Title

Functional Regression Analysis using an F Test for Longitudinal Data with Large Numbers of Repeated Measures

Authors

Yang, Xiaowei
Shen, Qing
Xu, Hongquan
et al.

Publication Date

2005-07-01

Functional Regression Analysis using an F Test for Longitudinal Data with Large Numbers of Repeated Measures

Xiaowei Yang[1,2], Qing Shen[3], Hongquan Xu[4], and Steven Shoptaw[1,2]

1. BayesSoft, Inc.
   3641 Midvale Avenue, #207, Los Angeles, CA 90034
   E-mail: XYang@BayesSoft.com
2. UCLA-Integrated Substance Abuse Programs
   11075 Santa Monica Blvd, Suite 200, Los Angeles, CA 90025
3. Edmunds.com Inc.
   2401 Colorado Ave., Suite 250, Santa Monica, CA 90404
4. UCLA, Department of Statistics
   PO Box 951554, Los Angeles, CA 90095-1554
   Email: hqxu@stat.ucla.edu

SUMMARY

Longitudinal data sets from certain fields of biomedical research often consist of several variables repeatedly measured on each subject yielding large number of observations. This characteristic complicates the use of standard longitudinal modeling strategies, such as random effects models and marginal models, where rigorous assumptions on intra-subject correlation structure are required. An innovative way to model the data is to apply functional regression analysis, an emerging statistical approach in which observations of the same subject are viewed as a sample from a functional space. No assumptions are needed for the intra-subject correlation structure. Shen and Faraway (*Satistica Sinica* 2004; 1239-1257) introduced an F test for linear models with functional responses. This paper illustrates how to apply this F test and functional regression analysis to the setting of longitudinal data where intra-subject repeated measures are viewed as discrete samples from an underlying curve with continuous function forms. A smoking cessation study for methadone-maintained tobacco smokers is analyzed for demonstration. In estimating the treatment effects, the functional regression analysis provides meaningful clinical interpretations, and the functional F test provides consistent results supported by a random-effects linear regression model.

KEY WORDS: Functional F Test; Functional Data Analysis; Functional Regression Analysis; Longitudinal Data Analysis.

1 INTRODUCTION

In biomedical research with longitudinal studies, subjects are repeatedly measured for a set of characteristics so that time-varying causal relationships between the responses and explanatory variables of interest can be modeled, e.g., growth trajectory and disease progression [1]. Intra-subject correlation across repeated measurements is of concern when considering approaches for longitudinal data analysis. When the correlation and mean structures are relatively simple or can be assumed to be certain parametric forms, satisfactory performance can be anticipated using standard longitudinal modeling strategies e.g., mixed effects models, marginal models, and Markov transition models [2]. In certain fields of study, such as substance abuse, environmental, and public health research, repeated measures are sometimes collected at high frequencies over long periods of time. For example, in a 12-week smoking cessation study, carbon monoxide levels were collected three times weekly on each methadone-maintained tobacco smoker [3]. To analyze such longitudinal data with large-scale time grids, it is often insufficient to assume a linear or polynomial form of the mean structure, a compound symmetric or autoregressive pattern of the covariance matrix, or a Markov process with stable transitional probabilities. Other multivariate-observation approaches, such as hierarchical models, latent variable models, and structure equation models, involve many parameters with unverifiable assumptions [4-6]. As of yet, there are few alternatives that successfully address the unique problems presented by data collected in longitudinal studies with high dimensionality. This paper evaluates a recently developed method of functional data analysis for this purpose.

In the emerging statistical research field, functional data analysis refers to a collection of strategies for analyzing functional data sets, such as curves, images, or shapes [7]. To a study observing seated automobile drivers' body motion patterns [8, 9], and to a study of urinary metabolites and a progesterone data set [10], various strategies of functional regression analysis have been applied.

Until very recently, functional data analysis and longitudinal data analysis have been viewed as distinct enterprises [11]. For longitudinal data with dense time grids, one could understand within-subject repeated measures as discrete samples from a functional curve over the studied time interval. A curve for each subject's response can be obtained via various smoothing techniques in connecting the discrete data points [12] and these individual subject response curves can be tested using functional data analysis. The approach to using functional data analysis provides an alternative with innovative insights to the practice of longitudinal data analysis. Unlike the using of long-form to represent longitudinal data in computer procedures fitting standard longitudinal models (e.g., PROC Mixed in SAS), functional regression analysis does not change the original wide-form format of the data structure, and thus is more intuitive to data analysts. With time-dependent coefficients, functional regression analysis captures the time-varying exposure-response relationship, thus providing a simpler data structure with intuitive interpretations. A time series plot of the estimated coefficient function vividly reveals how the effect of a predictor can change along the time axis. Most importantly, functional regression analysis draws more robust conclusions as it has features similar to nonparametric methods, requiring few assumptions on the intra-subject error correlation structure and mean structures for the studied population.

2. FUNCTIONAL LINEAR REGRESSION MODELS

A longitudinal study, usually collects continuous repeated measures, $\{ y_i(t_{ij}) ;\ i = 1,...,n$, $j = 1,...,m \}$, on a time grid, $\{t_1,...,t_m\}$, that is either exactly or approximately the same for all n subjects. One may restrict the same number of repeated measures to be collected on each subject. Ideally, these repeated measures can be viewed as discrete samples from a continuous response curve, $y_i(t)$. In this setting, a functional linear regression model has the form of

$$y_i(t) = x_i^T \beta(t) + \varepsilon_i(t),$$

where $x_i = (x_{i1},...,x_{ip})^T$ is a vector of fixed covariates or predictor variables, $\beta(t) = (\beta_1(t),...,\beta_p(t))^T$ is a vector of coefficient functions, and $\varepsilon_i(t)$ is an error function of Gaussian process with mean zero and unknown covariance function $r(s,t) = \text{cov}(\varepsilon_i(s), \varepsilon_i(t))$. Note that $\varepsilon_i(t)$ and $\varepsilon_k(t)$ are independent to each other when $i \neq k$.

