Statistics Surveys
Vol. 13 (2019) 150–180
ISSN: 1935-7516
https://doi.org/10.1214/19-SS126

Scalar-on-function regression for predicting distal outcomes from intensively gathered longitudinal data: Interpretability for applied scientists

John J. Dziak

The Methodology Center
The Pennsylvania State University, University Park, PA
e-mail: dziakj1@gmail.com

Donna L. Coffman

Department of Epidemiology and Biostatistics
College of Public Health
Temple University, Philadelphia, PA
e-mail: dcoffman@temple.edu

Matthew Reimherr

Department of Statistics
The Pennsylvania State University, University Park, PA
e-mail: mreimherr@psu.edu

Justin Petrovich

Department of Business Administration
St. Vincent College, Latrobe, PA
e-mail: justin.petrovich@stvincent.edu

Runze Li

Department of Statistics and The Methodology Center
The Pennsylvania State University, University Park, PA
e-mail: runzeli@psu.edu

Saul Shiffman

Department of Psychology
University of Pennsylvania, Pittsburgh, PA
e-mail: shiffman@pinneyassociates.com

Mariya P. Shiyko

Joyful and Creative Living, Boston, MA
e-mail: mshiyko@gmail.com

Abstract: Researchers are sometimes interested in predicting a distal or external outcome (such as smoking cessation at follow-up) from the trajectory of an intensively recorded longitudinal variable (such as urge to smoke).

This can be done in a semiparametric way via scalar-on-function regression. However, the resulting fitted coefficient regression function requires special care for correct interpretation, as it represents the joint relationship of time points to the outcome, rather than a marginal or cross-sectional relationship. We provide practical guidelines, based on experience with scientific applications, for helping practitioners interpret their results and illustrate these ideas using data from a smoking cessation study.

MSC 2010 subject classifications: Primary 62-02; secondary 62M10, 62G08.

Keywords and phrases: Distal outcomes, functional regression, intensive longitudinal data, scalar-on-function regression, trajectories.

Received November 2018.

Contents

1. Introduction

Modern longitudinal studies in medical and behavioral sciences often make use of advanced technologies, such as activity monitors and smartphones, to capture variables intensively over time in real life settings [4, 76, 91]. It is sometimes of scientific importance to study whether intensively-measured covariates can inform predictions of later outcomes of interest such as adverse events or diagnoses [31]. For example, in people who are attempting to quit smoking, the trajectory of self-reported withdrawal symptoms, such as urge to smoke and negative affect, has been shown to predict relapse risk [27, 58]. Another study [66] found that the pattern of change in a fetus's heartbeat, measured every 0.2 seconds over one minute after a vibroacoustic stimulation, predicted whether the mother would experience a high-risk birth. Similarly, participants' response patterns over time in a grip strength test may distinguish between different neurological disorders [53]. Longitudinal data collected over months or years can also be used for predicting a later outcome. For example, [26] studied the relation between weight gain during the first months of infancy and later neuropsychological performance in childhood, and [29] studied prediction of later height from children's early height trajectory.

In many of these contexts, an investigator wishes to predict a scalar variable Y_i, for each participant or other sampling unit, from the a trajectory, $X_i(t)$, of a longitudinally measured variable. The simplest way to do this would be to use a data reduction strategy. Specifically, one could summarize $X_i(t)$ as one or more univariate statistics to be used as regression predictors. For example, if $X_i(t)$ is numerical then the researcher could predict Y_i in a regression using the average $\bar{X}_i$ [e.g. 34], the fitted linear slope for each participant in X_i [e.g. 46, 89, 90], and perhaps also the variance or volatility in X_i [e.g. 10]. If X_i is binary, the researcher could use the count of events [e.g. 75]. However, each of these methods imposes a parametric form on the predictive relationship and discards some information on $X_i(t)$. A more general option is scalar-on-function regression, described by [19], [50], [62], [63], and [67], to name only a few. This is a form of functional regression, which allows a nonparametric relationship between a longitudinal function and an outcome. Other forms of functional regression exist, such as function-on-scalar, concurrent function-on-function, and lagged function-on-function, but they are beyond the scope of this paper [see 63].

1.1. Scalar-on-function regression

Assuming, as is convention, that the time interval is rescaled to be between 0 and 1, this model can be written as

$$g(E(Y_i|X_i)) = \theta_0 + \int_0^1 \theta_1(t)X_i(t)dt \tag{1.1}$$

where $g()$ is a link function such as the identity for linear regression or the logit for logistic regression [48]. Here, $X_i(t)$ is viewed as a smooth function evolving

over time, taking on some scalar value, X_{ij}, at time t_{ij}, though the index variable t need not always be time. In particular, scalar-on-function regression is also useful for spatial and imaging data, including neuroimaging [e.g. 20, 21]. Note that scalar predictors and additional functional predictors can also be included in the model [19], but in this paper we focus on the case of a single functional predictor $X_i(t)$ for simplicity.

It can be helpful to view Model (1.1) as a limit of the approximation

$$g(E(Y_i|\mathbf{X}_i)) = \theta_0 + \sum_{j=1}^{J} \theta_j X_{ij} \tag{1.2}$$

for a large number J of equally spaced measurement times [see 19, p. 832]. The smoothness assumption translates into constraining differences between consecutive θ_j to be small in order to manage collinearity and overfitting. Intuitively, a model such as (1.1) or (1.2) is extremely flexible and therefore able to represent many different possible predictive relationships. This makes it an attractive choice for researchers. However, there have been at least two challenges that have prevented it from being of more use in practice: accessibility and interpretability.

1.2. Challenges in scalar-on-function regression

In the past, software availability has been limited when fitting models such as Model (1.1) in cases when measurement times are irregular and vary across units. Software such as the `fda` R package [64] and older versions of the `refund` R package [12] required data to be specified as a matrix where columns are measurement times. This was a problem for many studies, such as ecological momentary assessment [76], which involve electronic measurements taken at random times. However, this challenge has been assuaged with the relatively recent release of two R packages that allow measurement times to differ for each participant: `funreg` [14] and newer versions of `refund` [e.g. 22]. The method used by `funreg`, which closely follows [19], is reviewed in the Appendix.

