

Grimm, K. J. (2008). Longitudinal associations between reading and mathematics. *Developmental Neuropsychology*, 33, 410-426.

Grimm, K. J., Ram, N. and Hamagami, F. (2011), Nonlinear Growth Curves in Developmental Research. Child Development, 82: 1357–1371. doi:10.1111/j.1467-8624.2011.01630.x

Hankin, B., Abramson, L., Moffitt, T., Silva, P., Mcgee, R., & Angell, K. (1998). Development of depression from preadolescence to young adulthood: Emerging gender differences in a 10-year longitudinal study. *Journal of Abnormal Psychology*, 107, 128-140.

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning: Data mining, inference, and prediction* (Second ed.). Springer.

James, G. (2010). Sparse Functional Data Analysis. In *The Oxford handbook of functional data analysis*. Oxford: Oxford University Press.

James, G., Hastie, T., & Sugar, C. (2000). Principal component models for sparse functional data. *Biometrika,* 587-602.

Laird, NM (1988). Missing data in longitudinal studies. *Statistics in Medicine,* 305 –315.

Laird, NM & Ware, JH. (1982). Random-effects models for longitudinal data. *Boimetrics,* 963 –974.

Leon, S. (2010). Vector Spaces. In *Linear algebra with applications* (8th ed., pp. 110-159). Upper Saddle River, New Jersey: Prentice Hall.

Levitin, D. J., Nuzzo, R., Vines, B. W., & Ramsay, J. O. (2007). *Introduction to functional data analysis. Canadian Psychology, 48(3), 135-155. doi:* 10.1037/cp2007014.

McCulloch, C., Searle, S., & Neuhaus, J. (2001). *Generalized, linear, and mixed models* (2nd ed.). New York: John Wiley & Sons.

Peugh, J., & Enders, C. (2004). Missing Data in Educational Research: A Review of Reporting Practices and Suggestions for Improvement. *Review of Educational Research*, 74(4), 525-556.

Porter, L., Crampon, W., & Smith, F. (1976). Organizational commitment and managerial turnover: A longitudinal study. *Organizational Behavior and Human Performance*, 15(1), 87-98. doi:10.1016/0030-5073(76)90030-1.

Ramsay, J. O., & Silverman, B. W. (2005). *Functional data analysis*. New York: Springer.

Raudenbush, S., & Bryk, A. (2002). *Hierarchical linear models: Applications and data analysis (2nd ed.)*. Thousand Oaks: Sage Publications.

Rubin, D. (1976). Inference and Missing Data. *Biometrika, 63*(3), 581-592.

Searle, S., Casella, G., & McCulloch, C. (1992). *Variance components*. New York: Wiley.

Snijders, T., & Bosker, R. (1999). *Multilevel analysis: An introduction to basic and advanced multilevel modeling*. London: Sage.

Swamy, P. (1971). *Statistical inference in random coefficient regression models*. Berlin: Springer-Verlag.

Ullah, S., & Finch, C. (2013). Applications of functional data analysis: A systematic review. *BMC Medical Research Methodology*, 43-43.

Ward, K.L. & Suk, H.W. (2015, July). *A comparison of different approaches to handling missing data in functional data analysis*. Poster presented at the 2015 International Meeting of the Psychometric Society, Beijing, China.

Wolfinger, R., Tobias, R., & Sall, J. (1994). Computing Gaussian Likelihoods and Their Derivatives for General Linear Mixed Models. *SIAM J. Sci. Comput. SIAM Journal on Scientific Computing,* 1294-1310.

Wu, H., & Zhang, J. (2006). *Nonparametric regression methods for longitudinal data analysis: Mixed-effects modeling approaches*. Hoboken, N.J.: Wiley-Interscience.

Yang, M. & Maxwell, S.E. (2014). Treatment effects in randomized longitudinal trials with different types of non-ignorable dropout. *Psychological Methods, 19*, 188-210.

Yao, F., Müller, H., & Wang, J. (2005). Functional Data Analysis for Sparse Longitudinal Data. *Journal of the American Statistical Association,* 577-590.

APPENDIX A

FIGURES

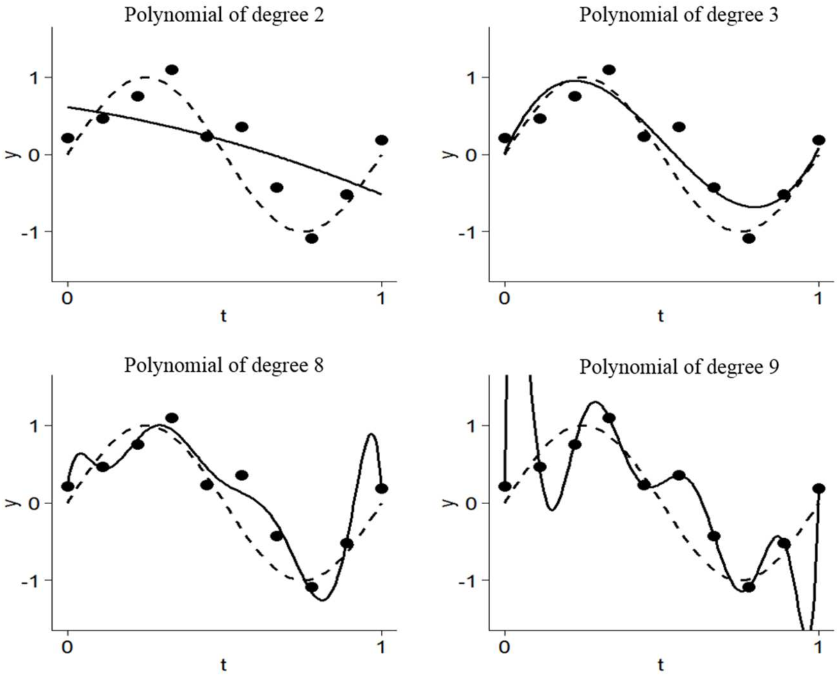

Figure 1. Plots of polynomials with various degrees fitted to simulated data. The dots indicate simulated data points, the dashed lines indicate the true underlying curve, and the solid lines are the fitted polynomials.

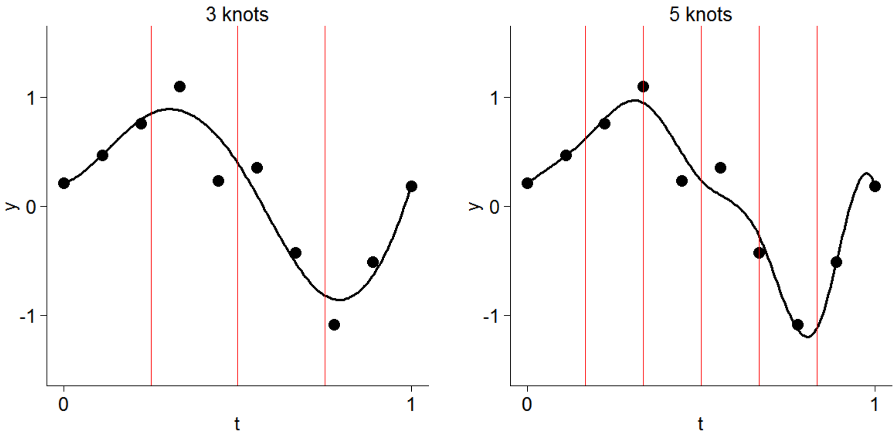

Figure 2. The left panel shows a regression spline of degree 4 with three equally spaced knots and the right panel shows a regressions spline of degree 4 with five equally spaced knots. Dots indicate simulated data points, solid lines are the fitted regression splines, and vertical lines indicates the locations of the knots.

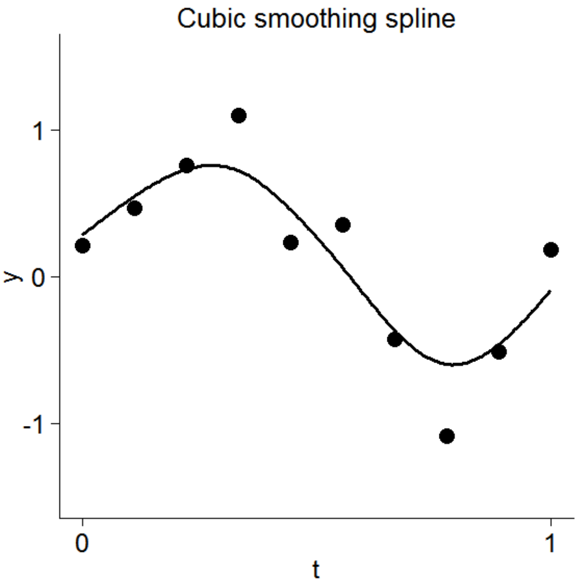

Figure 3. The estimated cubic spline smoothing with λ=0.45

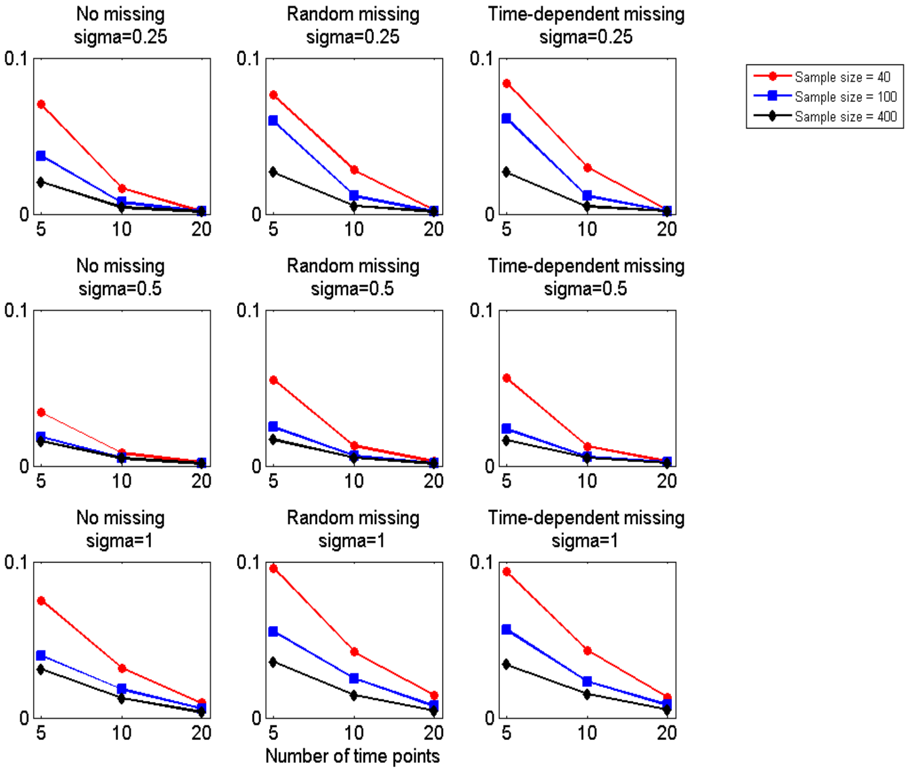

Figure 4. MSE scores for the estimated means curve across all conditions.