The coefficient function $\beta(t)$ can be estimated by the least squares method, which leads to

$$\hat{\beta}(t) = (X^T X)^{-1} X^T Y(t),$$

where $X = (x_1,...x_n)^T$ is the model matrix and $Y(t) = (y_1(t),...y_n(t))^T$ is the vector of response functions. The predicted (or fitted) responses are $\hat{y}_i(t) = x_i^T \hat{\beta}(t)$ and the

residuals are $\hat{\varepsilon}_i(t) = y_i(t) - \hat{y}_i(t)$. The residual sum of squares is $rss = \sum_{i=1}^{n} \int (y_i(t) - \hat{y}_i(t))^2 dt$.

In reality, only a finite number of measures (i.e., $y_i(t_{ij})$'s) exist for the i^{th} response curve (i.e., $y_i(t)$). To apply functional regression analysis to discrete observational data, Shen and Faraway [9] recommended analyzing the un-smoothed raw data directly over a common grid of time for different subjects. Otherwise, one may reconstruct the response curve from the observed data points to get estimates of $y_i(t)$ over a common grid $\{t_j ; j = 1,...,m\}$ via proper smoothing techniques, e.g., model-based cross-validation methods [12], kernel-based or spline-based nonparametric regression methods [13], and robust methods such as LOWESS [14]. The choice of different smoothing techniques would have little impact on the analysis if there are plentiful underlying response curves (i.e., $y_i(t)$'s) with quite smooth functional forms.

2.1. A Functional F Test for Hypothesis Testing and Model Selection

An important inference problem is to compare two nested linear models, ω and Ω, where $\dim(\omega) = q$, $\dim(\Omega) = p$, and model ω results from a linear restriction on the parameters of model Ω. There are relatively few satisfactory solutions available in the statistical literature to this situation. Ramsay and Silverman [7] suggested a naive approach by examining the point-wise F statistics on each time point for testing $\beta(t)$. This carries a serious problem with multiple-comparison and if Bonferroni corrections were applied to the significance level, power would be significantly compromised

considering that repeated measures are often strongly correlated. As pointed out by Faraway [8], traditional multivariate test statistics are inappropriate due to the influence of unimportant variation directions..

To overcome these issues, Shen and Faraway [9] proposed a functional F test. Define

$$F = \frac{(rss_\omega - rss_\Omega)/(p-q)}{rss_\Omega/(n-p)},$$

where rss_ω and rss_Ω are residual sum of squares under models ω and Ω, respectively. The null distribution of this statistic is $((n-p)/(p-q))\sum_{k=1}^{\infty} r_k \chi^2_{(p-q)} \Big/ \sum_{k=1}^{\infty} r_k \chi^2_{(n-p)}$, where $r_1 \geq r_2 \geq \ldots \geq 0$ are eigenvalues of the covariance function $r(s,t)$ and all the χ^2 random variables are independent of each other. This null distribution can be effectively approximated by an ordinary F distribution with degrees of freedoms $df_1 = [\lambda(p-q)]$ and $df_2 = [\lambda(n-p)]$, where $\lambda = (\sum_{k=1}^{\infty} r_k)^2 / \sum_{k=1}^{\infty} r_k^2$ is the *degrees-of-freedom-adjustment-factor* [9] and $[\cdot]$ represents the operator to get the largest integer.

In practice, when repeated measures are observed on an evenly spaced time grid $\{t_1,...,t_m\}$, we should replace the integration with summation, compute $rss = \sum_{i=1}^{n}\sum_{k=1}^{m}(y_i(t_k) - \hat{y}_i(t_k))^2 / m$ and estimate the degrees-of-freedom-adjustment-factor by trace$(E)^2$/trace(E^2), where E is the empirical covariance matrix computed from the alternative or full model.

It is important to note that the functional F test statistics make sense when the grid size m becomes large, while most multivariate test statistics [15, 16] would fail. Other important work addressing the functional testing problem was provided by Fan and Lin

[17], Eubank [18], and Abramovich et al. [19], but they only considered ANOVA type models and their test statistics were formed by orthogonal (Fourier or Wavelets) expansion coefficients of response curves. Eubank [18] proved that among different ways of combining the coefficients into a test statistic, the L^2 norm, a simple sum of the squared coefficients, is asymptotically equivalent to the uniformly most powerful test when the grid size m goes to infinity. This result provides important evidence that the functional F-test statistic, which uses L^2 norm of the residual curves, is not only computationally cheaper but also more powerful than other methods.