The remaining difficulty thus involves interpretation. The coefficient function, $\theta_1(t)$, does not have the same interpretation as an ordinary regression coefficient, even when only examining a single time point. Therefore, it is easy to make misleading conclusions about the interpretation of $\theta_1(t_0)$ at a given value, t_0. In particular, $\theta_1(t_0)$ does not express the marginal relationship between $X(t_0)$ and Y in Model (1.1). That is, one *cannot* say, as in simple linear regression, that the prediction $g(E(Y_i))$ is $\theta(t)$ units higher for a participant with $X_i(t) = x$ versus a participant with $X_i(t) = x - 1$. Such an interpretation would confuse a *joint relationship* for a *marginal relationship*, which can yield misleading conclusions [65]. To see this, notice that θ_j in Model (1.2) expresses the estimated relationship of X_j to Y while adjusting for the other measurements $X_{j'}$, *i.e.*, the contribution of X_j to Y holding the other X covariates constant. However, because the measurements are likely to be highly correlated within an individual, there is no reason why θ_j should have the same value,

or even the same sign, as the regression coefficient of Y on X_j alone. Furthermore, in Model (1.1), $\theta_1(t)$ and $X(t)$ are treated as smooth functions of time, so it is not even meaningful to speak of varying $X(t_0)$ while holding all other $X(t)$ constant. It would be more accurate to say that $\theta_1(t_0)$ tells approximately how many y-units $g(E(Y))$ would increase when $X_i(t)$ increased by an average of one x-unit within in a small neighborhood of time centered at t_0, holding other $X_i(t)$ roughly constant outside this neighborhood. However, it is not clear how this rather abstract explanation should be applied to answer a researcher's substantive questions.

In the fetal pulse example, [66] estimated $\theta_1(t)$ to look roughly like the negative of a cosine function, first becoming negative, then gradually positive, then returning to around zero over the course of the minute. They found that the fitted model had predictive or diagnostic value (i.e., $\theta_1(t)$ was not identically zero for all t). However, they did not attempt to interpret the specific clinical implications of the negative, positive, and near-zero time intervals.

James [30] sought to predict the five-year life expectancies (Y_i) of cirrhosis patients from their bilirubin levels (X_{ij}) over a period of 800 days, taking censoring into account in a way not described here. An increasing trajectory of bilirubin levels is thought to be indicative of disease severity and to predict liver failure [43, 73]. James found that the weight function $\theta_1(t)$ had a negative relationship with five-year survival for the beginning 200 and final 200 days of the interval of interest, reflecting the expected behavior of bilirubin as a marker of illness. However, paradoxically, he found that it seemed to have a weak positive relationship with five-year survival between around days 200 to 600. There was no clear reason why the meaning of bilirubin as a diagnostic indicator would fluctuate over time in this way. However, James cautioned that the plot "must be interpreted carefully" (p. 13), because those who have high bilirubin at one time will also have high bilirubin at other times. In this sense, perhaps the positive coefficient in the middle of the interval really meant something like *bilirubin is always a disease indicator, but is a less severe indicator in the middle of the interval than at other times.* It might even mean *bilirubin is always a disease indicator, but the worst situation is when bilirubin increases from the middle to the end of the interval.* A naïve interpretation such as *high bilirubin is associated with less disease severity towards the middle of the interval, but more disease severity at the beginning or end* could be very misleading.

1.3. Approach of this paper

This goal of this paper is to help researchers to more easily and insightfully interpret scalar-on-function regression models. We explore heuristics and potential pitfalls for interpretation of scalar-on-function regression models, and also describe alternative models that may be more useful in specific situations. We begin by briefly reviewing the notation and implementation of the scalar-on-function regression model and illustrating its use on a smoking cessation dataset. We then explain possible approaches for making $\theta_1(t)$ easier to interpret. After

this, we illustrate these issues within an important special case, namely the attempts to find a value of t for which $X(t)$ is considered maximally influential in predicting Y, and show why this can also be misleading if the question is not defined carefully. We additionally make recommendations for possible future research.

2. Notation and implementation

We assume a generalized linear model of the form in (1.1). Scalar predictor variables $S_{i1}, ..., S_{iL}$ representing subject-level non-time-varying covariates, such as gender or ethnicity, can additionally readily be included along with the scalar intercept θ_0, but for simplicity we do not address this here. The coefficient function, $\theta_1(t)$, is not assumed to have a known parametric form, and is instead estimated nonparametrically. We do assume here that the link function, g, is set to a known link function (e.g., logit); for a discussion of the more complicated case in which the form of g must also be nonparametrically estimated, see [50].

Although the the underlying trajectories, $X_i(t)$, are assumed to be continuous functions, practically they are observed at only a finite number of points and potentially with measurement error. Therefore, in practice the model is fit by first fitting smooth nonparametric functions $\hat{X}_i(t)$, and then using these as data to estimate the $\theta_1(t)$ function. Let $\hat{X}_i(t)$ at time t_{ij} be a smooth prediction of $X_i(t)$, rather than setting it to the observed X_{ij} itself [19, 92]. Thus, the actual fitted model is approximated as

$$g(E(Y_i|X_i) \approx \theta_0 + \int_0^1 \theta_1(t)\hat{X}_i(t)dt. \tag{2.1}$$

This changes the meaning of the model slightly, because $X_i(t)$ and $\hat{X}_i(t)$ are not technically the same variable, but for simplicity we do not address this distinction here. Goldsmith and co-authors [19] and Wang and co-authors [92] describe the relationship of smoothing to inference in functional data analysis further. We also do not consider the issue of registration (warping): roughly speaking, techniques for making sure that time points have comparable meanings across participants. This is very important for interpretability but is discussed elsewhere [see, e.g. 63, 95].

To fit Model (2.1), we approximate both $\theta_1(t)$ and the $\hat{X}_i(t)$ in terms of an appropriate finite basis. In particular, we use a functional principal components expansion [also known as Karhunen-Loève decomposition; 3] for the $X_i(t)$, and we use a penalized B-spline for $\hat{\theta}_1(t)$ [15] with the richness of each basis controlled by some model selection or tuning criterion. Thus, the problem is translated into a manageable parametric multiple linear regression problem; see [19], [30], [63]. Either a very small set of basis functions or a roughness penalty function, or both, is used to control the complexity of $\theta(t)$ and to control its sampling variance [16, 19, 32, 63].

