Journal of Applied Statistics

ISSN: 0266-4763 (Print) 1360-0532 (Online) Journal homepage: http://www.tandfonline.com/loi/cjas20

Applying functional data analysis to assess tele-interpersonal psychotherapy's efficacy to reduce depression

Henok Woldu, Timothy G. Heckman, Andreas Handel & Ye Shen

To cite this article: Henok Woldu, Timothy G. Heckman, Andreas Handel & Ye Shen (2018): Applying functional data analysis to assess tele-interpersonal psychotherapy's efficacy to reduce depression, Journal of Applied Statistics, DOI: 10.1080/02664763.2018.1470231

To link to this article: https://doi.org/10.1080/02664763.2018.1470231

JOURNAL OF APPLIED STATISTICS, 2018
https://doi.org/10.1080/02664763.2018.1470231

Taylor & Francis
Taylor & Francis Group

Applying functional data analysis to assess tele-interpersonal psychotherapy's efficacy to reduce depression

Henok Woldu[a], Timothy G. Heckman[b], Andreas Handel[a] and Ye Shen[a]

[a]Department of Epidemiology and Biostatistics, College of Public Health, University of Georgia, Athens, GA, USA; [b]Department of Health Promotion and Behavior, College of Public Health, University of Georgia, Athens, GA, USA

ABSTRACT

The use of parametric linear mixed models and generalized linear mixed models to analyze longitudinal data collected during randomized control trials (RCT) is conventional. The application of these methods, however, is restricted due to various assumptions required by these models. When the number of observations per subject is sufficiently large, and individual trajectories are noisy, functional data analysis (FDA) methods serve as an alternative to parametric longitudinal data analysis techniques. However, the use of FDA in RCTs is rare. In this paper, the effectiveness of FDA and linear mixed models (LMMs) was compared by analyzing data from rural persons living with HIV and comorbid depression enrolled in a depression treatment randomized clinical trial. Interactive voice response systems were used for weekly administrations of the 10-item Self-Administered Depression Scale (SADS) over 41 weeks. Functional principal component analysis and functional regression analysis methods detected a statistically significant difference in SADS between telphone-administered interpersonal psychotherapy (tele-IPT) and controls but linear mixed effects model results did not. Additional simulation studies were conducted to compare FDA and LMMs under a different nonlinear trajectory assumption. In this clinical trial with sufficient per subject measured outcomes and individual trajectories that are noisy and nonlinear, we found FDA methods to be a better alternative to LMMs.

ARTICLE HISTORY
Received 24 December 2016
Accepted 20 April 2018

KEYWORDS
IVR; generalized cross-validation; functional *f*-test; functional regression; tele-IPT; SADS

1. Introduction

The distinguishing feature of longitudinal data is that measurements are taken repeatedly from each subject over time and are, therefore, correlated [4,7,9]. Because of these intercorrelations, the analysis of longitudinal data requires that statistical methods account for correlations among the data [14,32]. Univariate and multivariate repeated measures analysis of variance (ANOVA) are two traditional methods of longitudinal data analysis and are most appropriate when the number of observations per individual is small [6,21]. Both methods are popular due to their simplicity and similarity to ANOVA. However, they are

CONTACT Ye Shen yeshen@uga.edu

Supplemental data for this article can be accessed here. https://doi.org/10.1080/02664763.2018.1470231

less applicable when the assumptions they impose are not satisfied [14]. For example, these methods often assume that measurements are taken for all subjects at the same time and that there are no missing observations, which are of limited flexibility in practice [20]. More recently, parametric linear mixed models (LMM), which are often better alternatives to the univariate and multivariate repeated measure analysis techniques, have found widespread use.

LMM, widely used in biomedical research, is a statistical model containing both random and fixed effects [4,7]. LMM offers many advantages over traditional univariate and multivariate repeated measures analysis. In LMM analysis, time of observation can differ across subjects, the covariance structure is more flexible, and the issue of missing data can be addressed. This method, however, assumes linear relationships between the response and covariates, normality of the random terms, and independence of errors [6,14,32]. In practice, however, it is often the case that the mean trend over time is too complex to be described by models with simple linear or lower order polynomial forms. Furthermore, it is common to encounter data that are contaminated when observed at each time point to become less noisy when smoothed [11]. When all or some of the LMM assumptions are violated, nonlinear modeling [3], semiparametric and nonparametric modeling approaches may be used [17,37,38]. Taking the nonparametric approach of modeling to longitudinal data a step further, functional data analysis (FDA), a nonparametric method that involves smoothing, may be an effective alternative.

FDA is a technique comprising a group of statistical methods that handle data as functions rather than single values [22,26,27]. Typically, these are functions of time but can also be space or other domains [15]. The methods assess values that are theoretically continuous functions but that may have been measured at many discrete points [11,16]. In spite of the many possibilities arising with functional data in randomized control trials (RCTs), the use of FDA has been relatively rare. One main reason for this was the unavailability of software packages to implement it. However, the fda [24], fda.usc [5], and other packages in R and similar function in Matlab [24] have greatly extended the application of FDA in the last two decades.

In this paper, we analyzed data from rural persons living with HIV infection and comorbid depression diagnoses who were enrolled in a randomized clinical trial assessing the efficacy of telephone-administered interpersonal psychotherapy to reduce depression in this group. Depression scores, obtained from participants through weekly interactive voice response (IVR) systems, though recorded as discrete values, deserve to be labeled as functional data in that they reflect a smooth variation in one's depression. We analyzed the data using FDA techniques and LMM and compared the effectiveness of the two statistical analysis approaches. Functional principal component analysis (FPCA) and functional regression, two FDA techniques, were used to analyze the data collected from participants enrolled in the clinical trial. Additional simulation studies were conducted to compare FDA and LMMs when frequent measures of outcome with noisy and nonlinear individual trajectories were obtained from a subject. Some of the results in this paper have previously been reported in the form of abstract at the 50th annual meeting of the Society for Epidemiologic Research (SER) 2017.