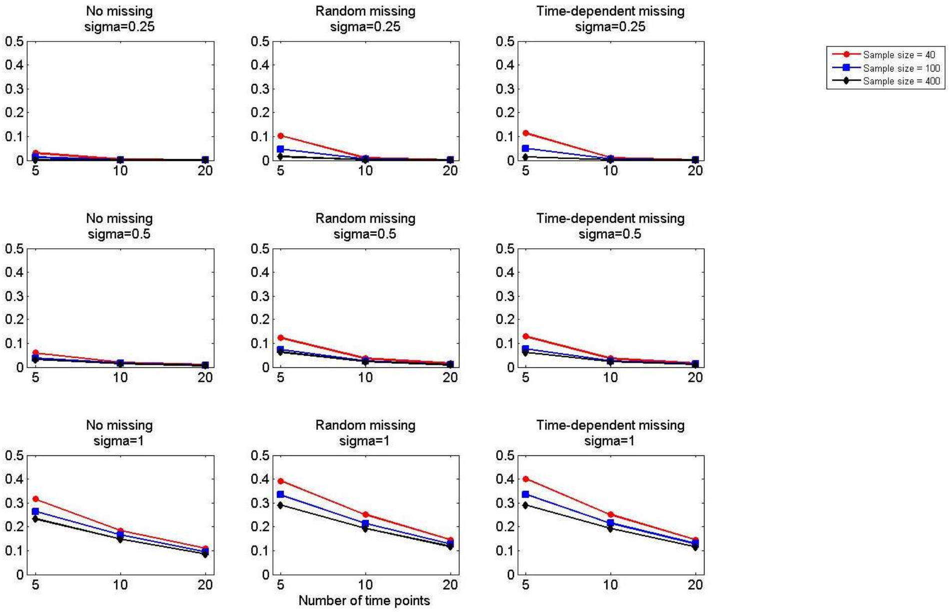

Figure 5. AMSE scores for the estimated individual curves across all conditions.

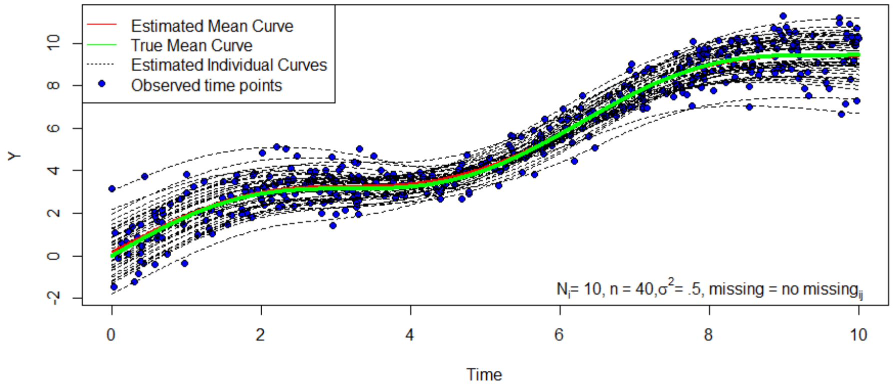

Figure 6. Model fit for a no missing data condition. This illustrates the model fit for 40 curves observed at 10 time points with no missing data, and error variance level of .5.

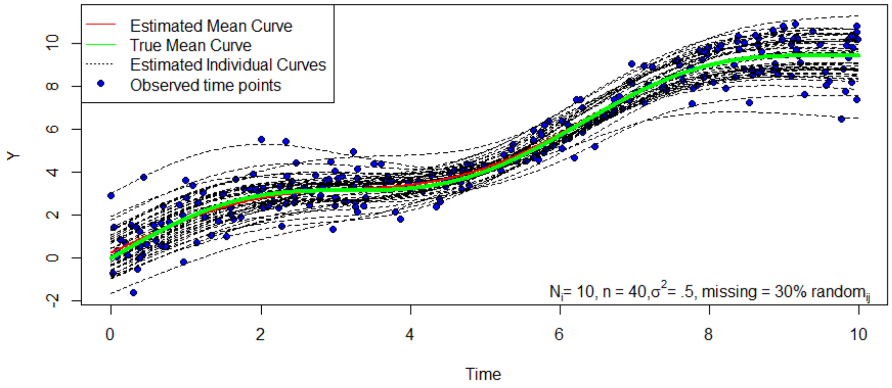

Figure 7. Model fit for a randomly missing data condition. This illustrates the model fit for 40 curves observed at 10 time points with randomly missing data, and error variance level of .5.

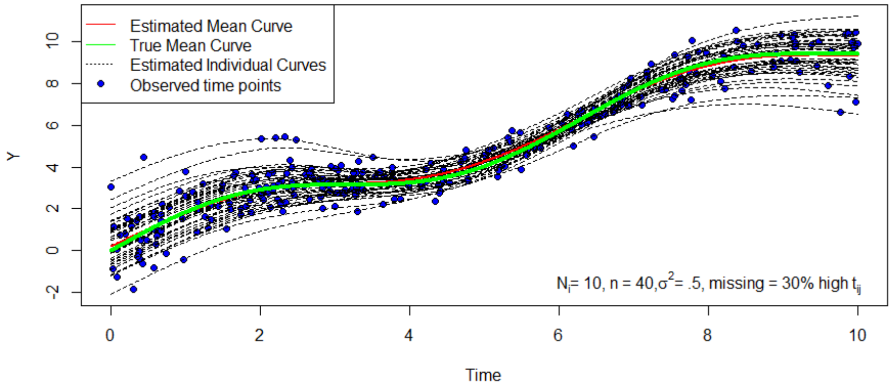

Figure 8. Model fit for a time-dependent missing data condition. This illustrates the model fit for 40 curves observed at 10 time points with time-dependent missing data, and error variance level of .5.

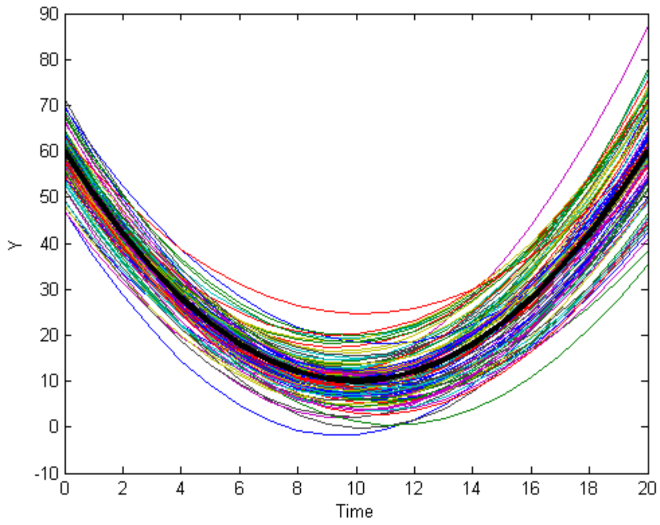

Figure 9. The shape of the quadratic grand mean trajectory (black) and individual trajectories (colored) used in the simulation study.

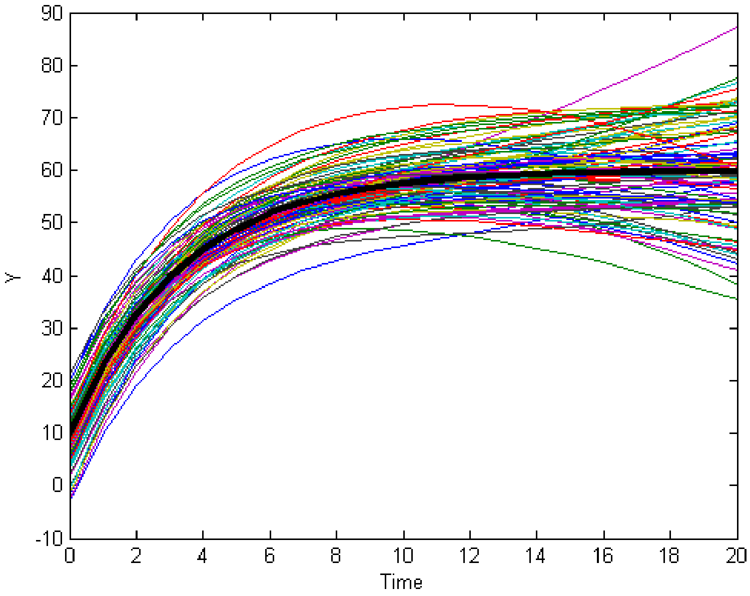

Figure 10. The asymptotic grand mean trajectory (black) and individual trajectories (colored) used in the simulation study.

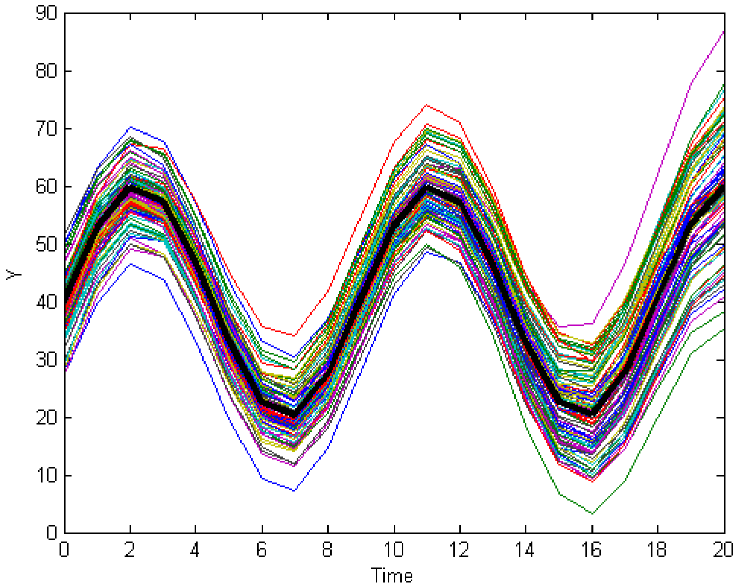

Figure 11. The periodic grand mean trajectory (black) and individual trajectories (colored) used in the simulation study.

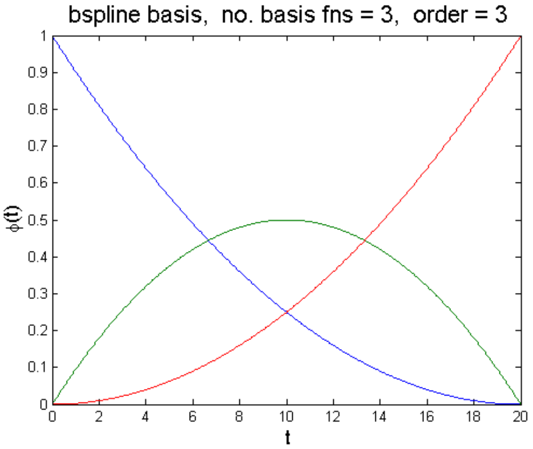

Figure 12. The three third-degree B-spline basis functions of degree 3 with equally spaced knots over [0, 20] used to generate the individual trajectories in the proposed simulation.