3. Empirical illustration

Ecological momentary assessment (EMA) is an innovative and increasingly important data collection strategy which provides intensive longitudinal data relevant to studying human experiences and problems in daily life [see 84]. One prominent use of EMA has been in monitoring the everyday experiences of participants suffering from addiction or trying to quit a harmful substance [e.g. 58, 74, 90]. In particular, tobacco smoking is known to be one of the largest preventable causes of death and of disability in the world [18, 25, 33], but overcoming nicotine dependence in order to quit smoking can be extremely difficult [33, 59]. Therefore, research into the causes of relapse and its prevention is very important [59].

In this section, we present a re-analysis of a real dataset from an EMA study on smoking cessation, previously analyzed elsewhere [13, 74, 77, 78, 79, 82]. Details of the study are described by Shiffman and co-workers [74] but are briefly summarized here. Analyses are done in R [60].

3.1. Original study

The 304 participants in this study were adults highly motivated to quit smoking. They were given a palmtop computer used as an electronic diary, and were asked to record their smoking behavior, as well as being given random prompts about five times daily to answer questions about their emotions and level of urge to smoke. They were asked to quit smoking on a particular day, about two weeks into the observational study, and they continued to record observations for about a month after this quit date. Past analyses focused on describing the context and antecedents of lapses into smoking, the risk factors for relapses (serious lapses, defined as smoking at least 5 cigarettes for 3 consecutive days), and the natural course of changes in urge to smoke over time. To explore the potential for scalar-on-function regression, we reanalyze the data with a new question: Among those who are able to avoid relapse for at least a week, can the trajectory of urge to smoke during the first week after quitting (the third week of the study overall), be used to predict their likelihood of continuing to avoid relapse for the rest of the month? In other words, can the trajectory of urge to smoke be used to give an early assessment of which participants are in high danger of relapse and which participants are doing well?

3.2. Analysis subsample

The subsample used for this analysis was 235 participants who provided data points before, during, and after the first post-quit week and who did not relapse during the first post-quit week. These criteria were set in an attempt to make sure that enough data would be available per subject to fit a scalar-on-function regression model, and that informative dropout would be avoided by focusing on those who were relapse-free for the first post-quit week. Only data

resulting from random prompts was analyzed, ignoring additional event-driven prompts described elsewhere [e.g. 78]. Also, within a subject, not all prompts were answered and only occasions in which urge was reported are considered here. The total number of usable observations per subject during the first post-quit week ranged from 2 to 53, with an average of 27. Among these 235 analyzed participants, 31 had relapsed after the first four weeks and 204 had not.

3.3. Predictor variable

The predictor variable $X_i(t)$ used in this analysis was self-rated urge to smoke, rated on a scale from 0 to 10 [78] and treated as a numerical variable. Feelings of craving or urge have been found to be important predictors of relapse [see, e.g. 10, 47]. $X_i(t)$ had a mean of 2.76, standard deviation of 2.98, and median of 2, averaging across times. However, it ranged over the entire scale available (0 to 10) with first and third quartiles at 0 and 5, suggesting that an average participant's urge was usually fairly low but with some occasions of high urge.

3.4. Baseline covariates

We also considered three subject-level baseline covariates: age (22 to 67 with a mean of 44), sex (101 male, 134 female), and baseline nicotine dependence. The latter was operationally measured using the continuous form of the Heaviness of Smoking Index [HSI; 5], based on the earlier categorical Heaviness of Smoking Index [24, 39]. It is calculated as a function of self-reported cigarettes smoked in a typical day (more indicating greater dependence) and of minutes until first cigarette on a typical morning (fewer indicating greater dependence). Participants reported 7 to 90 cigarettes per day (mean of 26), and first cigarette between 1 and 180 minutes after waking (mean of 17; note that the skewed distribution is taken into account in the formula for HSI.) This corresponded to an HSI index from 4.66 to 15.49 with a median of 8.77 and a mean of 8.83. The original HSI score had been a 0 to 6 categorical scale; on this scale the participants ranged from 0 to 6, with the median corresponding to 4. The numerical and categorical HSI scores were correlated at $r = .92$; the numerical version is used here.

3.5. Initial descriptive analyses

For each participant, a mean self-rated urge and a least-squares slope in self-rated urge were calculated. The mean self-rated urge $\bar{X}_i$ was correlated at $r = 0.126$ ($p = .055$) with binary relapse status, and the least-squares slope in self-rated urge for each subject was correlated at $r = 0.187$ ($p = .004$) with binary relapse status. These analyses suggest that participants whose self-rated urge was high near the end of the first post-quit week were at higher risk of relapse.

A parametric logistic regression was fit to predict relapse from mean urge, least-squares slope in urge, and the baseline covariates. Results are shown in

TABLE 1
Parametric logistic regression results from smoking cessation example

	Coefficient	Standard Error	z	p
Intercept	−3.85	1.29	−2.98	0.003
Sex (male=1)	0.37	0.42	0.88	0.378
Age (years)	−0.02	0.02	−1.02	0.306
Numerical HIS	0.31	0.12	2.69	0.007
Mean urge	0.11	0.09	1.19	0.233
Slope in urge	1.60	0.56	2.88	0.004

Table 1. Higher levels of initial dependence, and higher slope, each predicted higher probability of relapse.

For illustrative purposes, the mean trajectories of urge for the 31 who relapsed and for the 204 who did not, are shown in Figure 1. Note that Figure 1 is not directly applicable either to forecasting or to causal explanation, because it involves predicting the past from the future, but it is nonetheless helpful in understanding how relapsers and non-relapsers differ. The upper left pane shows the fit for each subsample using the R function `loess`, accepting defaults (local quadratic with span 0.75). However, the confidence intervals are likely to be too narrow, due to ignoring within-person correlation, as well as implicitly conditioning on the choice of the tuning parameter. As an alternative, the upper right pane shows a generalized estimating equations (GEE) fit of urge on time and time squared for each subsample as a whole, using working independence with a robust standard error to account for within-person correlation [42]. Dotted lines indicate 95% pointwise confidence intervals. The lower left and lower right panes show within-day means and confidence intervals, for a local constant GEE, using working independence GEE again, for non-relapsers and relapsers respectively. These confidence intervals do not correct for multiple comparisons, which would have been difficult given the current sample size at the subject level. Similarly, as is usual in nonparametric regression, the 95% pointwise confidence interval for a function at a particular value of t is much narrower than the interval which would be required to contain the function for the whole interval with 95% familywise confidence. Therefore, it is not possible to make statements about which days are statistically significantly different from other days. However, although the details of each plot differ, and the smaller number of relapsers leads to wide confidence intervals when correcting for within-subject correlation, each plot approach suggests a steady mean decline in urge for non-relapsers and much less of a decline for relapsers. That is, participants who would eventually be successful seem to enjoy a steady reduction in urge, while those who would eventually relapse seem to decline briefly and then reach a plateau [see 82].