The remainder of this paper is organized as follows. In Section 2, a description of the data collected as part of the RCT is provided. Notations and the FDA methods used are described in Section 3. In Section 4, FDA is applied to tele-IPT RCT study data and

findings are presented. Simulation results are reported in Section 5. The paper concludes in Section 6 with a discussion of study findings and future implications.

2. Sampling and data collection

Between August 2010 and September 2014, AIDS service organizations (ASOs) in 28 states distributed recruitment brochures to their rural clients living with HIV through face-to-face interactions, regular mail, and placement of study brochures in high-traffic areas of their facilities. The Rural Center for AIDS Prevention (RCAP) at Indiana University also distributed study advertisements through its listserv and publicized the study on its website. Patients enrolled into the RCT were from the states of Alabama, Arkansas, California, Colorado, Florida, Georgia, Illinois, Indiana, Kansas, Kentucky, Louisiana, Maine, Minnesota, Mississippi, Missouri, Montana, New Hampshire, New York, North Carolina, Ohio, Oklahoma, Oregon, Pennsylvania, South Carolina, Tennessee, Texas, Vermont, and Virginia. The IRBs of all participating institutions approved the projects protocol and all patients provided written informed consent. No adverse events were reported during the trial.

Potential participants contacted the research office via a toll-free telephone number or a project-specific e-mail address listed in recruitment materials. During this contact, research staff provided detailed information about the study, described the informed consent process, and gathered preliminary screening information, such as current age, county of residence, and contact information. County of residence was used to determine the participants U.S. Department of Agricultures Urban–Rural Continuum [31]. If the individual satisfied the rurality inclusion criterion and expressed interest in the study, study personnel mailed (or e-mailed) an informed consent form to the individual to be signed and returned to the research office. After receiving a signed consent document, research staff contacted the individual and conducted an eligibility screening, which generally took less than 30 min to complete. Telephone-based eligibility interviews administered the Primary Care Evaluation of Mental Disorders (PRIME-MD) [28] and the Modified Mini Mental State Examination (3MS) [29].

Inclusion and exclusion criteria. Inclusion criteria were: (1) 18 years of age; (2) self-reported diagnosis of HIV infection or AIDS; (3) residing in a county with U.S. Department of Agriculture Rural-Urban Continuum Code of 4–9; (4) diagnosis of DSM-IV Major Depressive Disorder (MDD), MDD in Partial Remission, or Dysthymic Disorder based on the Mood Module of the PRIME-MD; (5) the patient intended to stay in his or her current residence for 1 year; and (6) written informed consent. The sole exclusion criterion was serious cognitive or neuropsychiatric impairment based on the telephone-administered version of the 3MS (scores < 70). Individuals were not excluded on the basis of alcohol or substance use disorders, active bipolar disorder, psychotic symptoms, or current receipt of psychotherapy or pharmacotherapy. The minimization of exclusion criteria was used to assemble a sample high in external validity.

Of the 161 patients who satisfied inclusion and exclusion criteria, 81 were randomly assigned to the experimental treatment condition and 80 to standard care. Analyses in this paper used a completer-only approach ($N = 115$), 56 patients in the treatment condition and 59 standard care controls. Patients assigned to the treatment condition (tele-IPT) received nine weekly sessions of telephone-administered interpersonal psychotherapy

while SC controls received no active treatment but had access to community-based support services available to people living with HIV. The 10-item Self-Administered Depression Scare (SADS) was administered via IVR weekly for 41 weeks (during 9 weeks of intervention and 32 weeks of follow-up). SC controls completed weekly IVR-based SADS with their time-matched treatment counterparts. A more detailed description of the clinical trial's procedures and intervention-outcome findings for the primary outcome were published somewhere else [12,13].

The current study served as an excellent opportunity to assess the effectiveness of FDA to analyze SADS obtained from IVR for three reasons: (1) the frequency of measurements taken from each individual (41 measurements) was sufficiently dense for FDA; (2) because the intervention occurred over 9 weeks with 32 weeks of follow-up assessments, the rate of change (if any) was not expected to be constant across the 41-week assessment period (this complex relationship could be quantified through FDA as opposed to LMM; and (3) SADS are subjective and perhaps contain error when self-reported every week. Applying smoothing to depression scores was expected to reduce noisy individual trajectories, thereby resulting in less biased estimators.

3. Methods

3.1. Converting observed SADS scores from each patient to curves

In FDA, the first step is to predict individual trajectories using the subjects' available measurements. Once a curve is fitted to each subject, treating each curve as a sample unit, different analyses can be performed on the curves [1,23,34,36]. In our working data set, let y_{ij} represent the SADS of the ith HIV positive patient at the jth time point. Where $i = 1, 2, \ldots, N = 115$ be the number of patients in the study and $j = 1, 2, \ldots T = 41$ be observation times. Generally, the number and time of measurements taken could be different between subjects [26]. In the depression scores obtained through IVR, however, both frequency and time of measurements are the same for all individuals. In working with functional data, it is commonly assumed that y_{ij} is a noisy observation of $x_i(t_j)$ which is written as follows:

$$y_{ij} = x_i(t_j) + \varepsilon_{ij} \quad \text{or in matrix form } \mathbf{y} = x(\mathbf{t}) + \mathbf{e},$$

where ε_{ij} are the random noise commonly assumed to be independent and identically distributed as $N(0, \sigma^2)$ and all $\mathbf{y}$, $x(\mathbf{t})$, $\mathbf{t}$, and $\mathbf{e}$ are column vectors of length T [22]. Furthermore, the term $x_i(t_j)$ is assumed to be a smooth function and is estimated by the basis function expansion approach using the following equation:

$$\hat{x}_i(t) = \sum_{k=1}^{K} \hat{c}_{ik}\phi_k(t) \quad \text{in matrix form } \mathbf{x} = \mathbf{c}'\phi,$$

where $\hat{c}_{ik}$ denotes the coefficients of the expansion and ϕ_k is the chosen basis system consisting of K number of basic functions [1,15]. $\mathbf{c}$ is a vector of length K of the coefficients $\hat{c}_{ik}$ and ϕ is an n by K matrix whose elements are the basis functions $\phi_k(t)$. The underlying nature of the data is the main factor for deciding which basis function to use. Fourier basis are used for periodic data, while the spline basis for non-periodic data [8,26]. Because the

SADS measurements taken from patients in the clinical trial are non-period in nature, Cubic Bspline (spline of order four), predominantly used in applications, was used to convert each participant's depression score to functional smooth curves.