Model selection is an important issue in regression analysis. Stepwise model selection requires an easy way of calibrating the p-value of a predictor in the full model, i.e., to test the null hypothesis "$H_{0j}: \beta_j(t) = 0$ for $j = 1,...,p$" against the full model hypothesis "$H_1: Y(t) = X\beta(t) + \varepsilon(t)$." To test these hypotheses, one can fit each null model H_{0j} separately for $j = 1,...,p$, and then use functional F statistics $F_j = \dfrac{rss_{0j} - rss_1}{rss_1/(n-p)}$ to make a decision on accepting or rejecting the null model. As shown by Shen and Faraway [9], it is indeed unnecessary to fit all the p null models, because F_j can be derived from quantities obtained directly from the fitting of the full model H_1, i.e.,

$$F_j = \frac{(n-p)\int \hat{\beta}_j^2(t)dt}{(X^TX)^{-1}_{jj} rss_1},$$

where $(X^TX)^{-1}_{jj}$ denotes the j^{th} diagonal element of $(X^TX)^{-1}$, $\hat{\beta}_j(t)$ is the estimate of $\beta_j(t)$, and rss_1 is the residual sum of squares under the full model H_1. In practice, the

operation of integration is replaced by that of summation. The null distribution of the functional F statistic F_j can be approximated by an ordinary F distribution with degrees of freedom $df_1 = [\lambda]$ and $df_2 = [(n-p)\lambda]$, where λ is the degrees-of-freedom-adjustment-factor.

2.2. Diagnostic Check

It is important to identify outliers and highly influential curves (subjects) since including them in the analysis may give misleading results. As in the context of traditional linear regression for scalar responses, we define jackknife residuals and Cook statistics for functional regression. Let $H = X(X^TX)^{-1}X^T$ be the hat matrix and define leverage h_{ii} as the diagonal entry of H. Define studentized residual as

$$s_i = \frac{\sqrt{\int \hat{\varepsilon}_i^2(t)dt}}{\sqrt{(1-h_{ii})rss/(n-p)}},$$

and jackknife residual as

$$J_i = \frac{\sqrt{\int \hat{\varepsilon}_{(i)}^2(t)dt}}{\sqrt{[1+x_i^T(X_{(i)}^TX_{(i)})^{-1}x_i][rss_{(i)}/(n-p-1)]}},$$

where $X_{(i)}$ is the X matrix with the i^{th} row deleted, $\hat{\varepsilon}_{(i)}^2(t)$ is the i^{th} residual from the model without the i^{th} curve, and $rss_{(i)}$ is the residual sum of squares from the model without the i^{th} curve. Define Cook statistics as

$$D_i = \frac{\int(\hat{\beta}_{(i)}(t)-\hat{\beta}(t))^T(X^TX)(\hat{\beta}_{(i)}(t)-\hat{\beta}(t))dt}{rss}\cdot\frac{n-p}{p},$$

where $\hat{\beta}_{(i)}(t)$ is the estimate of $\beta(t)$ computed without the i^{th} curve. The null distribution of jackknife residuals can be approximated by an ordinary F distribution with degrees of freedom $df_1 = [\lambda]$ and $df_2 = [(n - p - 1)\lambda]$, which can be used to detect outliers.

Shen [20] showed that jackknife residuals and Cook statistics can be computed directly from the studentized residuals and leverages as follows:

$$J_i = s_i \sqrt{\frac{n - p - 1}{n - p - s_i^2}} \quad \text{and} \quad D_i = \frac{s_i^2}{p} \cdot \frac{h_{ii}}{1 - h_{ii}}.$$

3. APPLICATION TO A SMOKING CESSATION CLINICAL TRIAL

A 12-week clinical trial was performed to evaluate relapse prevention (RP) and contingency management (CM) as smoking cessation therapies for methadone-maintained tobacco smokers [3]. A total of 174 subjects were randomly assigned to one of four treatment conditions (Control; RP-only; CM-only; RP + CM). All subjects received nicotine replacement therapy in addition to their assignment to behavioral therapies: RP and/or CM. The repeated measures of most interest in this study were breath samples collected three times weekly, which were analyzed for carbon monoxide levels (parts per million) to indicate recent tobacco smoking abstinence. The observed carbon monoxide levels and their mean profiles for each group are depicted in Figure 1. Participants' age (Age), baseline carbon monoxide levels ($BaseCO$), and numbers of nicotine patches ($Patches$) were recorded as other predictors along with treatment conditions. This data set has been analyzed using standard linear mixed effects models,

which provide some known results that can be used to evaluate the performance of functional regression analysis.

<<Insert Figure 1>>

For significance testing, an insufficient approach was first applied to compare the carbon monoxide levels across treatment conditions on any given time point using the method of ANOVA [7]. As depicted by Figure 2, at eight points significantly different carbon monoxide levels were indicated by the point-wise ANOVA with p-values smaller than 0.001. Because of the problem of "multiple comparison" [21], a significance level of 0.001 was used instead of the usual level of 0.05. Although this method provides some useful information in comparing the responses between treatment conditions, it is relatively limited in making inferences on the overall treatment efficacy. There is no simple way of combining these multiple p-values. Moreover, the point-wise ANOVA ignored the patterns showing that the average carbon monoxide levels were almost consistently lower for the treatment conditions involving contingency management. Therefore, the challenge for this simple method is how to combine these tests to achieve a powerful overall test. Another "time-naïve" approach is to compress the intra-subject repeated measures into a composite score, such as a mean level of carbon monoxide for each smoker. Using this method, a significant difference ($F_{170}^{3} = 5.04$ and p-value=0.002) between conditions was observed. A problem with this aggregation approach, however, is that it erodes both statistical power and validity [22]. This testing method requires that missing data must be "missing completely at random," which is usually too rigorous. As

discussed by Yang and Shoptaw [23], there is evidence that missing values in this carbon monoxide level data were not missing complete at random.