3.6. Scalar-on-function regression

We fit a binary functional logistic regression in which outcome variable Y_i was relapse status at the end of the study, about four weeks after quit date. We adjusted for the baseline covariates of age, gender, and assessed smoking dependence. As mentioned earlier, of the 235 participants considered in the subsample,

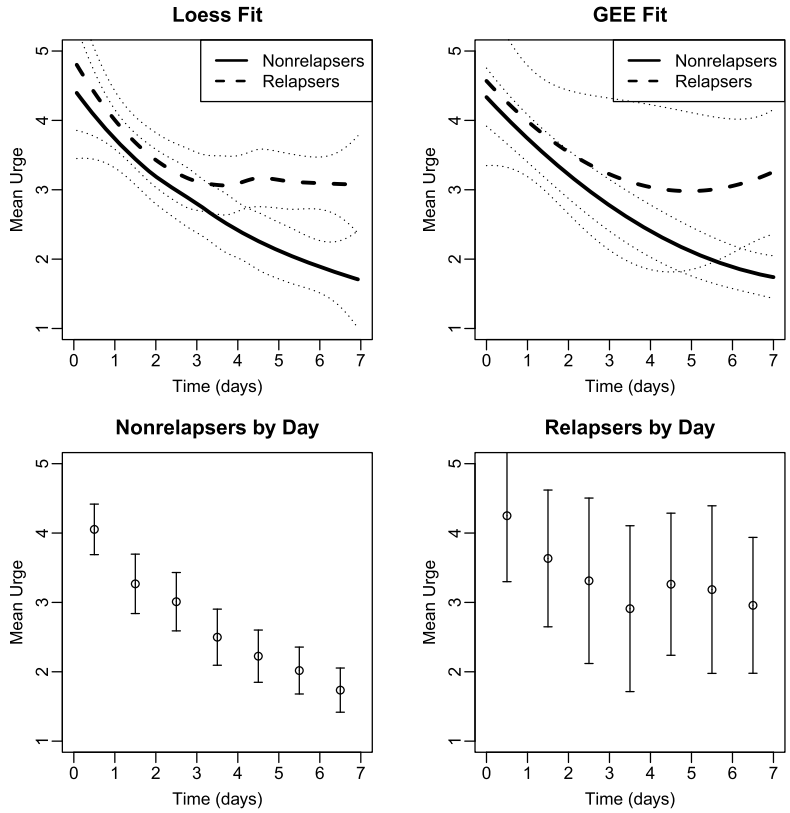

FIG 1. *Smoothed Mean Trajectories of Later Nonrelapsers and Later Relapsers Over Time.*

31 (13%) had relapsed and 204 had not. Because no one in the subsample relapsed during the first post-quit week, the relapses ranged from day 8 to day 24 after quit date. The length of time until relapse was not used in this illustrative analysis, although a survival-outcome scalar-on-function regression would have been possible.

Therefore, the model is

$$g(E(Y_i|X_i)) = \theta_0 + \theta_1 S_{1i} + \theta_2 S_{2i} + \theta_3 S_{3i} + \int_0^7 \theta_X(t) X_i(t) dt \qquad (3.1)$$

where $g()$ is the logit link, Y_i is relapse after four weeks, X_i is the trajectory of urge over the first weeks, and S_{1i}, S_{2i} and S_{3i} are sex (dummy-coded 1 for male), age, and HSI. $\theta_X(t)$ here is the coefficient function of interest, analogous to $\theta_1(t)$ in Model (1.1). As a caveat, possible interactions or nonlinear effects of the covariates were not considered because of the relatively small sample size of responders at the subject level, but they cannot be ruled out as a possibility.

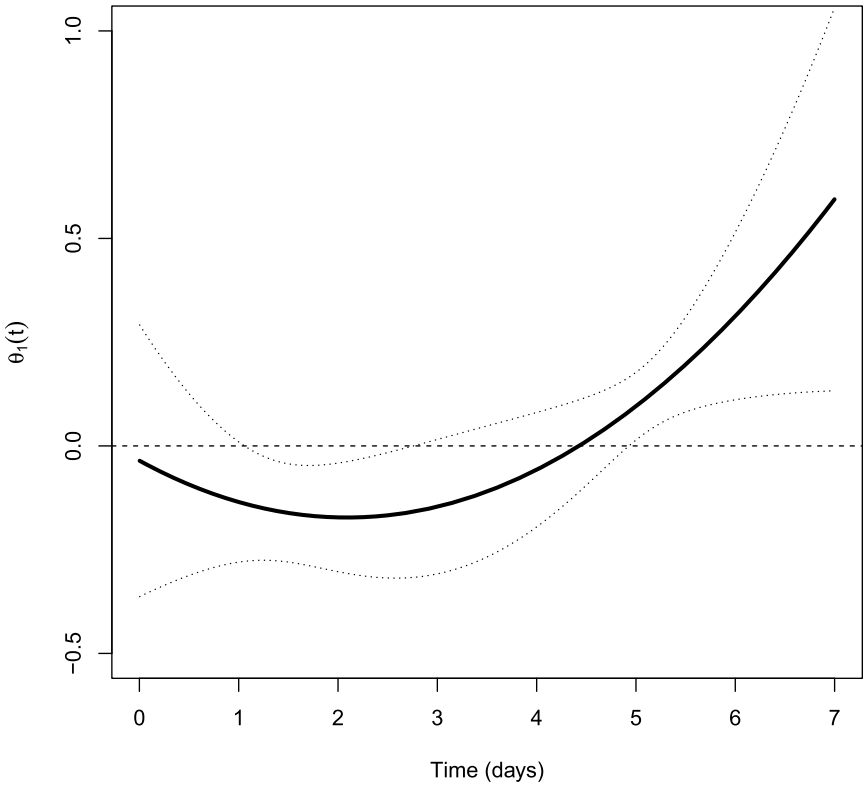

FIG 2. *Coefficient Function in Functional Logistic Regression Model for Predicting Relapse.*

The `funreg` package [14] was used for computation; the Appendix gives technical details on this package. The estimated coefficient for the intercept was −3.76. The estimated coefficients for male sex, age, and numerical HSI were approximately 0.40 (standard error 42), −0.02 (standard error .02), and 0.31 (standard error .12). Sex and age were not significant predictors, but higher HSI predicted higher probability of relapse ($p = 0.008$). The fitted coefficient function $\theta_X(t)$ for Model (1.1), with 95% pointwise confidence intervals, is shown in Figure 2. The pointwise confidence intervals overlap with zero from day 0 through 5, but the function is pointwise significant and positive after day 5. This seems to agree with the other analyses reported in that high self-rated urge during the end of the first post-quit week predicted higher later risk of relapse.