Following the choice of the basis function, there are two main approaches to convert data into smooth functional curves; the *regression smoothing* and *roughness penalty smoothing* [26,27,34]. The regression smoothing approach minimizes the *sum of squared errors (SSE)* where

$$SSE = \sum_{j=1}^{T} [y_{ij} - \hat{x}_i(t_j)]^2 \quad \text{or in matrix form } SSE = (\mathbf{y} - \Phi)'(\mathbf{y} - \Phi).$$

Differentiating with respect to $\mathbf{c}$ will result in the estimates of $\hat{\mathbf{c}} = (\Phi'\Phi)^{-1}\Phi'\mathbf{y}$. The vector for the fitted data is then estimated to be $\hat{\mathbf{y}} = \Phi(\Phi'\Phi)^{-1}\Phi'\mathbf{y}$. The regression smoothing method, however, does not allow one to control the level of smoothness and results in over-fitting for a large choice of K [22,23,27]. The roughness penalty smoothing approach which adds a penalty roughness for increasing values of K, however, allows one to overcome the problem of over-fitting. The roughness penalty in FDA aims to capture the notable characteristics of the data by using a powerful basis expansion, but avoids over-fitting the data by a penalty on the *roughness* of the function [34]. Roughness commonly used in practice is defined as follows:

$$PEN_{m+2}[x] = \int [D^{m+2}x(t)]^2 \, \mathrm{d}t,$$

where $D^{m+2}x(t)$ represent the $(m+2)$th derivative of $x(t)$. For our data set we worked with the most commonly practiced, second derivative squared (setting $m = 0$), where the *penalizedsumofsquares*(*PENSSE*) is given as follows [23]:

$$PENSSE_\lambda(x) = \sum_{i=1}^{N}\sum_{j=1}^{T}(y_{ij} - x_i(t_j))^2 + \lambda PEN_2[x].$$

The above expression can be written in matrix forms as follows: $PENSSE_\lambda(x) = (\mathbf{y} - \Phi\mathbf{c})'(\mathbf{y} - \Phi\mathbf{c}) + \lambda\mathbf{c}'\mathbf{R}\mathbf{c}$, where $\mathbf{R}$ which is an order K roughness penalty matrix is given as follows:

$$\mathbf{R} = \int [D^{m+2}\phi(t)][D^{m+2}\phi'(t)] \, \mathrm{d}t.$$

The estimated coefficient $\hat{\mathbf{c}}$ is then expressed as $\hat{\mathbf{c}} = (\Phi'\Phi + \lambda R)^{-1}\Phi'\mathbf{y}$. The smoothing parameter λ specifies the emphasis on the second term, penalizing curvature relative to goodness of fit in the sum of squared residuals in the first term. Large values of λ increasingly penalize roughness and $x(t)$ tend to become more linear. A value of λ closer to 0 does not penalize roughness and leads to over-fitting. The optimal value of λ can be obtained applying a data-driven approach called *generalizedcross − validation*(*GCV*)

[2,10,22], where GCV is given as follows:

$$\mathrm{GCV}(\lambda) = \left\{ \left(\frac{N}{N - df(\lambda)} \right) \left(\frac{SSE}{N - df(\lambda)} \right) \right\},$$

where $df(\lambda) = \mathrm{trace}\{(\Phi'\Phi + \lambda R)^{-1}\Phi' y\}$ and SSE is given as defined above. Choosing the minimum value obtained by GCV, individual discrete measurements of SADS from each subject were converted to continuous, smoothed SADS curves. The 56 patients in the treatment arm each had 41 weekly measurements, and formed 56 SADS curves for the tele-IPT group. Similarly, 59 patients with 41 weekly measurements formed 59 SADS curves for the control group.

3.2. Functional principal component analysis

Functional principal component analysis (FPCA), with a similar idea to the principal component analysis (PCA) in the multivariate cases, is used to explore the temporal/spatial variation in the fitted curves. What makes FPCA different from PCA is that the weights and data are functions not vectors. Let $x_i(t)$ be represented as a linear combination of $\sum_{k=1}^{K} c_{ik}\xi_k(t)$ with an optimal choice of K and the mean removed. Furthermore, let us assume the first most important m number of principal components is represented by the m-vector $\xi(t$ weight functions. The principal component score that corresponds to $\xi(t)$ for the ith curves is then represented as $f_i = \int \xi(t)x_i(t)\,\mathrm{d}t$. To find the first principal score defined as $f_{i1} = \int \xi(t)x_i(t)\,\mathrm{d}t$, the first weight function $\xi_1(t)$ that defines the largest component of variation is calculated so as to maximize $N^{-1}\sum_i f_{i1}^2(t)$ subject to $\|\xi_1\|^2 = \int \xi_1^2(t)\,\mathrm{d}t = 1$ constraint. The second principal score $f_{i2} = \int \xi(t)x_i(t)\,\mathrm{d}t$ is calculated similarly such that $\xi_2(t)$ which defines the second largest component variation is calculated so as to maximize $N^{-1}\sum_i f_{i2}^2(t)$ subject to $||\xi_2||^2 = \int \xi_2^2(t)\,\mathrm{d}t = 1$ constraint. This process is repeated until m, the first most important principal components, are achieved [22,23,27]. Thus, using FPCA we were able to extract the main difference in the curve trajectories between tele-IPT and control groups for our data.