<<Insert Figure 2>>

Because 20% of the carbon monoxide levels were missing due to either occasional omission or to premature withdrawal, multiple imputation [24] was applied. After a logarithmic transformation, repeated carbon monoxide levels for each participant could be assumed to be multivariate normally distributed. Specifying a normal prior distribution and an inverted Wishart distribution for the covariance matrix, we conducted multiple imputation using an iterative procedure called data augmentation, which was originally proposed by Schafer [25]. This process consists of two steps on each iterate. In the *imputation step*, for each person, we draw imputations of missing values conditionally on the observed values using a conditional normal distribution with parameters drawn in the previous iteration. In the parameter *estimation step*, new parameters on the mean and covariance of the multivariate normal distribution are updated, given the complete data set with current imputed values. Repeating these two steps until the procedure converged, four sets of imputations were obtained with an interval of 500 iterations, yielding four complete data sets, each with different imputations of the missing values. For each complete data set, the full model, including all the interesting predictors, was fitted using the method of least squares estimation,

$$y(t) = \beta_0(t) + CM \cdot \beta_1(t) + RP \cdot \beta_2(t) + CM * RP \cdot \beta_3(t) + BaseCO \cdot \beta_4(t) + Age \cdot \beta_5(t) + Patches \cdot \beta_6(t) + \varepsilon(t)$$

where $CM = 1$ (or 0) indicates whether a subject received contingency management (or not), $RP = 1$ (or 0) indicates whether a subject received relapse prevention (or not), and

$CM * RP$ is an interaction term. In this coding scheme, the control group was coded as "$CM = 0$ and $RP = 0$," and the RP+CM groups was coded by "$CM = 1$ and $RP = 1$." Since there was little difference between the four imputed data sets, the fitted curves for the coefficient functions were plotted in Figure 3 for the first imputed data set. The fitted coefficient functions of *RP* and Age are close to the zero function, indicating that the treatment effect of the relapse prevention and the age effect are negligible. Further, the interaction term *CM*RP* is not significant, indicating that contingency management does not interact with relapse prevention. Regression coefficient functions for *CM* and *Patches* are negative-valued throughout, suggesting favorable effects of contingency management and nicotine patch replacement. By contrast, the positive-valued coefficient function of the baseline carbon monoxide level implied that the higher the baseline carbon monoxide level, the more difficult to achieve tobacco abstinence.

<<Insert Figure 3>>

The functional F test statistics and their p-values of each predictor in this model are listed in Table 1. For all four complete data sets, only the terms, CM, $BaseCO$, and $Patches$ are significant at significance level of $\alpha = 0.05$. After removing insignificant terms (RP, $CM * RP$, and Age), the reduced model was fitted to the imputed data sets. The functional F test statistics and their p-values for the remaining terms were obtained (see Table 2).

<<Insert Table 1>>

<<Insert Table 2>>

As expected, all predictors were significant at $\alpha = 0.01$ level this time. Since all the four data sets consistently supported the same results, we accept this three-predictor functional regression model as the final model to make inferences:

$$y(t) = \beta_0(t) + CM \cdot \beta_1(t) + BaseCO \cdot \beta_2(t) + Patches \cdot \beta_3(t) + \varepsilon(t),$$

where the subscript indicating subjects is suppressed. The fitting of this model indicated that, after adjusting for the baseline levels ($BaseCO$) and number of nicotine patches applied ($Patches$), contingency management turned out to be significantly effective in helping this specific group of smokers achieving tobacco abstinence during treatment.

To check diagnostics for the above selected model, jackknife residuals and Cook statistics for all the imputed data sets were computed. The charts of these statistics from the first imputed data set are shown in Figure 4. The jackknife residuals for the participants numbered 92 and 93 are bigger than the critical value of the functional F distribution at significance level of $\alpha = 0.05$. Therefore, these two smokers may be declared as outliers. The record associated with the subject numbered 92 is also a highly influential point according to the Cook statistics. Checking the original records, both points with unusually high values for most of the observations were noted. After excluding these two "outliers," we re-analyzed the data using the above models and found consistent results.

<<Insert Figure 4>>

We also analyzed the four complete data sets after imputation by a linear mixed effects model with random intercept to model heterogeneities across subjects:

$$y_{ij} = \beta_0 + CM \cdot \beta_1 + RP \cdot \beta_2 + CM * RP \cdot \beta_3 + BaseCO_i \cdot \beta_4 + Age_i \cdot \beta_5 + Patches_i \cdot \beta_6 + u_i + \varepsilon_{ij}$$

,

where y_{ij} stands for the j^{th} carbon monoxide level of the i^{th} smoker, CM, RP, $RP*CM$, $BaseCO$, Age, and $Patches$ are fixed effects that are common for all observations on the same subject, $u_i \sim N(0,\sigma_u^2)$ is the random intercept effect explaining the heterogeneity across subjects, and ε_{ij}'s are identically independently distributed normal random errors. Consistent conclusions were observed by fitting this linear mixed effects model. *CM* (p-value<0.01) is significant while *RP* and *CM*RP* are not significant. *Age* (p-value=0.43) is not significant while *BaseCO* (p-value<0.01) and *Patch* (p-value<0.01) are significant.

As seen in this example, the scalar linear mixed effects model and the functional regression model differ in at least two ways. First, in the mixed effects models the fixed effects (i.e., β) are time independent, while in the functional regression model the effects (i.e., $\beta(t)$) are functions over time. Second, the mixed effects model assumes compound symmetry error correlation structure within each smoker, while the functional regression model assumes no specific simple forms on the intra-subject correlation structure. Thus, the functional regression model is more flexible and the conclusions drawn are more robust in general. The time series plots of the estimated coefficient functions in Figure 3 for the functional regression model provide more vivid information with intuitive clinical interpretation than the point estimates of the effects of CM, RP, $RP*CM$, $BaseCO$, Age, and $Patches$ in the mixed effects model. For example, there appears to be a slightly increasing negative effect of $Patches$ over time. Although functional regression analysis and scalar linear mixed effects models supported no strong overall age effect, a negative influence of age on carbon monoxide levels (higher ages associate with lower

carbon monoxide levels) was noticed starting the eighth week. It appeared that older smokers stayed longer in the study, and the longer they stayed, the more likelihood they were to achieve smoking abstinence as measured using carbon monoxide levels.