3.7. Significance test

A simple permutation test was used to provide a familywise bootstrap p-value for a global test of whether $\theta_X(t)$ was identically zero. Using 5000 permutations, the p-value was significant at $p = 0.0176$ if baseline covariates are not adjusted for,

or $p = 0.0092$ adjusting for baseline covariates. This p-value is slightly higher than the parametric model, although still significant. This may be because a nonparametric model can have lower power and precision in small samples than a well-chosen parametric model, a classic bias-variance tradeoff. The current sample is large in terms of total number of observations, but rather small when considered as a logistic regression at the subject level (only 31 relapsers). In some recent applications of intensive longitudinal data with extremely large datasets from the general population, statistical power limitations might not be an issue, and insuring practical significance and interpretability would become more important [28, 56]. Nonetheless, for studies such as this one involving specific groups of volunteers (e.g., people in the first days of a smoking cessation attempt), sample size limitations will still exist.

3.8. Interpretation of the coefficient function

More importantly, the interpretation of Figure 2 involves a similar dilemma in interpretation as the bilirubin data discussed earlier. Notice that in Figure 2 the model places a nonsignificant but seemingly negative weight θ_X on urge from days 0 to 5 and a then a positive weight on urge after day 5. How should this be interpreted? If the coefficient function at a given time is misinterpreted as a marginal relationship between urge at that time and relapse, a researcher might conclude that a high urge to smoke was not problematic, or was perhaps even beneficial, until after day 5. However, there is no obvious reason why the sign of the relationship between urge and relapse risk would change so abruptly on that or any other specific day. Figure 3 shows the actual marginal correlations. The dotted lines are the marginal correlation of the smoothed $\hat{X}_i(t)$ values with Y_i at each given t, with approximate 95% pointwise confidence intervals. The dots with error bars are correlations of the within-day mean data for each subject with that subject's relapse status. It can be seen that the estimated relationship between $\hat{X}_i$ and Y_i is always positive, as one would expect, although it is initially not statistically significant.

Thus, Figure 2 can only be correctly interpreted by considering the entire curve as a single unit. The fitted coefficient function puts positive weight on the later days and negative weight on the earlier days, not because urge goes from being a favorable to an unfavorable symptom, but because the slope of urge over time (i.e., its decrease or lack thereof) is the important predictive feature. This is reasonable in light of the descriptive analysis and Figure 1. Some analogous past studies using parametric models have also found the slope of a trajectory to be an important predictor. For example, [94] found that an increase of pain symptoms over time was an important predictor of relapse in individuals trying to recover from prescription opioid dependence.

At first, the correlations in Figure 3 seem small. However, it is reasonable for the correlations to be modest because urge at a single given time may be less predictive than the average change in urge over time. One reason is that urge at a specific time is likely to have high noise variability, both because urge

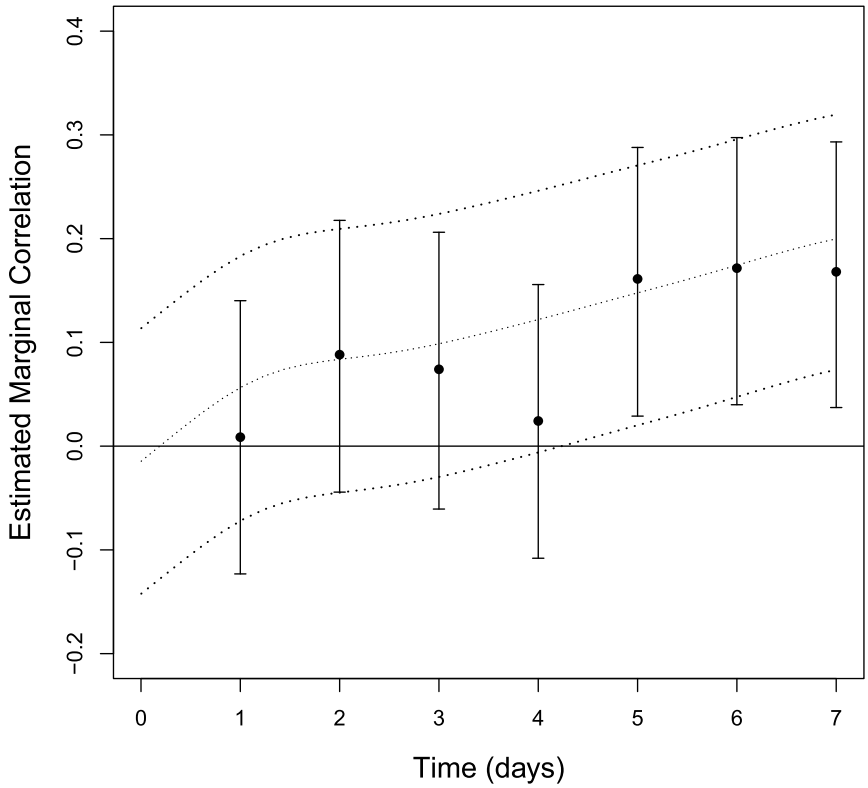

FIG 3. *Subject-level Correlation of Urge with Relapse over Time for Binned Means and for Smoothed Trajectories.*

changes rapidly over time and because different participants are in different environments at that exact time. Another reason is that there is no way to correct for subjective differences between participants in what subjective feeling they choose to describe as an urge of 1, 2, etc. Therefore, for best predicting relapse, perhaps it is necessary to consider all times together as in Figure 2. However, if it is desired to be able to interpret the predictive relationship of urge at a given day to later outcome, then Figure 3 is useful and Figure 2 is misleading. Thus, both the joint and marginal approach are useful, but for different questions, and a better understanding is obtained by considering them all instead of just one.