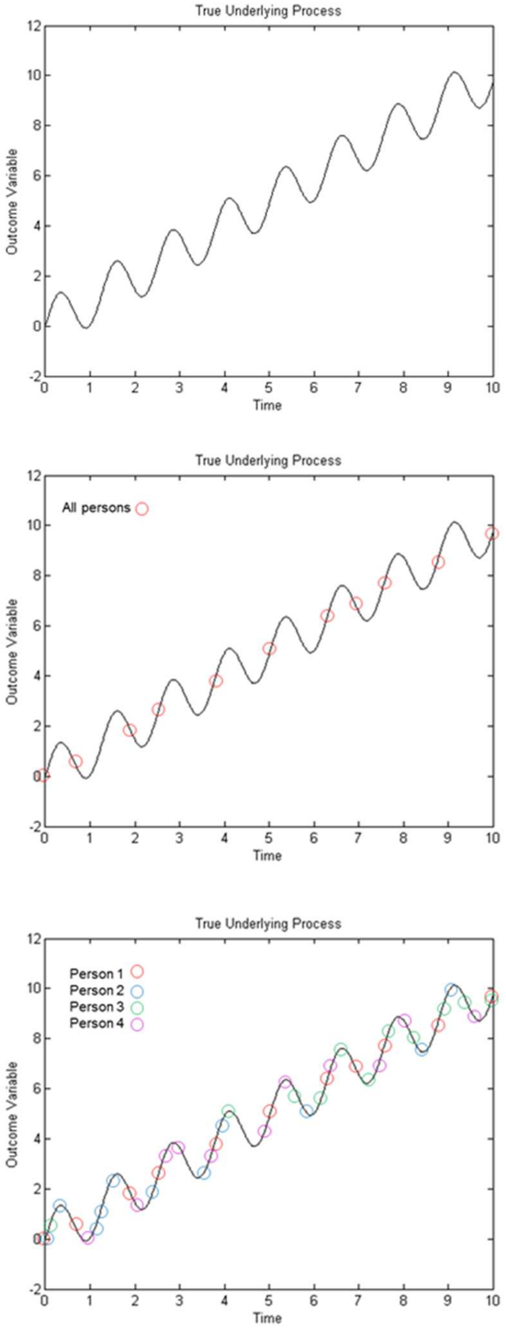

Figure 13. Plot depicting benefit of sampling participants at random time points versus fixed time points.

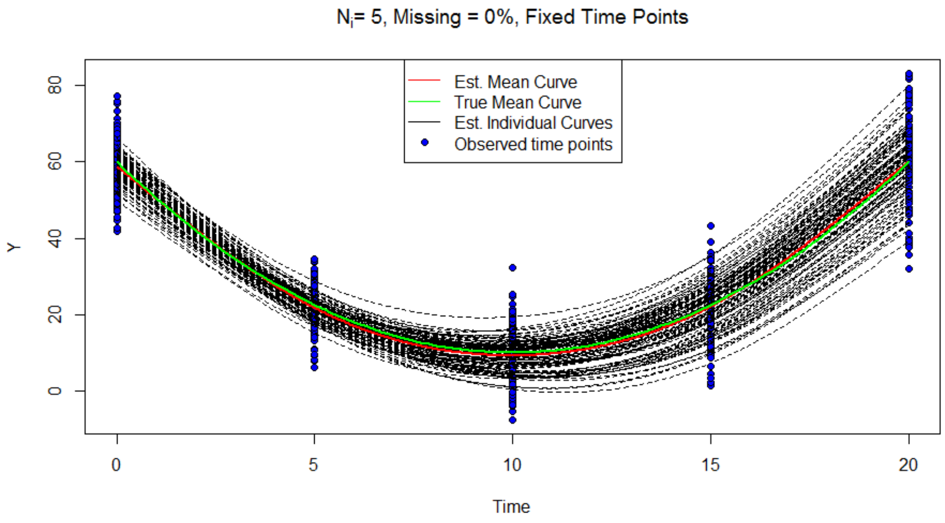

Figure 14. Plot of fitted model for the quadratic trajectory with fixed time points, no missing data, and 5 time points.

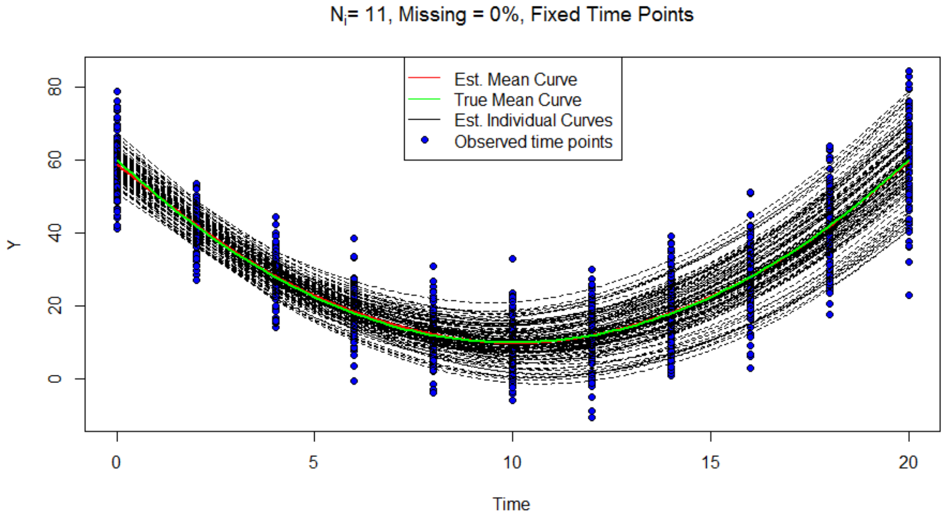

Figure 15. Plot of fitted model for the quadratic trajectory with fixed time points, no missing data, and 11 time points.

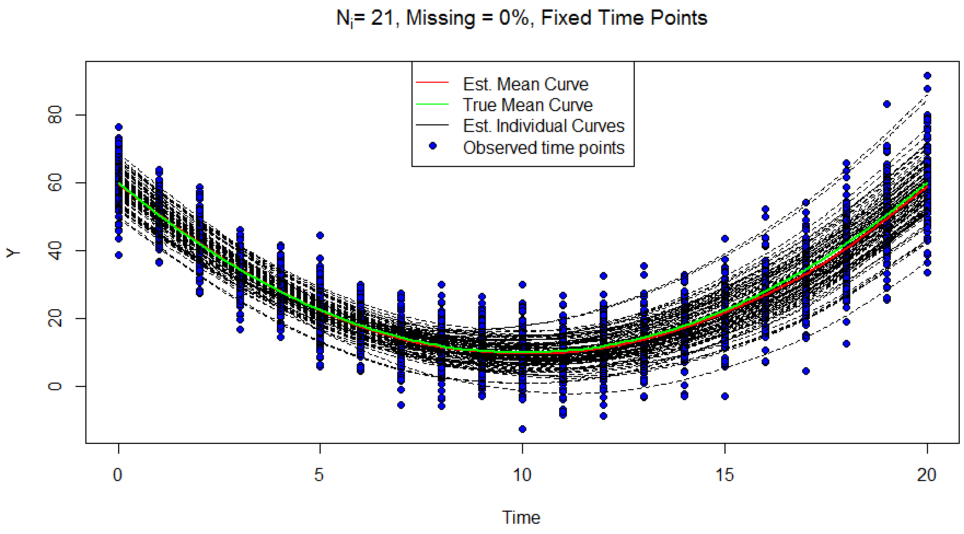

Figure 16. Plot of fitted model for the quadratic trajectory with fixed time points, no missing data, and 21 time points.

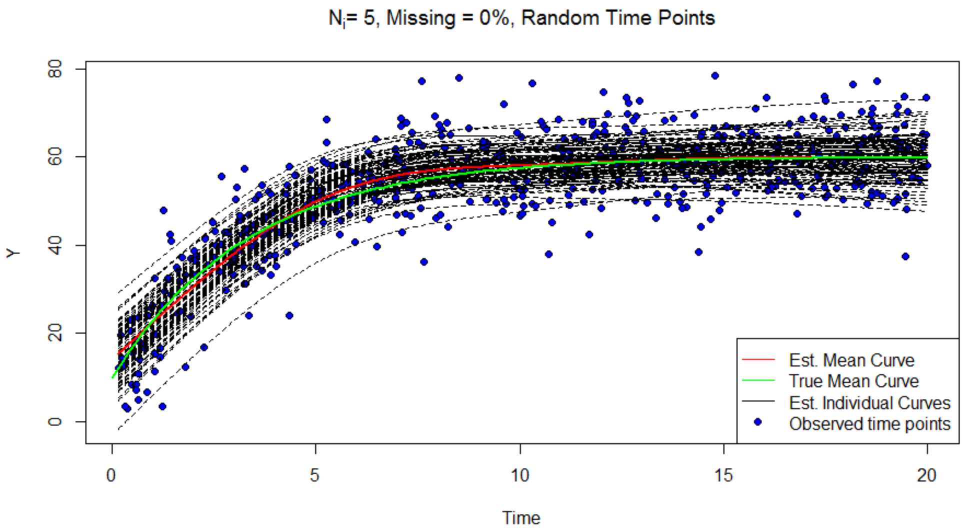

Figure 17. Plot of fitted model for the asymptotic trajectory with random time points, no missing data, and 5 time points.

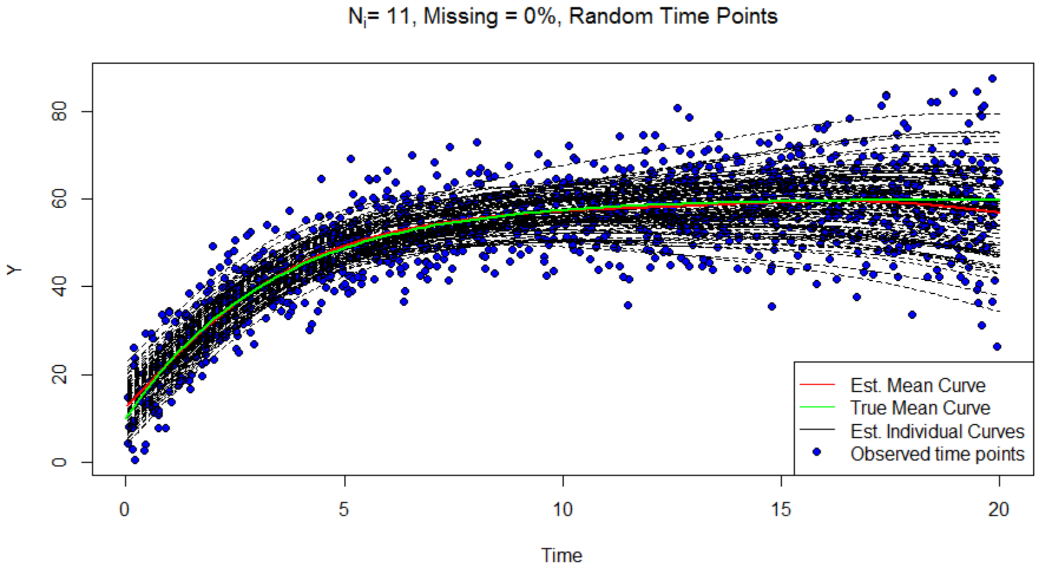

Figure 18. Plot of fitted model for the asymptotic trajectory with random time points, no missing data, and 11 time points.

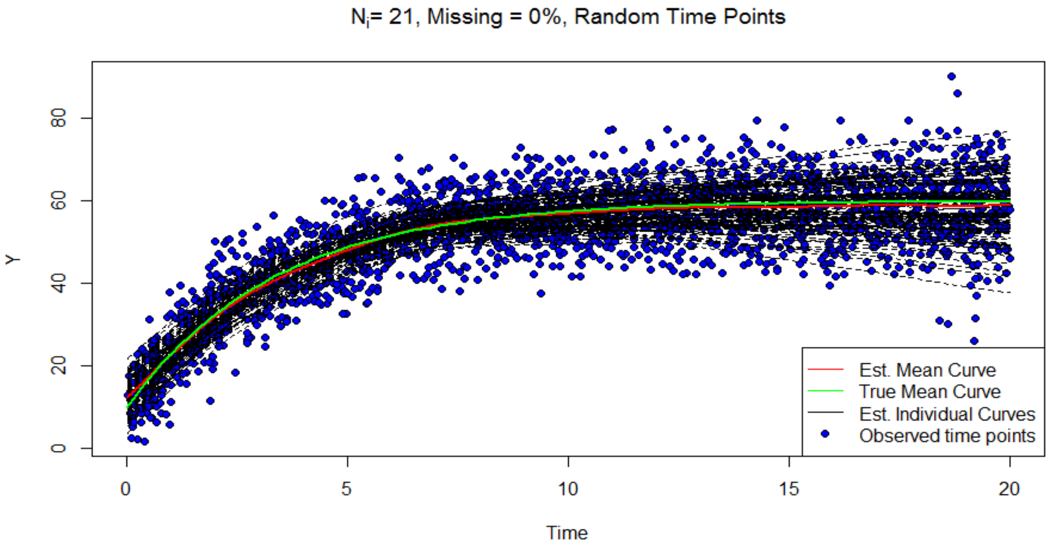

Figure 19. Plot of fitted model for the asymptotic trajectory with random time points, no missing data, and 21 time points.