4. DISCUSSION

As a companion to the work of Shen and Faraway [9], this paper demonstrates the functional F test for functional regression analyses using a longitudinal data analysis with a fairly large number of repeated observations measured on each object. The functional F test for functional linear regression models is appropriate for evaluating effectiveness of experimental conditions. Application of the technique to a longitudinal trial of smoking cessation techniques indicates that the use of functional regression analysis is a valid and promising alternative to traditional longitudinal modeling strategies.

Functional regression analysis also provides researchers in longitudinal study of biomedical and social-economic topics a promising alternative to gain deeper understanding of effects of interventions. Compared with traditional random-effects and marginal models, functional regression analysis treats data from the same participant as a single observation in a functional space and provides a natural and simpler data structure. Functional regression analysis also requires fewer restrictions on the intra-subject correlation structure, thus should be more robust to any violation of such assumptions. Missing values or unbalanced longitudinal data can be handled easily by applying smoothing techniques that do not require a common fixed time-grid. Additionally,

functional regression coefficients provide both intuitive and time-dependent estimators thereby yielding insights for studying time-varying relationships.

Similar ideas on the functional F test could be traced to Box [26], where the property of the F-test statistic in the two-way ANOVA for correlated data was studied in detail. Other ways of functional data analysis were provided by Fan and Lin [17], who used adaptive Neyman or thresholding tests on the Fourier or wavelet expansion coefficients of the estimated parameter function in order to compare groups of curves. Recently, these testing methods have been further extended by Nie [27] within the setting of functional regression analysis. As suggested by Eubank [18], these transform-based methods are complicated and may not ultimately boost power. Since the functional regression model, restricted to the finite time grid, becomes a standard multivariate problem, it is natural to try multivariate-based tests. Shen [20] carefully compared the performance of the functional F test with a traditional multivariate likelihood ratio test and its variation, such as a B-spline coefficient test, and found that the functional F test had at least the following advantages: (i) it works when the grid size becomes large; (ii) it is stable and not easily influenced by unimportant variation directions; (iii) it is always fairly powerful; and (iv) it is computationally cheap. These reasons provide strong rationale for applying functional regression analysis with functional F test in practice.

A limitation of functional regression analysis in this paper is that it models repeated measures that are assumed of Gaussian distribution. Although the large sample theory ensures the use of functional regression analysis in wider applications, more specific forms of functional models and functional F test statistics need to be developed for other types of longitudinal data. Another limitation of the functional regression

analysis for longitudinal data analysis comes from missing data problems, which are common in practice. In the smoking cessation data, missing data were assumed "ignorable" [24], so that multiple imputations could be created using an MCMC algorithm proposed by Schafer [25]. When assuming such a process, analyses based only on observed data, while ignoring missing values, would provide unbiased estimates. If missing data are ignorable, they can be interpolated or extrapolated from the underlying continuous functional curves in functional regression analysis. Unfortunately, this assumption of ignorability could not be verified in this smoking cessation study without follow-up investigations. It is urgent that functional regression analysis or functional data analysis be developed to analyze longitudinal data sets with non-ignorable missing values.

ACKNOWLEDGEMENTS

This work was partially supported by the National Institute of Drug Abuse through an SBIR contract N44 DA35513 and two research grants: R03 DA016721 and P50 DA 18185. We especially thank Hamutahl Cohen for her editorial assistance.

REFERENCES

1. Hand DJ, Crowder MJ. *Practical longitudinal Data Analysis*. Chapman and Hall, 1996.
2. Diggle PJ, Liang KY, Zeger, SL. *Analysis of Longitudinal Data*. Oxford University Press, 1994.
3. Shoptaw S, Rotheram-Fuller E, Yang X, Frosch D, Nahom D, Jarvik ME, Rawson RA, Ling W. Smoking cessation in methadone maintenance. *Addiction* 2002; **97**:1317-1328.
4. Singer SCJ, Willett, JB. *Applied Longitudinal Data Analysis: Modeling Change and Event Occurrence*. Oxford Press, 2003.
5. Bryk AS, Raudenbush SW. *Hierarchical Linear Models: Applications and Data Analysis Methods* (2nd edn.). Sage Publications, 2001.
6. Kaplan D. *Structural Equation Modeling: Foundations and Extensions*. Sage Publication, 2000.
7. Ramsay JO, Silverman BW. *Functional Data Analysis*. Springer-Verlag: New York, 1997.
8. Faraway, JJ. Regression analysis for a functional response. *Technometrics* 1997; **39**:254–261.
9. Shen Q, Faraway J. An F Test for Linear Models with Functional Responses, *Statistica Sinica* 2004; **14**:1239-1257.