3.9. Other possibilities for analysis

As a caveat, note that attrition during the time period of interest was not a concern for this example because only those participants who avoided relapse for the first week were considered. If the goal was to predict relapse during the same time period as urge was being monitored, a researcher might use a joint model

of longitudinal process with a survival process [e.g. 55, 68, 97] or a landmarking approach [e.g. 88] but these are beyond the scope of this paper. Also, because all participants had a set quit date, treated as day zero in this analysis, time was already considered to be on a common scale for each participant and no special curve registration or warping method [63, 95] was considered necessary.

The analysis of the original study included not only "relapse" events but also "lapse" events, in which a participant uses one or a few cigarettes but does not return to daily smoking. A richer model could have included these events, perhaps using the presence or absence of an early lapse as an additional covariate for predicting ultimate relapse, or expanding the outcome variable to include three levels (full abstinence, some lapse, actual relapse) instead of only two. However, in this illustrative example we focus on predicting relapse only, as that is the more severe outcome.

If the $X_i(t)$ functions being used for prediction are all small variations upon a known basic shape, then it is possible to extract interpretable features using substantive knowledge rather than use a nonparametric approach. For example, [53] compared certain specified features of intensively observed curves, such as rate of increase or decrease during a particular phase of a task, as predictor variables. Describing group differences in terms of these features, instead of a nonparametric regression function, was advantageous because their meaning can be easily understood and interpreted by clinicians. Of course, in order for this approach to work, it is necessary for the $X_i(t)$ curves to be fairly regular and to be comparable enough across subjects for these features to have a common meaning. When making decisions based on features extracted from functions, it is important that the features be both predictive and interpretable, and how to find such features is an important topic of ongoing research [see 41, 70]. Sometimes, even from only a few measurement points, many features of substantive interest can be derived [36]. As a general rule, it is reasonable to conjecture that that the nonparametric approach is not as helpful as having a parametric model which is well-grounded in good theory, but more helpful than having a poorly informed parametric model.

4. Approaches towards improving interpretability

The challenge of making even nonparametric scalar-on-function regression more interpretable has received some attention, although not yet enough.

4.1. The FLiRTI method

In particular, [32] suggested the "Functional Linear Regression that's Interpretable" (FLiRTI) approach, an implementation of Model (1.1) which uses an approach like the LASSO [86] to bias the estimate of $\theta_1(t)$ towards a piecewise linear or quadratic function that is easier to describe. In the simplest version of FLiRTI, the time interval is divided into many subintervals, on many of which the $\theta_1(t)$ function is assumed to be zero, and model selection techniques are

used to select which values of t have nonzero $\theta_1(t)$, so that the predictor can be envisioned as influencing the response only during those time periods. More generally, instead of selecting regions with $\theta_1(t) \neq 0$, they select regions where $\theta_1(t)$ has a nonzero dth derivative for some $d \geq 0$; for instance, $\theta_1(t)$ may be piecewise linear. They demonstrate that this costs very little in terms of predictive ability. More recently, [72] also describe a somewhat similar way of using a LASSO-like penalty for functional regression. However, merely simplifying the shape of $\theta_1(t)$ does not resolve the problem of distinguishing joint from marginal relationships.

4.2. Eigenfunction decomposition

James [30] considered another approach to analyzing model (1.1). Instead of reconstructing and plotting $\theta_1(t)$, one could plot the most important eigenfunctions of $X_i(t)$ and report the regression coefficient of each one in predicting y_i. This might provide a description of just what features in the data may be associated with higher or lower y, which can provide helpful insights [see 98]. However, many researchers will find this approach rather unfamiliar and may have some difficulties in interpreting the results. Furthermore, there is no guarantee that the eigenfunctions of $X_i(t)$ are the patterns which are most important and interpretable for explaining variability in Y_i.

4.3. Heuristic rules

In an attempt to find more broadly applicable rules for interpreting the results of scalar-on-function regression, it may be helpful to first consider some rather abstract special cases. The most extreme possible case is one which that $X_i(t)$ values are constant over time within each subject i. Then the regression coefficients from the multiple regression Model (1.2) cannot be uniquely estimated because of perfect collinearity. The same intuition holds for Model (1.1); if $X_i(t)$ is constant in t then it can be factored out of the integral $\int_0^1 \theta_1(t)X_i(t)dt$ and the fitted values will be entirely determined by the integral of $\theta_1(t)$ rather than its shape or value at any point. In such a situation, the marginal effects would be the same at all t, and the joint effect will be unidentifiable.

In contrast, now suppose that $X_i(t)$ is not constant in t, but that the θ_j coefficients in Model (1.2) are all equal. Thus, $E(Y_i)$ is determined by the sum of the X_{ij}, or equivalently, after rescaling, by their mean. The equivalent in terms of Model (1.1) would be to have $\theta_1(t)$ equal to a constant κ for all t. This is because under such an assumption,

$$\begin{aligned} g(E(Y_i)) &= \theta_0 + \int_0^1 \theta_1 X_i(t)dt \\ &= \theta_0 + \kappa \int_0^1 X_i(t)dt \\ &= \theta_0 + \kappa \bar{X}_i \end{aligned}$$

where $\bar{X}_i$ is the average value of $X_i(t)$. Just as in classic regression, $\bar{X}_i$ and Y_i are then positively correlated, negatively correlated or uncorrelated if θ_1 is positive, negative, or zero, respectively. Here $\theta_1(t) = \kappa$ is a special parametric case within the nonparametric scalar-on-function regression model (1.1) Thus, we already have three heuristic rules for interpreting scalar-on-function regression:

1. If the plot of $\theta_1(t)$ oscillates wildly between extremely high positive and negative values, with a very wide confidence interval, either a stronger model restriction is needed (stronger penalty function or more limited basis functions) or the X_{ij} may be too highly correlated within subject.
2. If the plot of $\theta_1(t)$ resembles a flat line with zero intercept and zero slope, this is consistent with a model in which Y_i is unrelated to X_{ij}.
3. If the plot of $\theta_1(t)$ resembles a flat line with a nonzero intercept and zero slope, this is consistent with a model in which Y_i is predicted by the overall level of X_i but not by the direction of changes in X_{ij}, and this relationship is positive or negative depending on whether the intercept is positive or negative.