3.3. Functional regression with functional response and functional F-Test

Functional regression is no different than ordinary regression, except that either the response or the feature variables, or both, have to be functional variable(s). The focus of this paper is in functional regression with functional response (SADS) and scalar covariate (binary covariate for the treatment assignment). To answer the main objective of the IVR HIV RCT study, if tele-IPT intervention helped to reduce the depression in HIV positive rural patients with comorbid depression, the following functional linear regression model was fitted [34]:

$$y_i(t) = B_0(t) + \sum_{j=1}^{2} x_{ij}B_j(t) + \epsilon_i(t),$$

where $y_i(t)$ is a functional response and x_{ij} is 1 if the ith individual is from tele-IPT and 0 if from the control group. $B(t)$ are the coefficients associated with group j at time t. Because we have a functional response, the estimated coefficients obtained are also functions,

i.e. functions of time. Finally to test $H_0 : \mu(t)_{\text{Tele-IPT}} = \mu(t)_{\text{control}}$ Vs $H_1{:}\mu(t)_{\text{Tele-IPT}} \neq \mu(t)_{\text{control}}$, where $\mu(t)_{\text{Tele-IPT}}$ and $\mu(t)_{\text{control}}$ represent the mean of tele-IPT and control groups at time *t*, *functional F-test*, which is an extension of the *classical F-test* was performed. This procedure uses a point-wise critical value of 0.05 obtained using permutation test and a more conservative maximum critical value of 0.05 as reference lines [35,36]. The *F*-test statistic (*F*-observed) at each time point is then compared against the critical values on the reference line. When the *F*-test statistic value exceeds the reference line, this indicates the difference between the two groups is statistically significant.

4. Data application

A sample of fitted curves for the raw data and the smoothed curves for both tele-IPT patients and control is presented in Figure 1. The vertical line in the tele-IPT curves represents the 9th week of the study period. This corresponds to the termination of the 9-week intervention. Figure 1 shows that there was a gradual decrease in depression scores during the intervention in tele-IPT patients; however, among control subjects, no clear pattern of change emerged.

Results of the FPCA, which show the main temporal variation between the fitted curves of the tele-IPT and control groups, are shown in Figure 2. The first functional principal component (FPC) for the tele-IPT group (Figure 2-left) explains more than 80% of the variation between the fitted curves. From the first FPC of the tele-IPT group, we see that overall SADS scores decreased during the first 9 weeks (intervention period), followed by a gradual decline to Week 20 that was maintained for the remainder of the follow-up period. For control group, the first FPC (Figure 2-right) explains more than 90% of the variation between the fitted curves. It can be seen from the first FPC of the control group that, overall, the trend showed only a minor change from the baseline SADS values through follow-up.

Results of the functional linear regression model are presented in Figure 3. Taking the functional nature of the coefficients into consideration, it is desirable to show the graph of the coefficients as a function of time for each group separately. The intercept plot (*a*) corresponds to the mean SADS for the control group at a given time *t*. The coefficients associated with tele-IPT (*b*) are all negative. This indicates that on average, SADS scores were lower among tele-IPT than the control group throughout the study period. Furthermore, it can be observed that average depression scores were declining during the 9-week intervention period for tele-IPT subjects and that the decline mitigated across follow-up. Conversely, for controls, it can be seen from the plot of the intercept that, with the exception of the period from Week 28 to Week 35, SADS scores remained largely unchanged.

To test our hypothesis that there was no difference in the mean SADS between the two groups, we performed functional *F*-test. As shown in Figure 3(d), the observed *F*-statistic (solid line) is higher than the point-wise 0.05 critical value (bottom dotted line), and the most conservative maximum 0.05 critical value (dotted horizontal line) during the first 11 weeks. This indicates statistically significant group difference in SADS scale scores during this period. Similar differences were observed in Weeks 19–21. These results match those observed from the FPCA and the functional regression fit.

The predicted mean SADS curves (Figure 3(c)) for the tele-IPT and the control were further investigated. The two curves never crossed each other throughout the study period.