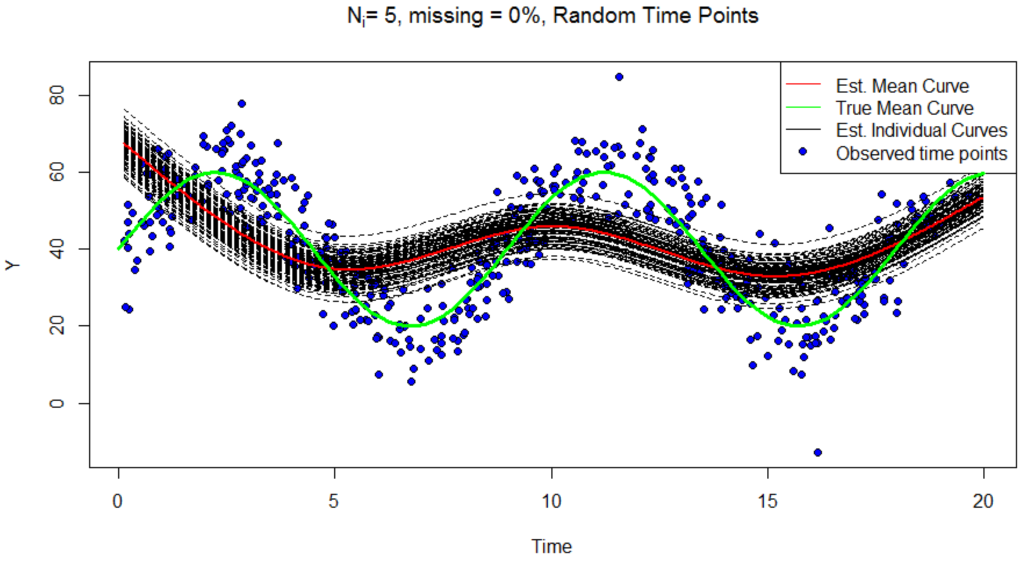

Figure 20. Plot of fitted model for the periodic trajectory with random time points, no missing data, and 5 time points.

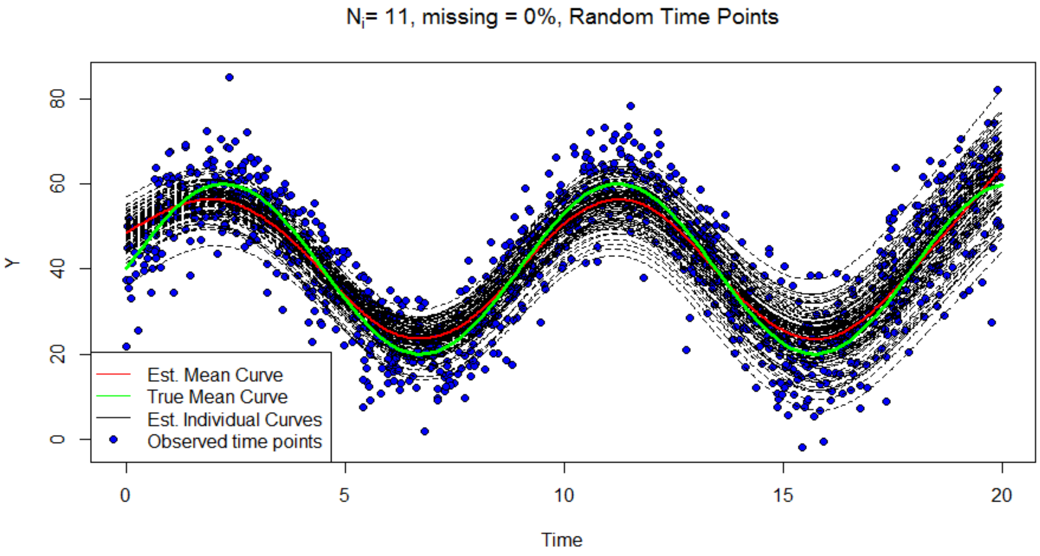

Figure 21. Plot of fitted model for the periodic trajectory with random time points, no missing data, and 11 time points.

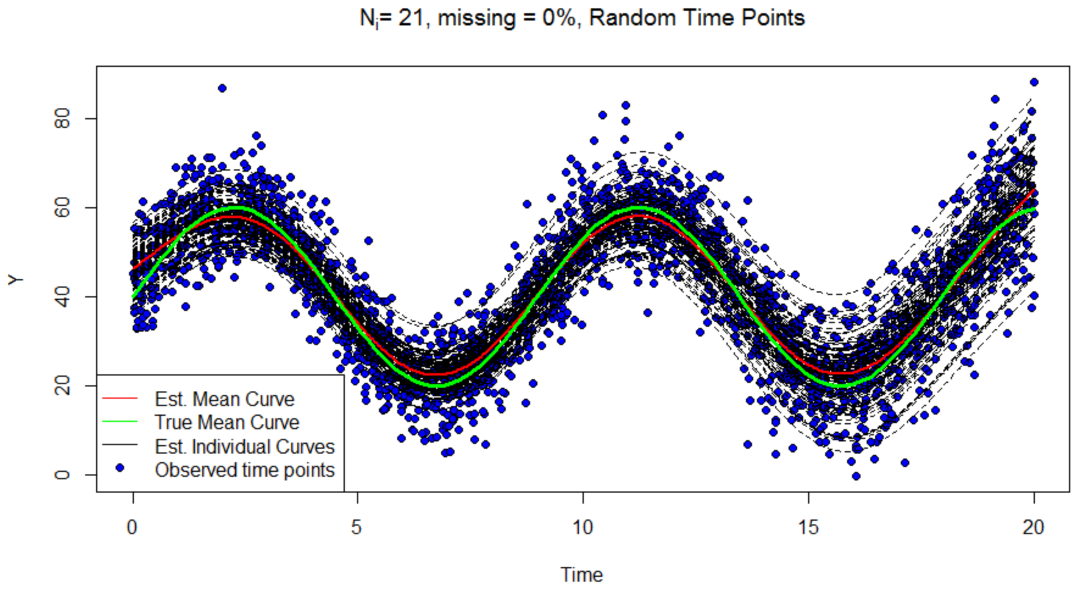

Figure 22. Plot of fitted model for the periodic trajectory with random time points, no missing data, and 21 time points.

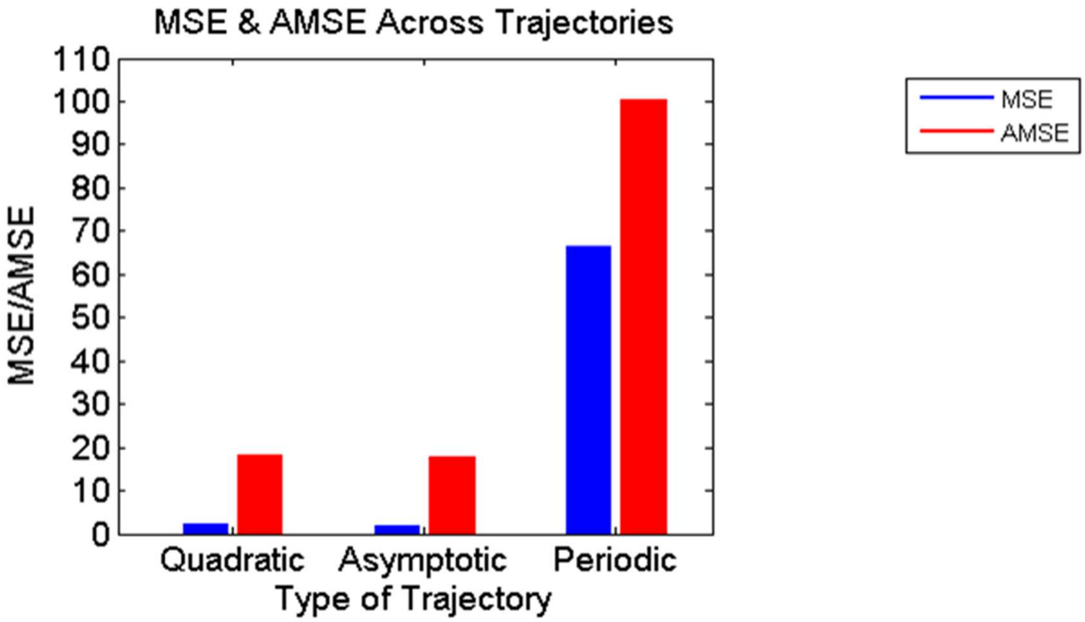

Figure 23. Plot of MSE and AMSE values for the three shapes of trajectory: quadratic, asymptotic, and periodic.

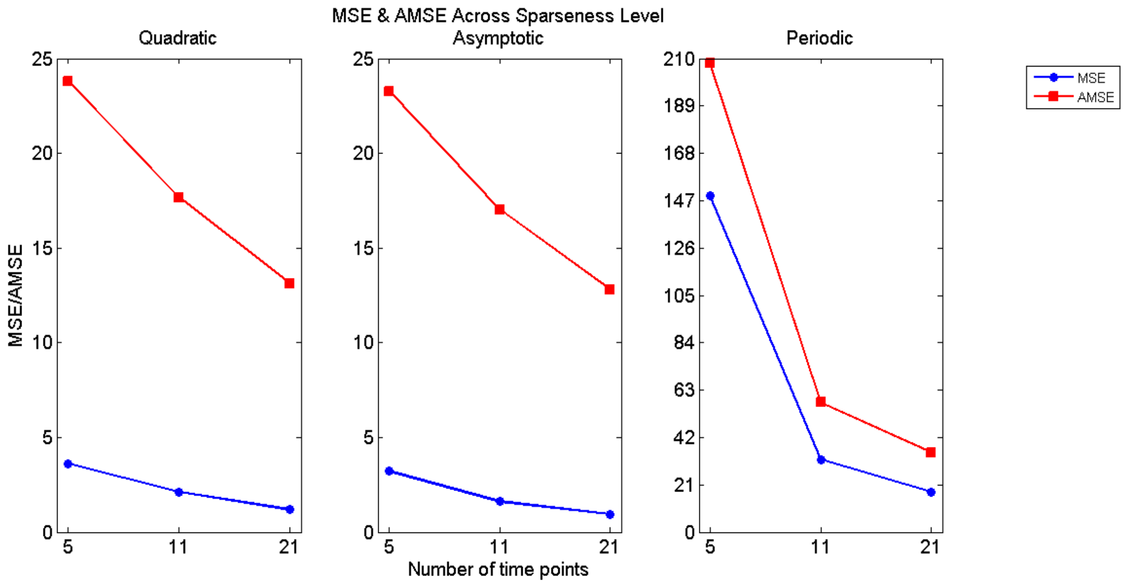

Figure 24. Plot of MSE and AMSE values for the three levels of number of time points: 5, 11, and 21.