10. Brumback BA, Rice JA. Smoothing Spline Models for the Analysis of Nested and Crossed Samples of Curves (with Discussion), *Journal of the American Statistical Association* 1998; **93**:961-994.
11. Zhao X, Marron JS, Wells MT. The Functional Data Analysis View of Longitudinal Data. *Statistica Sinica* 2004; **14**:789-808.
12. Rice JA, Silverman BW. Estimating the mean and covariance structure nonparametrically when the data are curves. *Journal of the Royal Statistical Society Series B* 1991; **53**, 233–243.
13. Wahba G. *Spline Models for Observational Data*, Society for Industrial and Applied Mathematics, Philadelphia, PA, 1990.
14. Cleveland W. Robust Locally Weighted Regression and Smoothing Scatterplots, *Journal of the American Statistical Association* 1979; **74**:829-836.
15. Johnson R, Wichern D. *Applied Multivariate Statistical Analysis* (5th edn.). Prentice Hall: New Jersey, 2002.
16. Rencher AC. *Methods of Multivariate Analysis* (2nd edn.). John Wiley & Sons: New York, 2002.
17. Fan J, Lin S-K. Tests of significance when data are curves. *Journal of the American Statistical Association* 1998**; 93**:1007-1021.
18. Eubank RL. Testing for no effect by cosine series methods. *Scandinavian Journal of Statistics* 2000; **27**:747–763.
19. Abramovich F, Antoniadis A, Sapatinas T, Vidakovic B. Optimal testing in a fixed-effects functional analysis of variance model. *International Journal of Wavelets, Multiresolution and Information Processing* 2004; **2**:323-349.

20. Shen Q. *Linear Models for a Functional Response*. Ph.D. Thesis, University of Michigan, Ann Arbor, MI, 1999.

21. Hsu J. *Multiple Comparison: Theory and Methods*. Chapman & Hall / CRC, 1996.

22. Hall SM, Delucchi KL, Velicer W, Kahler C, Ranger-Moore J, Hedeker D, Tsoh J, Niaura R. Statistical analysis of randomized trials in tobacco treatment. *Nicotine and Tobacco Research* 2001; **3**:193-202.

23. Yang X, and Shoptaw S. Assessing missing data assumptions in longitudinal studies: an example using a smoking cessation trial. *Drug and Alcohol Dependence* 2005; **77**:213-225.

24. Rubin DB. *Multiple Imputation for Nonresponse in Surveys*. John Wiley & Sons: New York, 1987.

25. Schafer JL. *Analysis of Incomplete Multivariate Data.* Chapman and Hall: London, 1997.

26. Box GEP. Some Theorems on Quadratic Forms Applied in the Study of Analysis of Variance Problems: II. The Effect of Inequality of Variance and Correlation between Errors in the Two-way Classification. *Annals of Mathematical Statistics* 1954; **25**:484-498.

27. Nie K. *Hypothesis for High Dimensional Data*. Ph.D. Thesis, Department of Biostatistics, University of California, Los Angeles, 2004.

Table 1. Observed functional F test statistics (and P-values) for each covariate

Data Set	*Intercept*	*CM*	*RP*	*CM*RP*	*BaseCO*	*Age*	*Patches*
Impute 1	98.9(*)	5.98(*)	0.89(.45)	1.11(.34)	24.8(*)	1.25(.29)	24.9(*)
Impute 2	98.0(*)	4.89(*)	0.78(.51)	1.17(.32)	24.2(*)	1.83(.14)	21.6(*)
Impute 3	104.5(*)	5.99(*)	0.71(.54)	1.07(.36)	26.1(*)	1.18(.32)	30.9(*)
Impute 4	96.2(*)	5.01(*)	0.91(.43)	1.29(.28)	25.7(*)	1.05(.37)	24.8(*)

Note: * P-values are smaller than 0.01.

Table 2. Functional F test statistics for each covariate in the final functional regression model

Data Set	*Intercept*	*CM*	*BaseCO*	*Patches*
Impute 1	254.6	14.71	25.1	27.3
Impute 2	239.3	13.75	24.7	24.0
Impute 3	272.0	15.35	26.6	33.4
Impute 4	250.7	14.04	26.3	26.8

Note: All p-values are smaller than 0.01

FIGURE CAPTIONS

Figure 1. Mean levels of the carbon monoxide across the treatment groups. For each plot, the y-axis indicates log(1+y) transform of the original level of carbon monoxide (p.p.m.), the x-axis indicates number of clinic visit for study participants (1, …, 36). Both individual profiles and the mean profile are plotted for each of the four treatment conditions: Control, RP-only, CM-only, and RP+CM (RP=Relapse Prevention, CM=contingency Management).

Figure 2. The average and standard deviation (SD) curves for the log-scaled carbon monoxide levels. On this plot, the four mean curves of the log-scaled carbon monoxide levels and the corresponding point-wise standard errors are drawn for each of the four treatment conditions: Control, RP-only, CM-only, and RP+CM (RP=Relapse Prevention, CM=Contingency Management). Vertical bars indicate the estimated standard errors of average carbon monoxide levels. The stars ("*") over the x-axis mark the time points (i.e., visit numbers) where the carbon monoxide levels are significantly different indicated by a point-wise ANOVA (P-value<0.001). Y-axis indicates values of carbon monoxide levels after log(1+y) transform. X-axis represents number of clinic visit for study participants (1 ,…, 36).

Figure 3. Estimated regression coefficient functions in functional regression analysis for the first imputed data set. The top panel shows the regression coefficient functions corresponding to effects of CM treatment, RP treatment and their interaction (CM*RP); the bottom panel depicts the regression coefficient functions corresponding to baseline

carbon monoxide level (BaseCO), smoker' age (Age), and number of nicotine patches a smoker has received during the study (Patches). Y-axis indicates values of regression coefficients and x-axis indicates number of clinic visit for each smoker (1, …, 36).