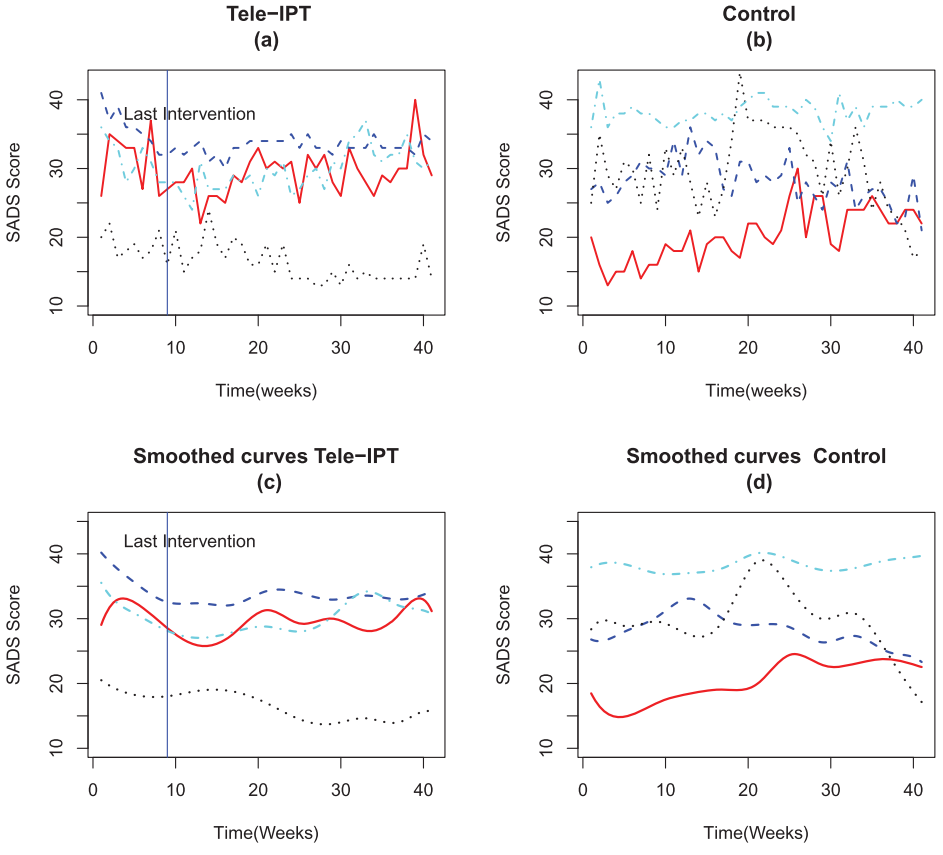

Figure 1. (a) and (b) show a sample of four patients curves for tele-IPT and control groups respectively from the raw data, while (c) and (d) show their curves when smoothed. The vertical line at Week 9 represents the end of the 9-week tele-IPT intervention.

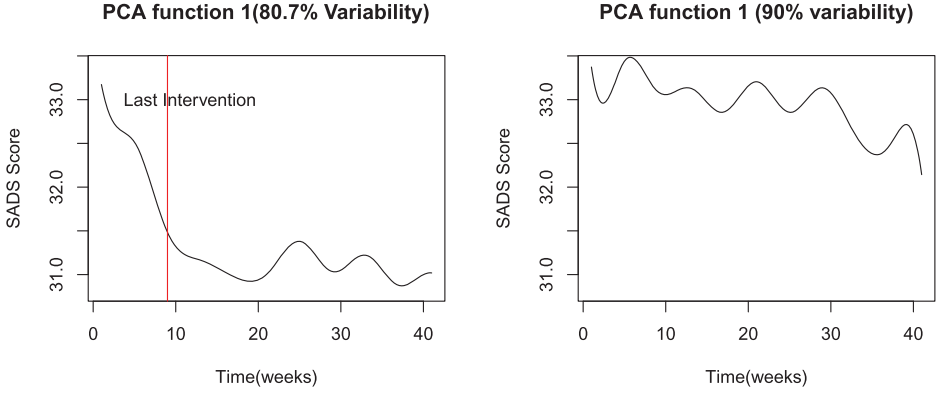

Figure 2. Left plot shows the first functional principal component (FFPC) which explains more than 80% of the variability among the tele-IPT; and the right plot is the FFPC which explains 90% of the variability among the control group.

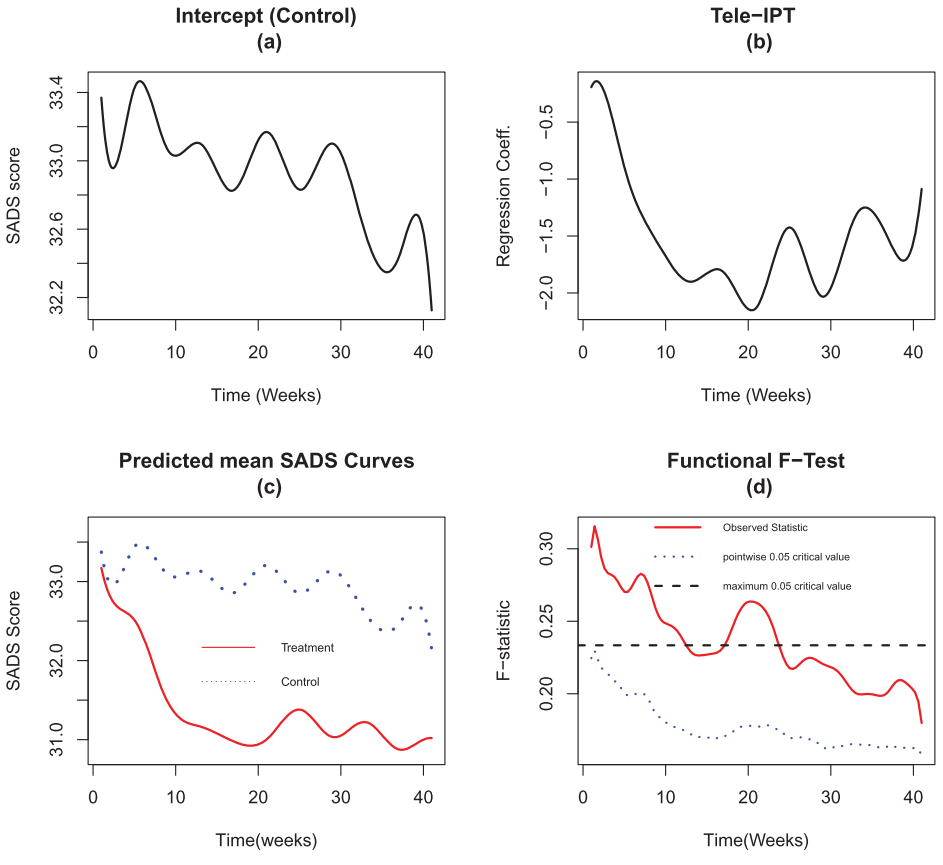

Figure 3. Results of the functional regression coefficients fit for the control (a) and Tele-IPT (b). Predicted mean depression score curves for tele-IPT and control from the functional regression fitted model are also shown in (c). Permutation *F*-Test (d) for testing the difference of depression score between tele-IPT and Control arms at each time point. The dashed horizontal line gives the permutation 0.05 critical value for the maximum of the *F*-statistic and the dotted the permutation critical value of the point-wise statistics.