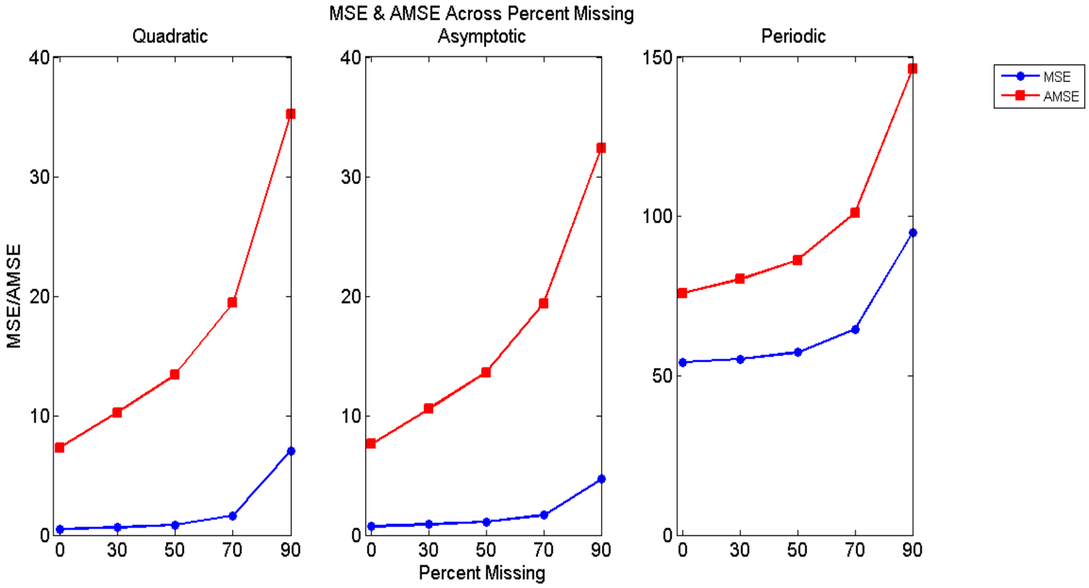

Figure 25. Plot of MSE and AMSE values for the five degrees of percent missing: 0, 30, 50, 70, and 90.

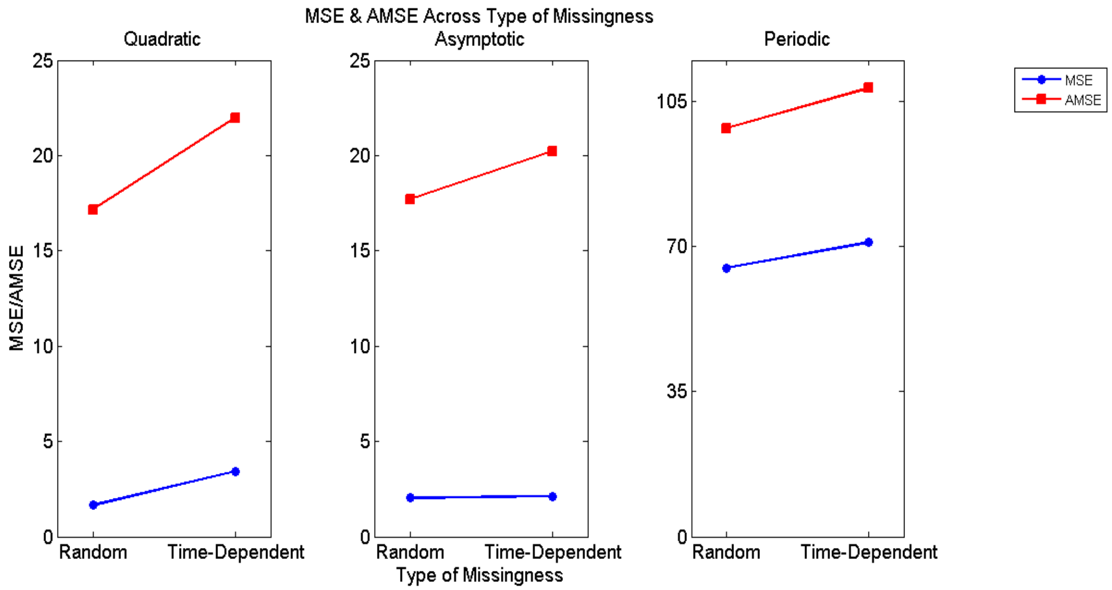

Figure 26. Plot of MSE and AMSE values for the two types of missingness: random and time-dependent.

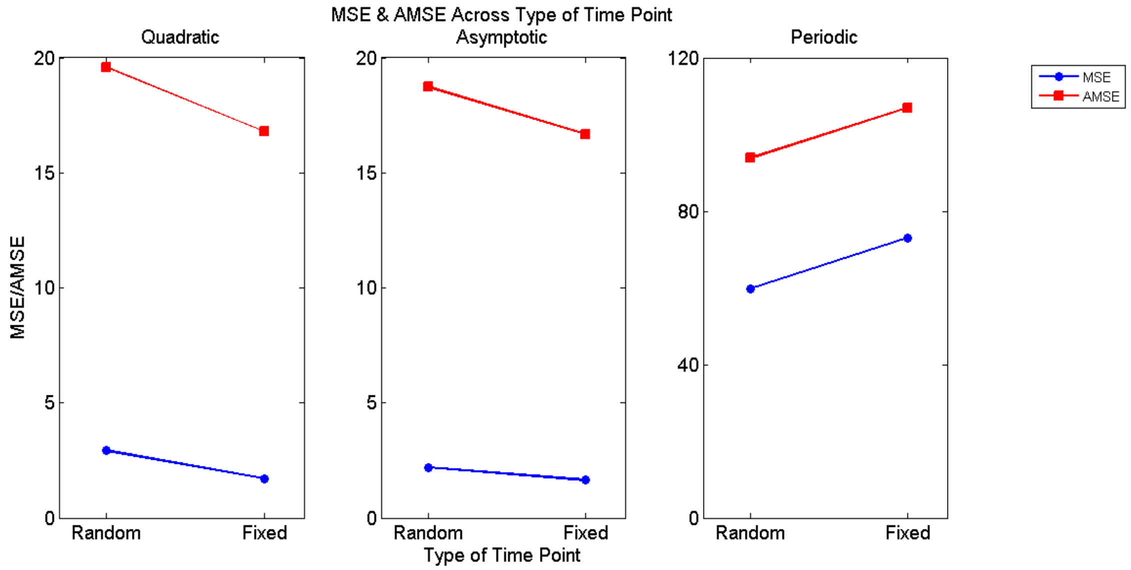

Figure 27. Plot of MSE and AMSE values for the two types of irregularity of time points: random and fixed.

.

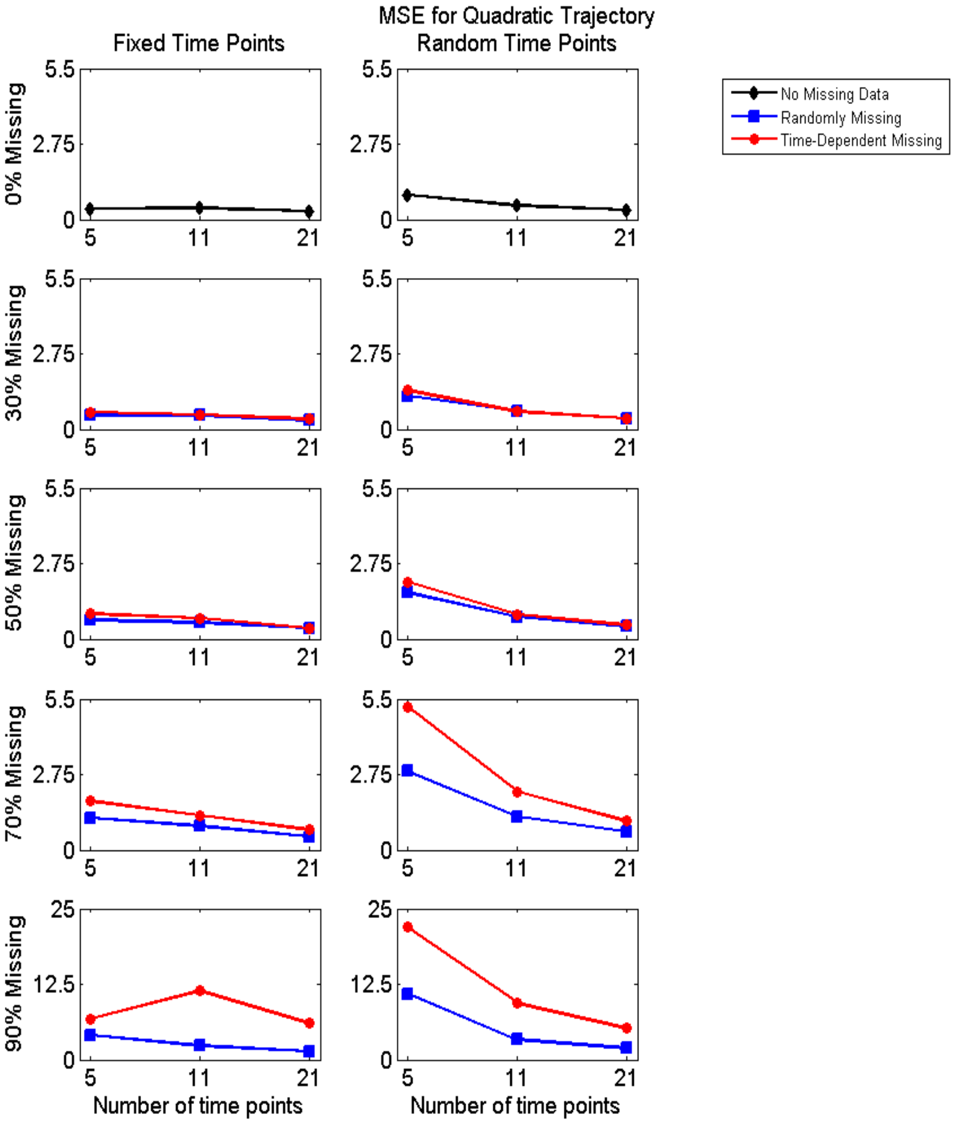

Figure 28. Plots of MSE values for the quadratic trajectory across the levels of sparseness, percent missingness, type of missingness, and type of time point.