Figure 4. Diagnostics for the first imputed data set. The left panel draws jackknife residuals and the right panel depicts Cook statistics calculated from the functional regression model including three predictors: *CM*, *Baseco*, and *Patches*. In both plots, the x-axis corresponds to the labels of the 174 participants in the study. The y-axis corresponds to either the values of jackknife residuals or Cook statistics. Two subjects (numbered 92 and 93) have jackknife residuals noticeable high and one subject (numbered 92) associates with the highest Cook statistics.

Figure 1. Mean plots of the carbon monoxide levels across the treatment groups

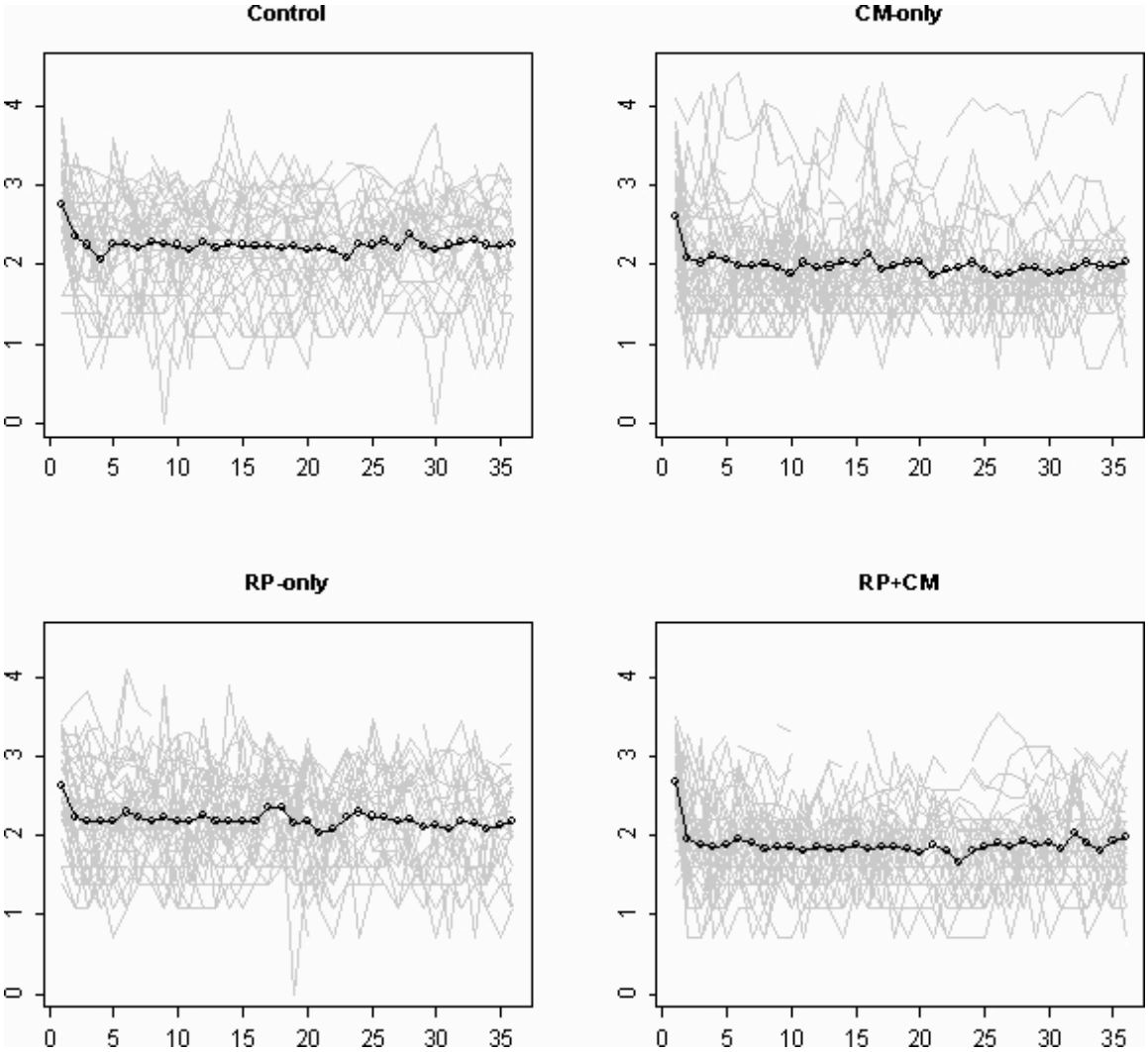
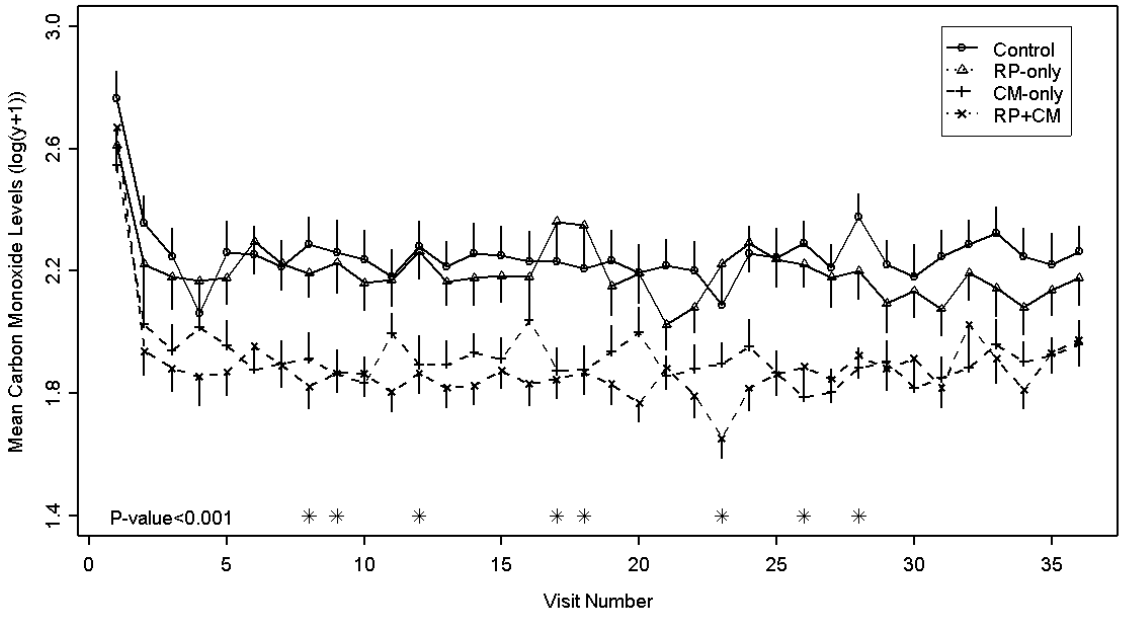
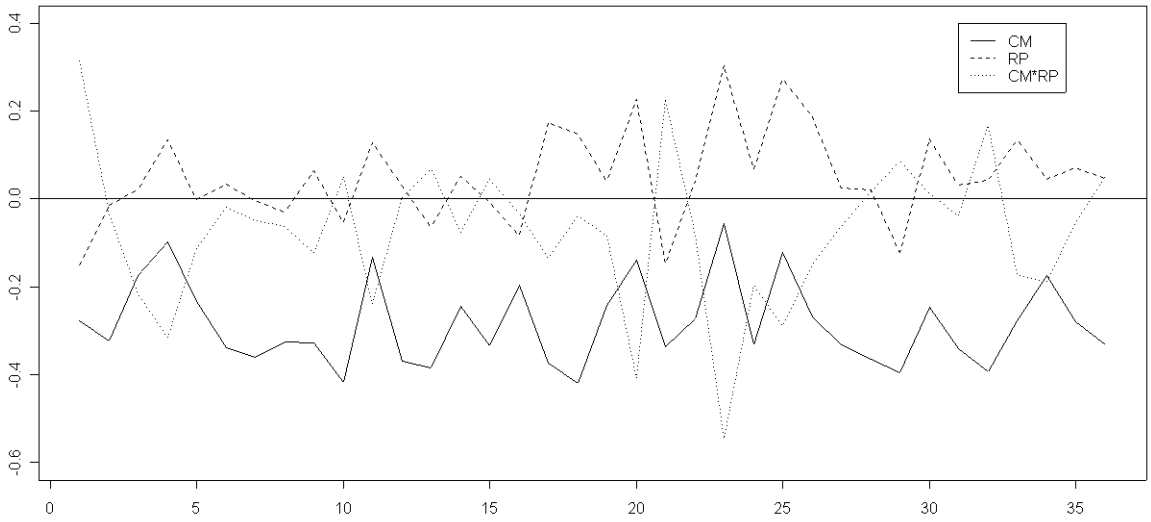