Furthermore, the mean SADS score was observed to be in a steady decline during the first 9 weeks (intervention period) followed by a slower decrease up to Week 20 for the tele-IPT. For the control arm, the mean SADS score showed no discernible changes, except during Weeks 28–35.

A linear mixed effects model was fitted to the same data to assess the effectiveness of tele-IPT to decrease depression in depressed HIV-infected rural persons. The overall difference between the two groups was not significant ($f = 1.09, df = 1, 4598, p = .2973$). The rate of SADS change, Treatment Arm X Time interaction, was also insignificant, although it approached significance ($f = 3.63, df = 1, 4598, p = .0567$). Although the results contradict those of the FDA, they were not unexpected. The noisy individual trajectories and the mean trends for the two groups were complex to be captured by any lower order polynomial relationships. Furthermore, because the intervention accounted for only the first 9 weeks of the assessment period for tele-IPT patients, one would not expect a steady rate of change, if any, throughout the study period.

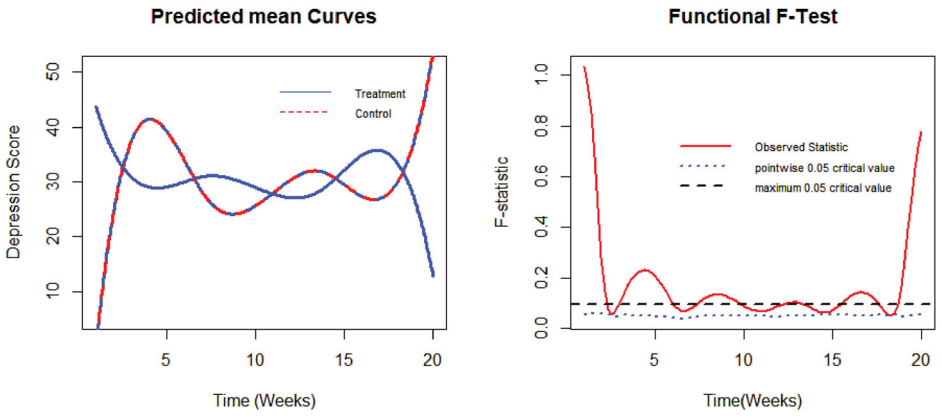

Figure 4. The predicted mean curves for treatment and control are shown in (a). The functional f-test for the mean difference between the two groups at each time point is shown in (b). The results in this figure are obtained from a randomly selected one data set of the 500 data sets generated for simulation studies setting $t = 20$.

5. A simulated illustration

In reality, it is common to encounter data in which the relationship between two variables cannot be explained linearly [18,19,25,30]. The IVR data collected as part of this clinical trial are no exception. For example, McGrath *et al.* [18] used cosine function to model the seasonal change in the birth weight and length of Danish neonates using a sample of data collected from 1973 to 2003. Fetal heart rate tracing record within 1 hour prior to delivery was also observed to have a sinusoidal like pattern in Modanlou *et al.* [19]. Assuming linearity and ignoring the presence of noise in the actual data may lead one to conclude there is no treatment effect when, in reality, there is. To test whether FDA methods provide more in-depth insights than LMM in identifying the difference between groups observed over time when the underlying assumptions of LMM are violated, a simulation study was conducted. Two different data sets were generated using nonlinear functions, one representing a hypothetical treatment group and another control group. The function used to generate data from the treatment group is: (a) $y_{ij} = 30 - 36 * \sin(0.5 * \pi * (x_i(t_j) - \frac{1}{4})) + v_{ij}$, where v_{ij} is $\mathcal{N}(\mu_1, \sigma_1^2)$, $i = 1, \ldots, n$, represent the number of subjects and $j = 1, \ldots, t$ represent the observation time. For the control group the function used is: (b) $y_{ij} = 30 - 36 * \cos(0.8 * \pi * (x_i(t_j) - \frac{1}{4})) + v_{ij}$, where v_{ij} is $\mathcal{N}(\mu_2, \sigma_2^2)$. For a given t, 500 data sets were generated using the functions stated above from each treatment group assuming measurements were taken at the same time for all subjects.

To assess the mean difference across time between the two groups, functional regression followed by function F-test was performed. Results from one data set selected randomly are shown in Figure 4. As seen in Figure 4(a), the predicted mean curves for the two groups were quite different for most of the time period. These differences were detected in the functional F-test (b) where the observed F-statistic (solid line) was above the dashed (most conservative) and dotted reference lines.

LMM which assume linearity and absence of noise in the data was fitted to the simulated data with the same objective of assessing the mean differences between the two groups

Table 1. Evaluating the performance of LMM and FDA by comparing the percentage of false positive (Type I error).

Number of observation time per subject	Linear mixed model (LMM)	Functional data analysis (FDA)
$t = 30$	3.0%	1.9%
$t = 40$	4.2%	1.1%
$t = 50$	5.2%	4.4%
$t = 30$	2.8%	1.1%
$t = 40$	4.0%	1.0%
$t = 50$	6.4%	4.9%

Note: The results are obtained from 500 simulations for each given t and assuming the error terms for the treatment and control respectively have $\mathcal{N}(0, 8)$ and $\mathcal{N}(0, 10)$ for the first set of simulation and $\mathcal{N}(0, 9)$ and $\mathcal{N}(0, 9)$ for the second set of simulation.

across the time interval. Summary of the results from all the 500 generated data sets after fitting FDA and LMM models are presented as supplementary material in Table S1. From Table S1, one can see that, FDA methods were able to detect the group mean difference with high power when the noise added is relatively small regardless of the number of observations taken per subject. Conversely, LMM did a poor job in detecting the group mean differences. Taking the very nonlinear nature of the functions we used for this simulation study, however, this is not far from expected. When larger noise is added to the data, one can see even the FDA methods to have less power for detecting the group mean differences. The power also gets smaller when t gets larger.