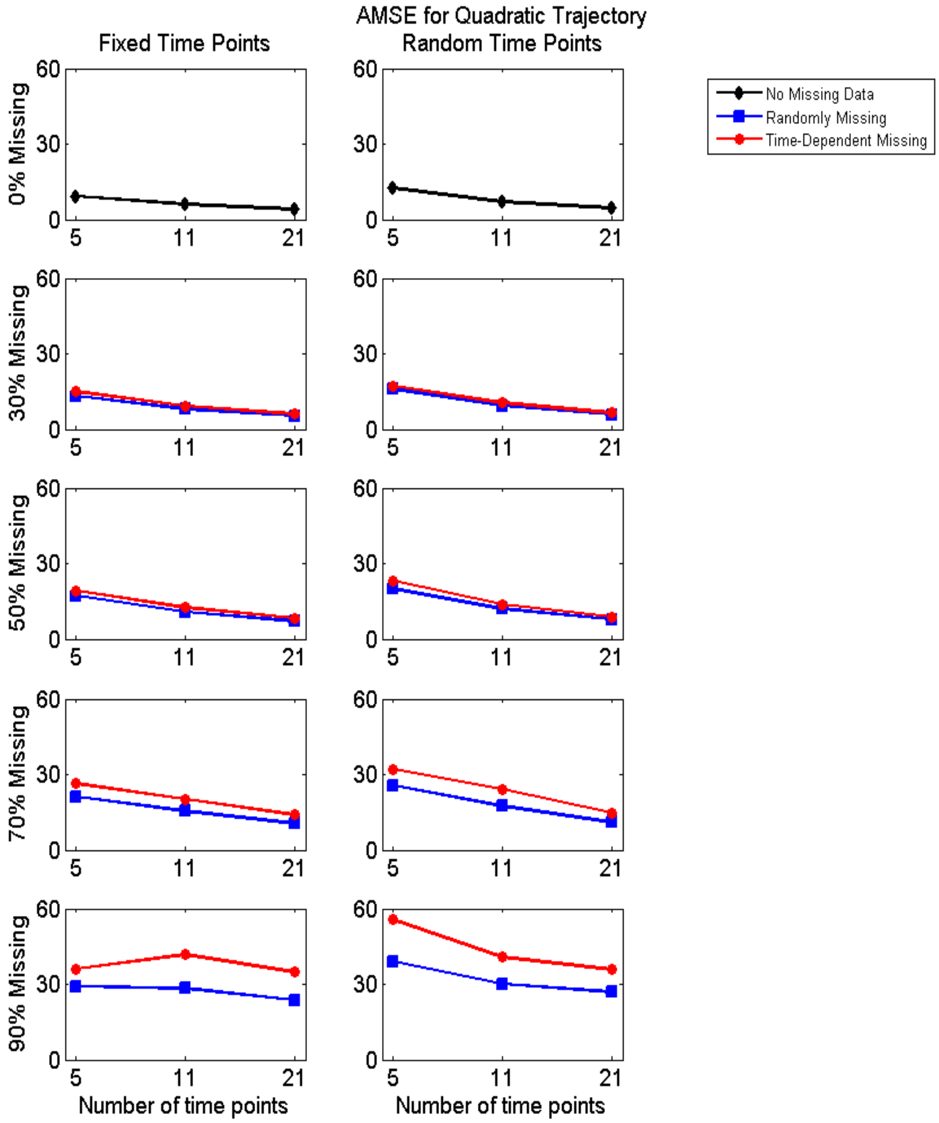

Figure 29. Plots of AMSE values for the quadratic trajectory across the levels of sparseness, percent missingness, type of missingness, and type of time point.

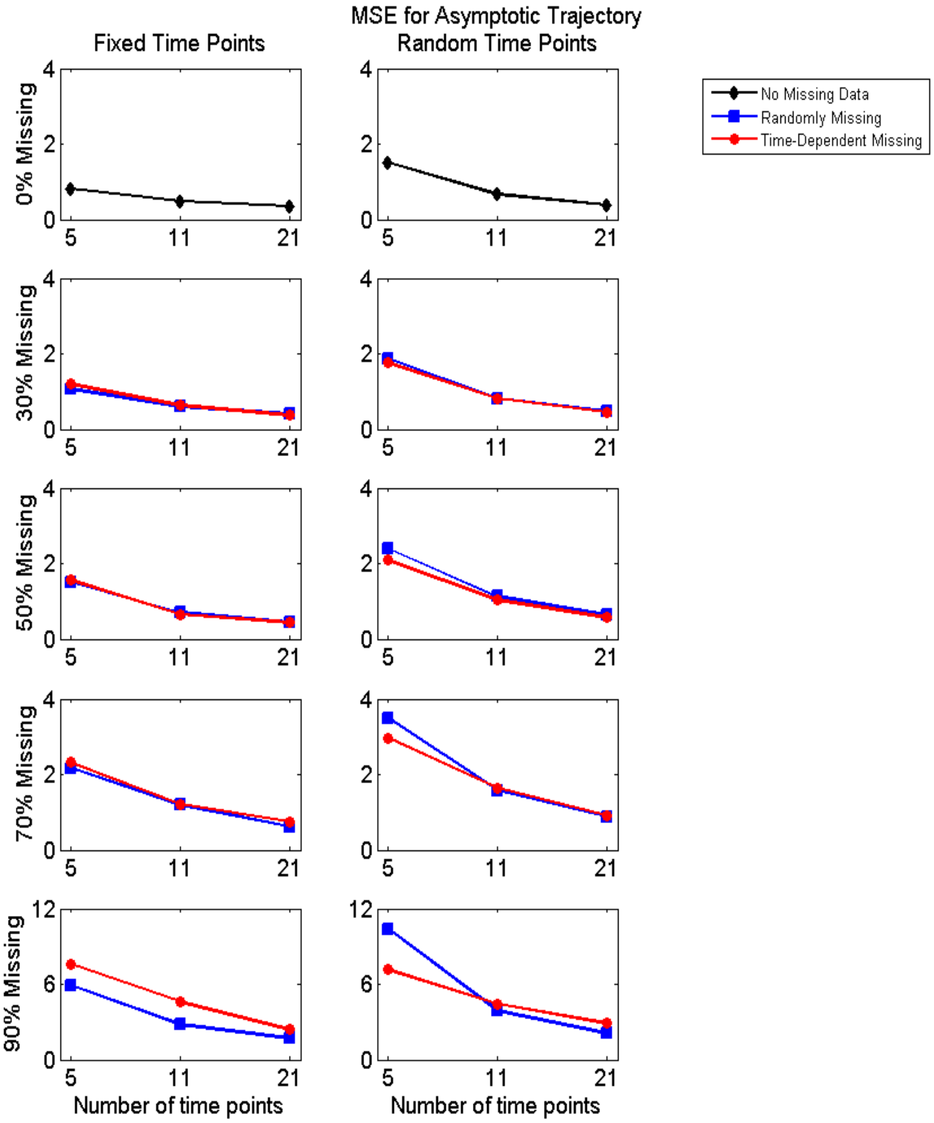

Figure 30. Plots of MSE values for the asymptotic trajectory across the levels of sparseness, percent missingness, type of missingness, and type of time point.

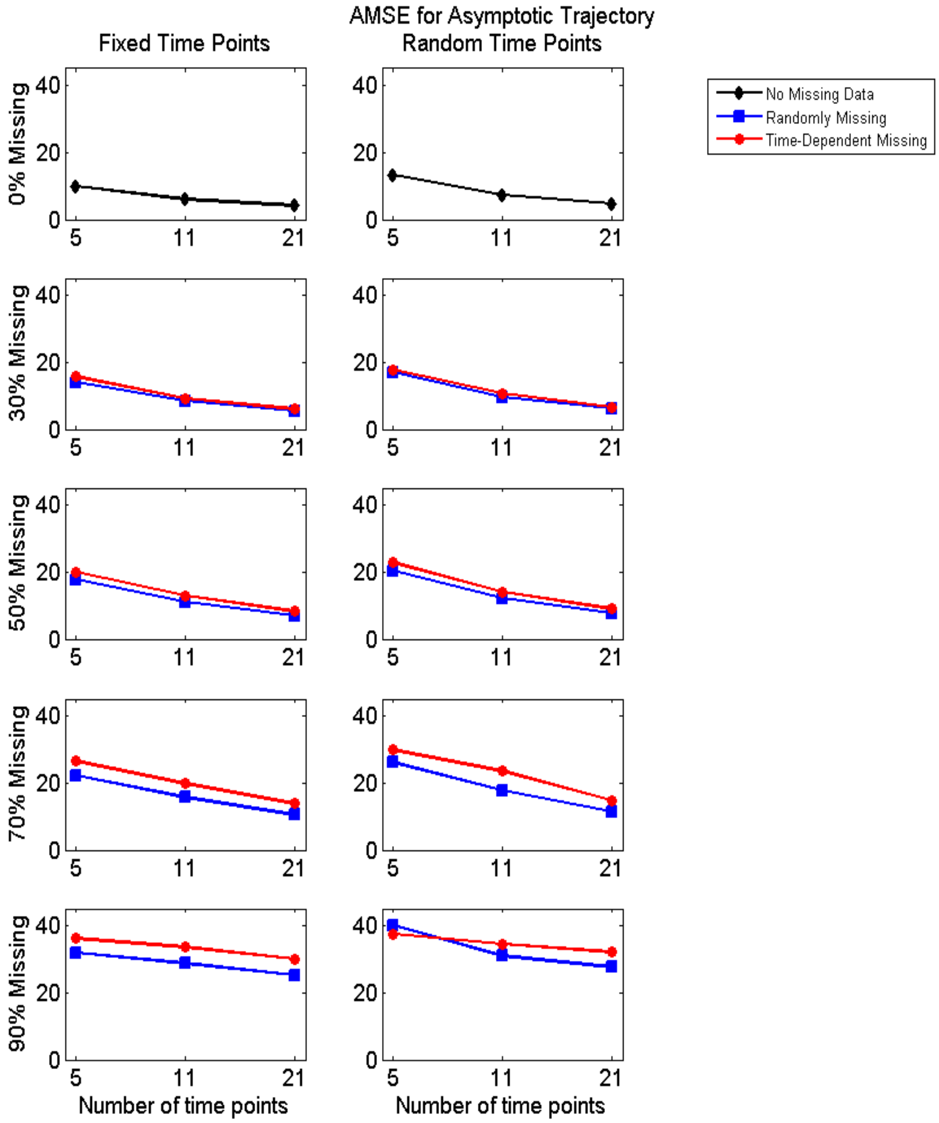

Figure 31. Plots of AMSE values for the asymptotic trajectory across the levels of sparseness, percent missingness, type of missingness, and type of time point.

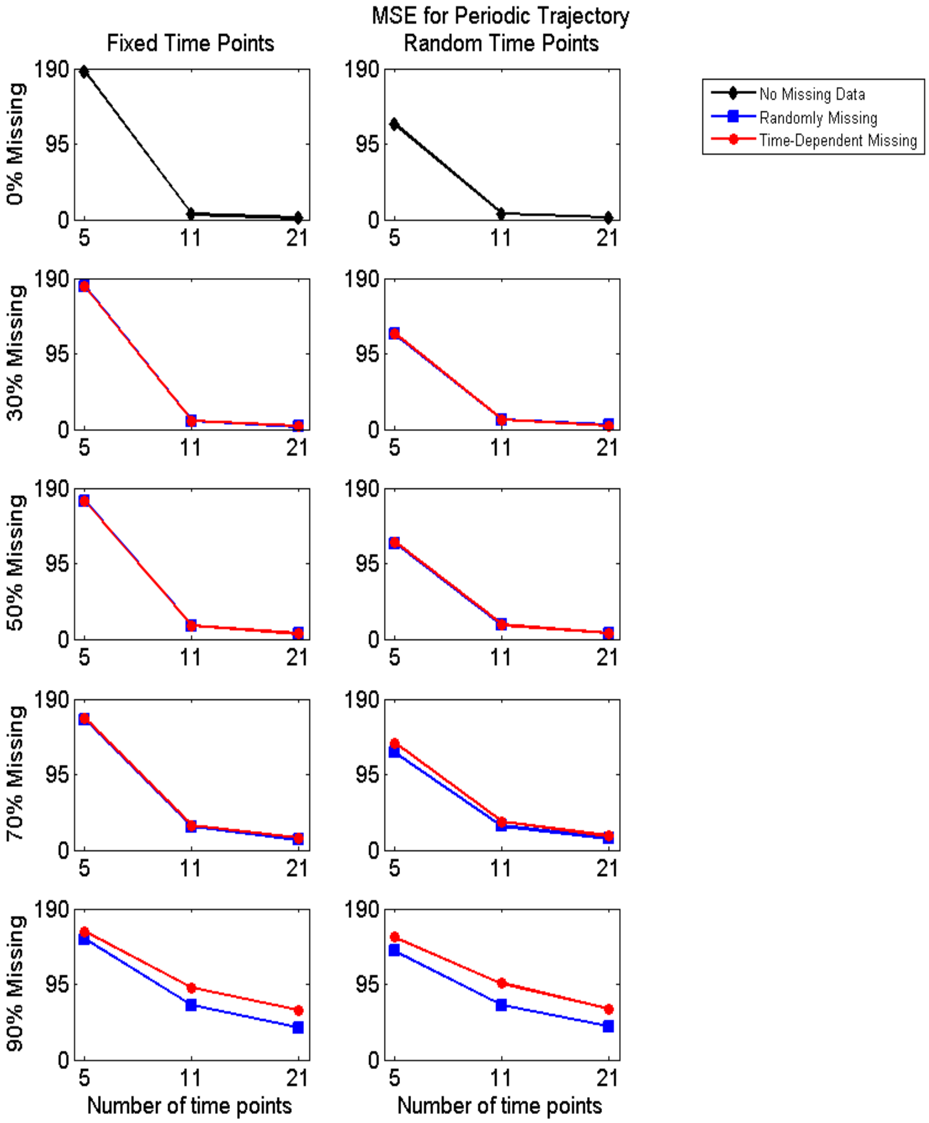

Figure 32. Plots of MSE values for the periodic trajectory across the levels of sparseness, percent missingness, type of missingness, and type of time point.