Figure 2. Average and SD curves for the log-scaled carbon monoxide levels



Note: (1) The vertical bars indicate the estimated standard errors for the average carbon monoxide levels. (2) The stares over the x-axis indicates the time points where the p-value of the point-wise ANOVA is smaller than 0.001.

Figure 3. Estimated Regression Coefficient Functions in the Functional Regression Analysis



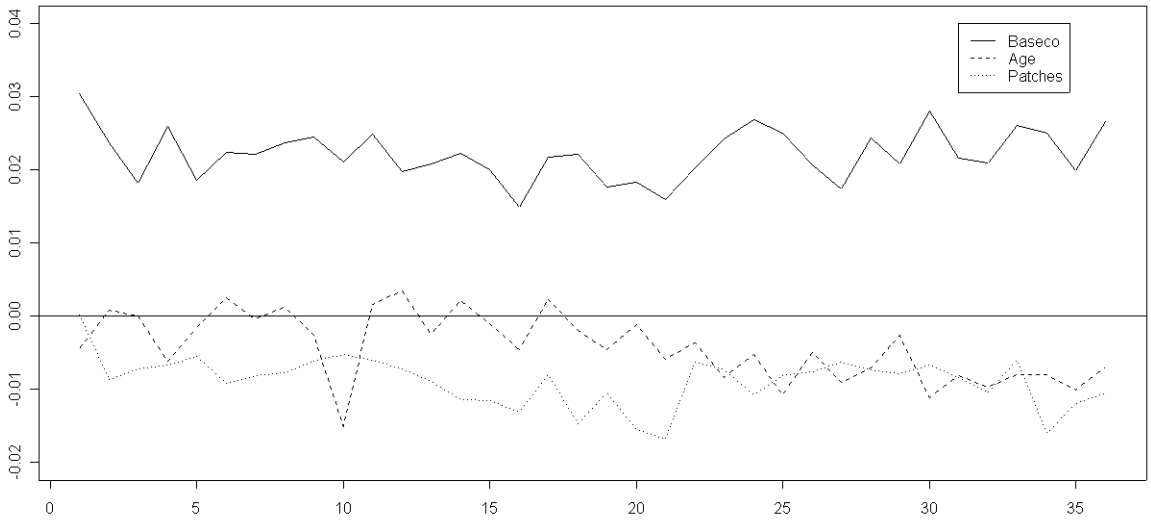


Figure 4. Diagnostics for the First Imputed Data Set