Furthermore, to assess the empirical false positive rate of the two methods, using linear function $y_{ij} = 3 + 0.3 * (x_i(t_j)) + v_{ij}$ where v_{ij} is $\mathcal{N}(0, \sigma^2)$, $i = 1, \ldots, n$ represent number of subjects and $j = 1, \ldots, t$ observation times per subject, we first generated 500 data sets from each group. Setting the number of observation per subject to $t = 30, 40$, and 50, we assessed the percentage of false positive under each method. Summary of the simulation results is shown in Table 1. Generally false positive rate range from 2.8% to 6.4% for the LMM and from 1.0% to 4.9% for FDA.

6. Discussion

Technological advancements have transformed intervention science and the treatments used to prevent and treat numerous diseases. New approaches to analyzing longitudinal data sets collected in the course of longitudinal research are also emerging. Perhaps due to convention and investigators familiarity with these approaches, many clinical trials continue to use data analytic approaches (e.g. repeated measures ANOVA or MANOVA and LMM) even though one or more key assumptions are not satisfied [20,22,38]. The application of these inappropriate statistical techniques could yield inaccurate findings which, in turn, may lead to the development of misguided policies. In clinical trials, FDA, a relatively novel technique that handles longitudinal data and reduces the noises through smoothing, can be used as an alternative to LMM in analyzing RCT data [33].

In this paper, we implemented FDA methods to analyze data from an AIDS mental health clinical trial testing and compared findings from these analyses to those from LMM. Using a roughness penalty smoothing method, we first converted the noisy raw curves to smooth curves. Subsequently, using FPCA, we showed that the dominant pattern of outcome changes within and between the two groups. Among the tele-IPT patients, the

reduction in depressive symptoms was clearly depicted in the FFPC. This finding was corroborated by the functional regression and functional *F*-test results. As stated previously, intervention among the treatment group was only during the first 9 weeks followed by 32 follow-up weeks. All the FDA results were able to capture the decrease observed during these 9 weeks followed by a period of little change. Conversely, LMM findings were statistically insignificant. Our simulation study further demonstrated the superior performance of FDA methods relative to LMM in detecting the mean difference between the two groups when individual trajectories were noisy and the pattern of the relationship was nonlinear.

FDA methods may help to uncover the hidden relationship in the data in situations were LMM cannot. When analyzing longitudinal data from clinical trial, however, care should be taken not to overestimate the effect of the intervention. FDA methods, though they reduce bias by trying to reduce the noise in the data, also introduced some bias as a result of smoothing. In clinical trials, it is customary to corroborate the statistical significance with methods that assess the clinical significance, such as Cohens D. Our analyses of longitudinal data using FDA only revealed the statistical significance of group differences but did not address the important issue of clinical or practice significance. However, from our observation, the mean difference in patients' depressive symptoms seem relatively small and of questionable clinical significance. One reason for this might be, the very brief duration of the intervention, given the many medical, psychological, and social difficulties challenging study patients.

As a summary, in this clinical trial study, we found FDA methods to provide a better understanding of the repeated measures outcome compared to LMMs. Functional data analysis methods appear to be a suitable alternative to linear mixed models when analyzing longitudinal data collected as part of clinical trials, particularly when individual trajectories are noisy and nonlinear.

Acknowledgments

The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

Research reported in this publication was supported by the National Institute Of Mental Health of the National Institutes of Health under Award Number R01MH087462.

References

[1] A. Baddeley, P. Gregori, J. Mateu, R. Stoica, and D. Stoyan, *Case Studies in Spatial Point Process Modeling*, Springer, New York, 2006.
[2] P. Craven and G. Wahba, *Smoothing noisy data with spline functions*, Numer. Math. 31 (1978), pp. 377–403.
[3] M. Davidian and D.M. Giltinan, *Nonlinear Models for Repeated Measurement Data*, Vol. 62, CRC Press, Boca Raton, FL, 1995.