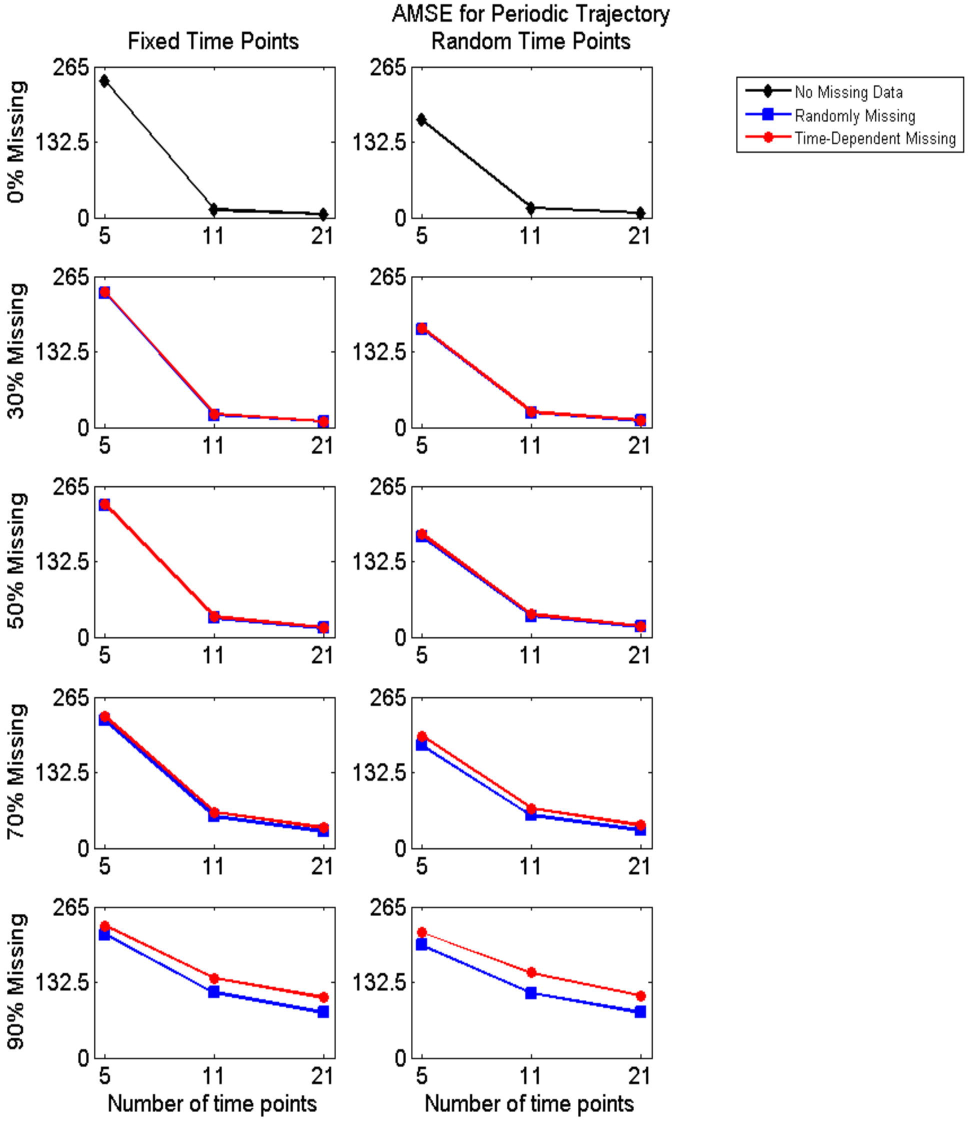

Figure 33. Plots of AMSE values for the periodic trajectory across the levels of sparseness, percent missingness, type of missingness, and type of time point.

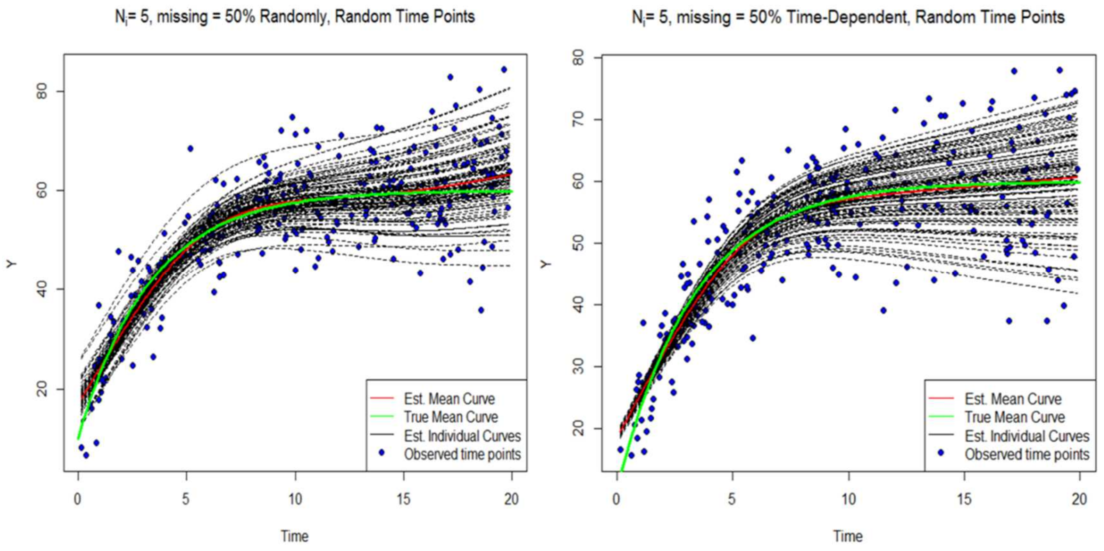

Figure 34. Plots of randomly missing data compared to time-dependent missing data. This shows a case where more data are above the true mean function and thus curve the estimated mean function up at later time points. This shows why time-dependent missingness was only slightly worse than random missingness.

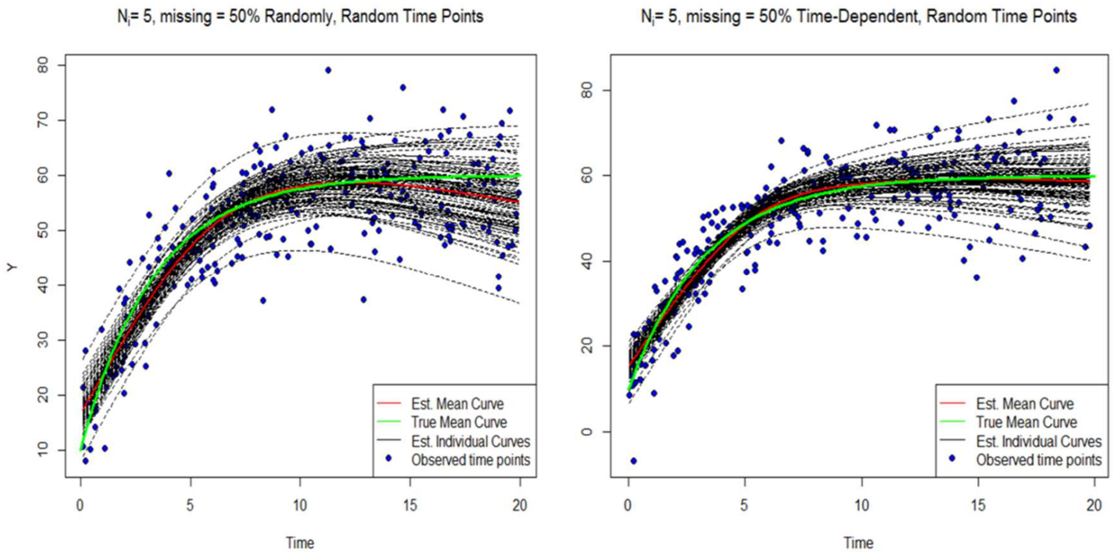

Figure 35. Plots of randomly missing data compared to time-dependent missing data. This shows a case where more data are below the true mean function and thus curve the estimated mean function down at later time points. This shows why time-dependent missingness was only slightly worse than random missingness.

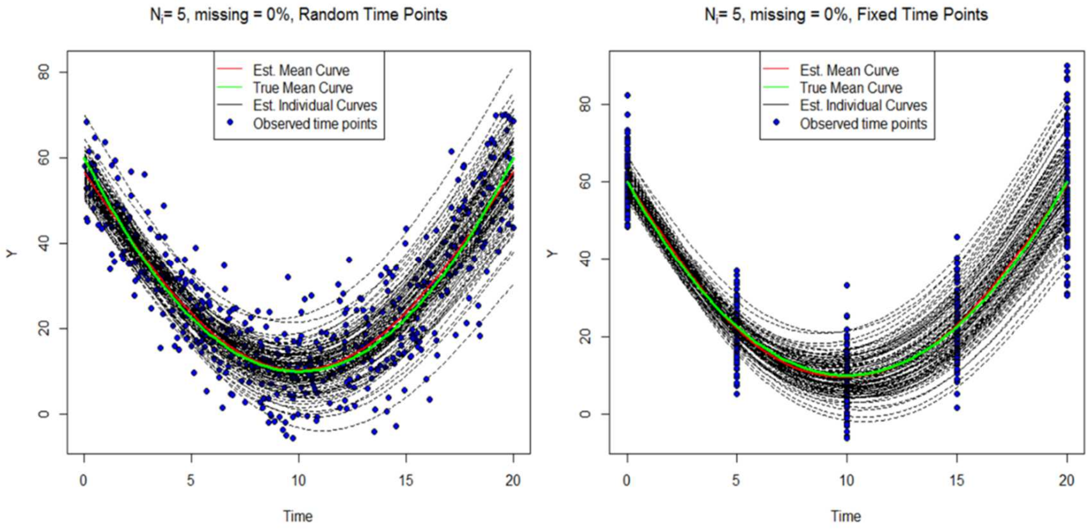

Figure 36. Plots of random time points compared to fixed time points for the quadratic trajectory. This shows the estimated mean curve for the random time points deviating from the true mean curve at the boundaries. The estimated mean curve and true mean curve show no deviation at the boundaries for fixed time points.