Consideration of some other simple scenarios will provide more of these rules. In Model (1.2), instead of supposing the θ_j to be zero, now suppose that they are equal to the coefficients of some meaningful linear contrast in X. For example, suppose that $\theta_j = t_j - t_{\text{mid}}$ where t_{mid} is the middle of the interval of interest (e.g., 0.5 if time is scaled from 0 to 1). Then this model is equivalent to saying that the within-subject regression slope of the x_i determines $E(y)$. The large-J equivalent of this model is then Model (1.1) with $\theta_1(t) = \kappa(t - t_{\text{mid}})$ for some constant κ. This suggests another simple case.

4. If the plot of $\theta_1(t)$ resembles a straight line with nonzero slope and which is zero near the middle time point, this is compatible with a model in which y_i is predicted by the within-subject slope of x_{ij} but not its overall mean.

It would seem rather unusual for the mean of the x_{ij} to have no predictive value at all if the within-subject slope had predictive value. However, one can obtain a more complex rule by adding the coefficients together. Adding $\kappa_1 + \kappa_2 \int x_i(t)dt$, in which the mean of x is the predictor of interest, to $\kappa_3 + \kappa_4 \int (t - t_{\text{mid}})x_i(t)dt$, in which the slope of x is the predictor of interest, we get $\kappa_5 + \int (\kappa_6 + \kappa_7 t)\, x_i(t)dt$. Thus, we have another rule of thumb:

5. If the plot of $\theta_1(t)$ resembles a straight line with nonzero slope, and does not pass through 0 midway through the interval, it is compatible with a model in which both the within-subject mean and the within-subject slope of x_{ij} are important predictors of y_i.

If one of these simpler cases appears roughly adequate, then an appropriate parametric model [46, 89] could be more useful and interpretable than a nonparametric curve. However, if no such case applies, then a richer model is necessary. Further research may suggest further heuristic rules for interpreting more complex situations. Alternatively, one might consider using one or more

derivatives of $x_i(t)$ as the predictor, instead of $x_i(t)$ itself. In this vein, even if the coefficient function is not a straight line, fairly simple oscillations can still be interpretable. For example, a straight line indicates the importance of the slope of the trajectory, while a parabola would point towards the importance of *acceleration*. Higher levels of oscillation can be interpreted similarly by linking them to higher order derivatives or more intricate contrasts, though one must be attentive to the risk of overfitting.

These rules for comparing nonparametric models to parametric models are only heuristic. It is reasonable to ask whether formal tests are available. Significance tests exist for whether a functional coefficient $\theta_1(t)$ is identically zero [8, 38, 50], and these could perhaps be adapted to test whether $\theta_1(t)$ is zero after accounting for a parametric effect. Alternatively, information criteria have been used to choose model complexity in functional regression [50] and the same idea could presumably be adapted to compare a spline model with a given number of knots to a particular parametric model. However, these ideas require more research before being generally recommended.

If a researcher's substantive questions are more about marginal than joint relationships between particular time points and the outcome, then a function-on-scalar regression approach [see 63, 81, 85, 96] may be more appropriate. To see why, consider that it is possible for a time-invariant variable to have a time-varying effect (i.e., to be the predictor in a regression with functional response). An example is biological sex, which has a different relationship with height at different stages of child growth. If one's question is about association rather than causation or future prediction, a researcher might try reversing the roles of the variables to perform a function-on-scalar regression using the scalar outcome as the predictor.

$$E(X_{ij}) = \beta_0(t_{ij}) + \beta_1(t_{ij})Y_i \tag{4.1}$$

Because this is not intended as a causal model, the essential difference between Model (1.1) on the one hand, and (4.1) on the other, is not which variable comes first in time. The difference is which kind of correlation is of most interest: joint relationship conditioning on other measurement times in (1.1), versus marginal relationship ignoring other measurement times in (4.1).

A researcher must think carefully about which approach addresses the question of interest. To explore this further, we consider a common question about longitudinal data, that of finding what time period is most important, whether in a predictive or causal sense. It will be seen that precise formulation of the question and model determines which methods can or can not be used.

5. Identifying a critical period

One possible motivation for scalar-on-function regression or similar techniques is in looking for a period of time during which $X(t)$ is somehow most strongly predictive of Y. This period may have special causal importance; for example, important developmental psychological studies have investigated critical periods in which an organism is especially influenced by its environment in developing

certain capacities or abilities [see 37]. Some researchers in developmental psychology use the idea of a "critical" or "sensitive" period to describe a time period in which the risk of a particular adverse outcome is particularly high for members of a vulnerable group [e.g. 17], or a developmental period in which environmental traumas (e.g., child abuse or separation from parents) have the largest harmful effect on a later health outcome [1, 2, 51, 57]. The term "sensitive" is often preferred to "critical" to emphasize that some adaptability and resilience can remain even after this period has passed [51]. Nonetheless, much public health research stresses the importance of early experiences and early interventions [e.g. 6, 11, 45, 51, 52]. Given a limited amount of intervention resources, intervening at the right time to prevent difficulties from developing may be much more effective than intervening to treat them later.

Note that the idea of a most important period could be of interest not only for describing developmental changes over a period of years, but also clinical changes over a shorter period of time. For example, in a smoking cessation study, [47] found many participants reported a jump in self-reported craving during the day of the quit attempt and that a larger jump predicted higher relapse risk.

Note that sensitive periods are assumed to occur at roughly the same time for each participant. They therefore differ from turning points (unexpected external events which dramatically change later trajectories; see [54]) and tipping points (abrupt changes due to an input reaching an unobserved threshold; see [9, 80]), although there may be some similarities.

It is sometimes of interest to find a period in which changes have the most causal influence, but other times it is instead of interest simply to identify a period which provides the best statistical measurement or prediction regardless of causation. Intensity of urge to smoke experienced during the first few waking minutes of a day might be a clearer indicator of addiction severity than the intensity of urge to smoke, say, just after lunch or just before dinner, if only because it is known that the individual has been abstaining for several hours in the first case, but may or may not have been abstaining in the other cases. The goal of finding either a causal or a predictive sensitive period seems to suggest scalar-on-function regression as a possible approach.