[4] L.J. Edwards, *Modern statistical techniques for the analysis of longitudinal data in biomedical research*, Pediatr. Pulmonol. 30 (2000), pp. 330–344.
[5] M. Febrero-Bande and M. Oviedo de la Fuente, *Statistical computing in functional data analysis: The R package fda.usc*, J. Stat. Softw. 51 (2012), pp. 1–28.
[6] G.M. Fitzmaurice, N.M. Laird, and J.H. Ware, *Applied Longitudinal Analysis*, Vol. 998, Wiley, Hoboken, NJ, 2012.
[7] G.M. Fitzmaurice and C. Ravichandran, *A primer in longitudinal data analysis*, Circulation 118 (2008), pp. 2005–2010.
[8] Functional Data Analysis, *Functional data analysis-welcome*. Available at http://www.psych.mcgill.ca/misc/fda/, accessed October 10, 2016.
[9] R.D. Gibbons, D. Hedeker, and S. DuToit, *Advances in analysis of longitudinal data*, Annu. Rev. Clin. Psychol. 6 (2010), pp. 79–107.
[10] C. Gu, *Cross-validating non-gaussian data*, J. Comput. Graph. Stat. 1 (1992), pp. 169–179.
[11] P. Hall, H.G. Müller, and J.L. Wang, *Properties of principal component methods for functional and longitudinal data analysis*, Ann. Stat. (34) 2006), pp. 1493–1517.
[12] T.G. Heckman, B.D. Heckman, T. Anderson, T.I. Lovejoy, J.C. Markowitz, Y. Shen, and M. Sutton, *Tele-interpersonal psychotherapy acutely reduces depressive symptoms in depressed HIV-infected rural persons: A randomized clinical trial*, Behav. Med. 43 (2017), pp. 285–295.
[13] T.G. Heckman, J.C. Markowitz, B.D. Heckman, H. Woldu, T. Anderson, T.I. Lovejoy, Y. Shen, M. Sutton, and W. Yarber, *A randomized clinical trial showing persisting reductions in depressive symptoms in HIV-infected rural adults following brief telephone-administered interpersonal psychotherapy*, Ann. Behav. Med. 52 (2018), pp. 299–308.
[14] D. Hedeker and R.D. Gibbons, *Longitudinal Data Analysis*, Vol. 451, John Wiley & Sons, Hoboken, NJ, 2006.
[15] J. Illian, E. Benson, J. Crawford, and H. Staines, *Principal component analysis for spatial point processes-assessing the appropriateness of the approach in an ecological context*, in *Case Studies in Spatial Point Process Modeling*, A. Baddeley, P. Gregori, J. Mateu, R. Stoica, and D. Stoyan, eds., Springer, New York, NY, 2006, pp. 135–150
[16] J.B. Illian, J.I. Prosser, K.L. Baker, and J.I. Rangel-Castro, *Functional principal component data analysis: A new method for analysing microbial community fingerprints*, J. Microbiol. Methods 79 (2009), pp. 89–95.
[17] D. Lin and Z. Ying, *Semiparametric and nonparametric regression analysis of longitudinal data*, J. Am. Stat. Assoc. 96 (2001), pp. 103–126.
[18] J. McGrath, A. Barnett, D. Eyles, T. Burne, C.B. Pedersen, and P.B. Mortensen, *The impact of nonlinear exposure-risk relationships on seasonal time-series data: Modelling Danish neonatal birth anthropometric data*, BMC. Med. Res. Methodol. 7 (2007), p. 1.
[19] H.D. Modanlou and Y. Murata, *Sinusoidal heart rate pattern: Reappraisal of its definition and clinical significance*, J. Obstet. Gynaecol. Res. 30 (2004), pp. 169–180.
[20] P.J. Morgan, D.R. Lubans, C.E. Collins, J.M. Warren, and R. Callister, *12-month outcomes and process evaluation of the SHED-IT RCT: An internet-based weight loss program targeting men*, Obesity 19 (2011), pp. 142–151.
[21] E. Park, M. Cho, and C.S. Ki, *Correct use of repeated measures analysis of variance*, Korean. J. Lab. Med. 29 (2009), pp. 1–9.
[22] J.O. Ramsay, *Functional Data Analysis*, Wiley Online Library, New York, NY, 2006.
[23] J.O. Ramsay, G. Hooker, and S. Graves, *Functional Data Analysis with R and MATLAB*, Springer, New York, NY, 2009.
[24] J. Ramsay, H. Wickham, S. Graves, and G. Hooker, *FDA: Functional data analysis*, R Package, 2013. Available at http://cran.r-project.org/package = fda.
[25] D. Ruppert, R.J. Carroll, and M.P. Wand, *Semiparametric Regression*, Cambridge University Press, Cambridge, 2005.
[26] N. Shaadan, A.A. Jemain, and S.M. Deni, *Data Preparation for Functional Data Analysis of PM10 in Peninsular Malaysia*, Proceedings of the 21st National Symposium on Mathematical Sciences (SKSM21): Germination of Mathematical Sciences Education and Research towards Global Sustainability, Vol. 1605, AIP Publishing, 2014, pp. 850–855.

[27] H. Sørensen, J. Goldsmith, and L.M. Sangalli, *An introduction with medical applications to functional data analysis*, Stat. Med. 32 (2013), pp. 5222–5240.
[28] R.L. Spitzer, J.B. Williams, K. Kroenke, M. Linzer, F. Verloin deGruy, S.R. Hahn, D. Brody, and J.G. Johnson, *Utility of a new procedure for diagnosing mental disorders in primary care: The PRIME-MD 1000 study*, JAMA 272 (1994), pp. 1749–1756.
[29] E.L. Teng and H.C. Chui, *The modified mini-mental state (3ms) examination*, J. Clin. Psychiatry 48 (1987), pp. 314–318.
[30] A. Tremblay and A.J. Newman, *Modeling nonlinear relationships in ERP data using mixed-effects regression with R examples*, Psychophysiology 52 (2014), pp. 124–139.
[31] U.S.D. of Agricultrue, *Rural-urban continuum codes*, 2015. Available at http://www.ers.usda.gov/data-products/rural-urban-continuum-codes.aspx, accessed May 19, 2015.
[32] G. Verbeke and G. Molenberghs, *Linear Mixed Models for Longitudinal Data*, Springer, New York, NY, 2009.
[33] R.M. West, K. Harris, M.S. Gilthorpe, C. Tolman, and E.J. Will, *Functional data analysis applied to a randomized controlled clinical trial in hemodialysis patients describes the variability of patient responses in the control of renal anemia*, J. Am. Soc. Nephrol. 18 (2007), pp. 2371–2376.
[34] F. Yang, A. M. Parkhurst, C. N. Lee, T. M. Brown-Brandl, and P. E. Hillman, *Using functional data analysis to evaluate effect of shade on body temperature of feedlot heifers during environmental heat stress*Annual Conference on Applied Statistics in Agriculture, 2013. Available at http://newprairiepress.org/agstatconference/2013/proceedings/12.
[35] X. Yang, Q. Shen, H. Xu, and S. Shoptaw, *Functional regression analysis using an f test for longitudinal data with large numbers of repeated measures*, Stat. Med. 26 (2007), pp. 1552–1566.
[36] K. Yaraee, *Functional data analysis with application to MS and cervical vertebrae data*, Ph.D. diss., University of Alberta, 2011.
[37] S.L. Zeger and P.J. Diggle, *Semiparametric models for longitudinal data with application to CD4 cell numbers in HIV seroconverters*, Biometrics 50 (1994), pp. 689–699.
[38] D. Zhang, X. Lin, J. Raz, and M. Sowers, *Semiparametric stochastic mixed models for longitudinal data*, J. Am. Stat. Assoc. 93 (1998), pp. 710–719.