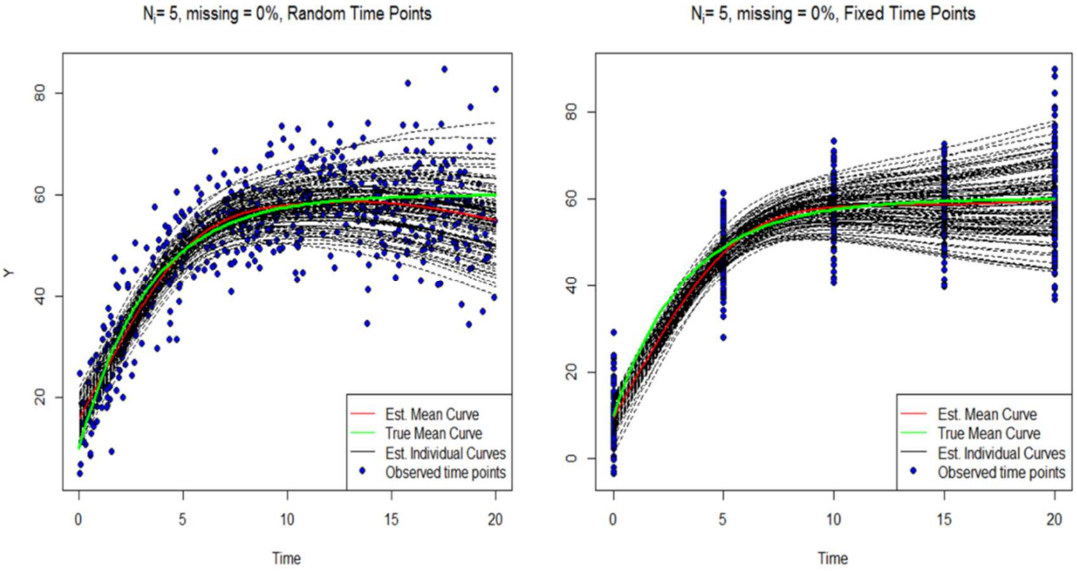

Figure 37. Plots of random time points compared to fixed time points for the asymptotic trajectory. This shows the estimated mean curve for the random time points deviating from the true mean curve at the boundaries. The estimated mean curve and true mean curve show no deviation at the boundaries for fixed time points.

APPENDIX B

TABLES

Table 1

Percent missingness per quintile used for generating time-dependent missing data

Percent	Quintiles				
Missing	1st	2nd	3rd	4th	5th
30%	10%	20%	30%	40%	50%
50%	30%	40%	50%	60%	70%
70%	50%	60%	70%	80%	90%
90%	82%	86%	90%	94%	98%

Table 2

MSE and AMSE values across the no missing data conditions

Type of Time Points	Trajectory Shape	MSE			AMSE		
		N_i=5	N_i=11	N_i=21	N_i=5	N_i=11	N_i=21
Fixed	Quadratic	0.39	0.42	0.31	9.37	5.99	4.11
	Asymptotic	0.82	0.49	0.35	10.01	6.10	4.11
	Periodic	186.47	6.42	2.26	239.34	13.70	6.55
Random	Quadratic	0.90	0.51	0.34	12.42	7.15	4.66
	Asymptotic	1.51	0.67	0.39	13.37	7.42	4.71
	Periodic	119.93	7.70	2.82	171.04	16.96	7.98

Table 3

MSE values across the fixed time point conditions

Percent Missing	Trajectory Shape	Type of Missingness					
		Randomly Missing			Time-Dependent Missing		
		$N_i = 5$	$N_i = 11$	$N_i = 21$	$N_i = 5$	$N_i = 11$	$N_i = 21$
30%	Quadratic	0.51	0.52	0.34	0.63	0.54	0.38
	Asymptotic	1.06	0.612	0.41	1.20	0.63	0.37
	Periodic	180.41	10.83	3.91	179.79	10.88	4.02
50%	Quadratic	0.69	0.61	0.39	0.92	0.75	0.41
	Asymptotic	1.51	0.70	0.46	1.56	0.66	0.43
	Periodic	173.76	16.69	6.80	172.73	16.80	6.74
70%	Quadratic	1.16	0.88	0.50	1.79	1.27	0.74
	Asymptotic	2.17	1.19	0.62	2.32	1.22	0.75
	Periodic	164.17	29.06	13.33	165.91	31.00	15.66
90%	Quadratic	4.03	2.33	1.38	6.77	11.39	6.08
	Asymptotic	5.94	2.80	1.69	7.62	4.61	2.40
	Periodic	152.25	68.44	40.59	161.14	90.51	62.13

Table 4

MSE values across the random time point conditions

Percent Missing	Trajectory Shape	Type of Missingness					
		Randomly Missing			Time-Dependent Missing		
		$N_i = 5$	$N_i = 11$	$N_i = 21$	$N_i = 5$	$N_i = 11$	$N_i = 21$
30%	Quadratic	1.22	0.68	0.39	1.41	0.64	0.39
	Asymptotic	1.87	0.81	0.49	1.76	0.82	0.45
	Periodic	119.46	12.30	4.73	120.10	12.08	4.64
50%	Quadratic	1.69	0.80	0.46	2.08	0.89	0.51
	Asymptotic	2.40	1.14	0.64	2.10	1.04	0.57
	Periodic	120.00	17.68	7.68	122.15	17.96	7.72
70%	Quadratic	2.88	1.21	0.67	5.22	2.12	1.05
	Asymptotic	3.50	1.58	0.89	2.97	1.65	0.92
	Periodic	122.60	29.93	14.48	134.79	35.91	17.74
90%	Quadratic	10.87	3.33	1.95	21.92	9.39	5.18
	Asymptotic	10.41	3.90	2.12	7.16	4.45	2.90
	Periodic	136.79	68.67	41.85	154.36	96.18	63.96

Table 5
AMSE values across the fixed time point conditions

Percent Missing	Trajectory Shape	Type of Missingness					
		Randomly Missing			Time-Dependent Missing		
		$N_i = 5$	$N_i = 11$	$N_i = 21$	$N_i = 5$	$N_i = 11$	$N_i = 21$
30%	Quadratic	13.37	8.17	5.37	14.95	9.20	5.97
	Asymptotic	14.03	8.37	5.48	15.67	9.22	6.01
	Periodic	235.84	21.88	10.03	236.90	23.10	10.68
50%	Quadratic	17.22	10.84	7.01	19.28	12.63	8.22
	Asymptotic	17.71	11.11	7.04	19.95	12.90	8.26
	Periodic	231.02	33.71	15.45	232.85	35.78	16.77
70%	Quadratic	21.23	15.44	10.43	26.35	20.25	14.12
	Asymptotic	22.25	15.69	10.55	26.58	19.83	13.94
	Periodic	223.94	55.36	29.15	230.46	63.14	35.86
90%	Quadratic	29.25	28.29	23.71	36.17	41.72	34.91
	Asymptotic	31.94	28.77	25.21	36.21	33.52	29.96
	Periodic	215.86	114.64	79.00	231.58	139.18	105.27

Table 6

AMSE values across the random time point condition

Percent Missing	Trajectory Shape	Type of Missingness					
		Randomly Missing			Time-Dependent Missing		
		$N_i = 5$	$N_i = 11$	$N_i = 21$	$N_i = 5$	$N_i = 11$	$N_i = 21$
30%	Quadratic	16.11	9.34	5.93	17.35	10.43	6.58
	Asymptotic	16.94	9.64	6.15	17.72	10.65	6.66
	Periodic	172.71	25.79	11.89	174.67	26.98	12.24
50%	Quadratic	19.87	11.96	7.70	23.07	13.86	8.85
	Asymptotic	20.34	12.26	7.73	22.76	14.08	9.07
	Periodic	175.74	36.52	17.55	180.19	39.49	19.08
70%	Quadratic	25.60	17.47	10.96	32.09	24.08	14.80
	Asymptotic	26.19	17.80	11.42	29.89	23.51	14.81
	Periodic	180.08	57.71	31.47	196.33	69.69	39.83
90%	Quadratic	39.09	30.18	26.98	55.75	40.74	35.85
	Asymptotic	40.07	31.04	27.60	37.53	34.57	32.02
	Periodic	198.04	113.20	79.60	220.09	148.94	108.83

Table 7

Pseudo-R^2 values across the no missing data conditions

Type of Time Points	Trajectory Shape	Pseudo-R^2		
		$N_i = 5$	$N_i = 11$	$N_i = 21$
Fixed	Quadratic	0.95	0.93	0.91
	Asymptotic	0.94	0.90	0.88
	Periodic	0.88	0.86	0.88
Random	Quadratic	0.91	0.90	0.90
	Asymptotic	0.88	0.86	0.85
	Periodic	0.47	0.86	0.88

Table 8

Pseudo-R^2 values across the fixed time point condition

Percent Missing	Trajectory Shape	Type of Missingness					
		Randomly Missing			Time-Dependent Missing		
		$N_i = 5$	$N_i = 11$	$N_i = 21$	$N_i = 5$	$N_i = 11$	$N_i = 21$
30%	Quadratic	0.95	0.93	0.91	0.95	0.93	0.91
	Asymptotic	0.94	0.91	0.88	0.95	0.92	0.89
	Periodic	0.88	0.84	0.87	0.86	0.84	0.87
50%	Quadratic	0.95	0.93	0.92	0.95	0.93	0.92
	Asymptotic	0.94	0.91	0.89	0.95	0.92	0.90
	Periodic	0.86	0.81	0.86	0.83	0.81	0.86
70%	Quadratic	0.96	0.94	0.92	0.95	0.93	0.92
	Asymptotic	0.94	0.91	0.89	0.95	0.93	0.91
	Periodic	0.82	0.74	0.82	0.71	0.71	0.82
90%	Quadratic	0.94	0.89	0.89	0.93	0.88	0.88
	Asymptotic	0.91	0.85	0.84	0.93	0.88	0.87
	Periodic	0.68	0.50	0.61	0.57	0.49	0.61

Table 9

Pseudo-R^2 values across the random time point condition

Percent Missing	Trajectory Shape	Type of Missingness					
		Randomly Missing			Time-Dependent Missing		
		$N_i = 5$	$N_i = 11$	$N_i = 21$	$N_i = 5$	$N_i = 11$	$N_i = 21$
30%	Quadratic	0.92	0.91	0.90	0.92	0.91	0.90
	Asymptotic	0.88	0.87	0.86	0.89	0.88	0.87
	Periodic	0.46	0.84	0.87	0.41	0.84	0.87
50%	Quadratic	0.92	0.91	0.90	0.90	0.91	0.90
	Asymptotic	0.88	0.87	0.86	0.88	0.89	0.88
	Periodic	0.43	0.81	0.86	0.37	0.80	0.86
70%	Quadratic	0.90	0.91	0.91	0.88	0.89	0.91
	Asymptotic	0.86	0.88	0.87	0.84	0.86	0.89
	Periodic	0.37	0.74	0.82	0.31	0.71	0.82
90%	Quadratic	0.85	0.84	0.85	0.85	0.83	0.83
	Asymptotic	0.74	0.77	0.78	0.81	0.80	0.81
	Periodic	0.30	0.51	0.62	0.28	0.50	0.62

Table 10
Signal-to-noise ratio (SNR) values

Trajectory Shape	SNR		
	$N_i = 5$	$N_i = 11$	$N_i = 21$
Quadratic	15.81	11.51	9.97
Asymptotic	13.27	8.58	7.07
Periodic	7.31	7.64	7.60