5.1. Challenges in locating sensitive periods

Although these critical or sensitive periods are substantively very important, locating them using functional regression would involve some challenges, especially because measurements are correlated over time. As a simple example, suppose that each subject's $X(t)$ was an autoregressive moving average process of order one as follows: $X_{i,t} = \gamma_0 + \gamma_1 X_{i,t-1} + e_t$, with independent normal errors e_t, and for simplicity consider only 5 discrete time points; note that the continuous time analog would be the Ornstein-Uhlenbeck process. Also for simplicity, assume $\gamma_1 \geq 0$. This process could be represented by a very simple path diagram: $X_1 \to X_2 \to X_3 \to X_4 \to X_5$. Let $y = X_5$ and suppose it is desired to

predict y from X_1, X_2, X_3, and X_4 in a model of the form (1.2). As the path diagram suggests, the X's are a Markov chain. That is, among the predictors, X_4 has the highest correlation with y, and, furthermore, in a model which contains X_4, the earlier X's have no predictive value. Thus, the true values for the coefficients in Model (1.2), if fit to this population, would be zero for X_1, X_2, and X_3, and positive only for X_4. In contrast, the true values for the marginal correlation coefficients of X_5 with any of X_1 through X_4, or of the coefficients in a model like (4.1), would all be positive, but they would start out small and increase over time. This would be the case even if the γ_1 coefficients differed at each time point (assuming $\gamma_t \geq 0$), so that some of the arrows in the path diagram represented stronger relationships than others. No matter how strong or weak X_2 is as a predictor of X_3, it can never be a stronger predictor of y than X_3 is, because the effect of X_2 is entirely mediated by X_3. Even if, say, time 2 represents a transition period of enormous importance so that the X_2 to X_3 coefficient is much larger than any other coefficient, the model specifies that an effect on X_5 of a disturbance at X_2 must always be fully mediated by X_3 and by X_4. That is, X_5 is independent of X_1 through X_3 conditionally upon X_4, and so $\text{Corr}(X_5, X_2) = \text{Corr}(X_5, X_4)\text{Corr}(X_4, X_2) \leq \text{Corr}(X_5, X_4)$.

Thus, the practical importance of X_2, such as for timing an intervention, cannot be found by examining marginal correlations with X_5. However, it would not necessarily be detected in a joint regression either, because such a regression conditions on all of the X's and does not explicitly model the relationship between them. Making the intuition of a most predictive time period rigorous would require a more careful definition (e.g., the largest autoregressive coefficient) and/or a richer model (e.g., the possibility for lagged causal effects). It may be that in a causal sense, one cigarette smoked at age 13 is a greater danger for causing addiction than one cigarette smoked at age 30. However, in predicting nicotine addiction at age 31, smoking level at age 30 is almost certain to be the stronger statistical predictor. This is not said to dismiss the concept of a sensitive period but to reiterate that a data-driven search for a sensitive period would depend heavily on the precise definition being used.

5.2. Implications for scalar-on-function or function-on-scalar regression

With these insights, we can imagine what would occur if either Model (1.1) or Model (4.1) were fit to a process with strong positive autocorrelation, in an attempt to predict a later value of that process. Scalar-on-function regression might well find that the coefficient function is close to zero except near the end of the interval, when it jumps to a positive value. The shape might not actually be quite that simple because of sampling and measurement error, error perhaps exacerbated by the autocorrelation. Still, it would very likely be generally increasing and not always noticeably positive at first. Function-on-scalar regression, or a smoothed marginal correlation plot like Figure 3, would probably find a gradually increasing positive curve. A researcher might therefore mistakenly

conclude that the most recent period was the most critical in a causal sense, even though it was essentially no different from any of the other periods. This suggests another heuristic rule to our list of rules for scalar-on-function regression:

6. If $\theta_1(t)$ tends to be increasing, or increases sharply near the end of the interval, and if y represents a measurement taken after the end of the interval, consider whether the finding may simply represent the fact that the most informative measurements in predicting the next step of an autocorrelated process are the most recent.

Another interpretation of this concept is that scalar-on-function regression will not be a useful explanatory (as opposed to merely predictive) model for a distal y, unless it is plausible that early values of x could have an effect on y that is not fully mediated by later values of x. Otherwise, the earlier x observations will no longer seem to have an effect in a joint model conditioning for the later observations, no matter their actual causal importance. This is essentially the same problem as would occur when conditioning on a consequence of the treatment variable when performing an analysis of covariance [see, e.g. 40].

5.3. Other possibilities

Apparently, functional regression is not always the best way to model a sensitive period. Perhaps a better way might be explicitly modeling the change in the covariate process over time, rather than conditioning on the entire process as in functional regression. For example, a parametric model of the changes in $X(t)$ over time is possible using a dynamical systems approach. An analysis in which the dynamic changes of craving for smoking are modeled over time is given by [87]; possibly this approach could be extended to predict a relapse outcome.

Other possibilities also exist. [35] provides a computational model for simulating the effect of early and recent events on later behavior. Causal modeling, whether frequentist or Bayesian, may also be useful [see, e.g. 49]. [44] assume that the predictive information in X can be summarized by its values at one or a few crucial time points, which are the same for every subject, and discuss how to estimate the location of this point in time (e.g., for timing medical checkups). Also, [83] describe a method for comparing the information available for making a decision based on predicting an outcome at different potential decision times, for purposes of deciding whether to change treatment plans. In any case, the most appropriate methodology must depend on the precise research question and goal.

6. Conclusion

In scalar-on-function regression, it is possible to find a coefficient function $\hat{\theta}_1(t)$ which provides a good nonparametric estimate of the best weighting function for using all of the observed data in a longitudinal trajectory to predict a distal

outcome. However, it can sometimes be difficult to interpret what this estimated coefficient function means to a substantive researcher. In a way, this is typical for any kind of regression model fitted to correlated data from a non-randomized experimental study: that is, just because a model provides good predictions does not mean that it provides causal or mechanistic insight. However, in scalar-on-function regression, there are many intercorrelated X measurements for each Y measurement, as well as a very rich model space, so the problem of interpretation is especially salient. Sometimes the nonparametric coefficient function estimated using penalized functional regression may suggest a simple parametric form. When this is the case, researchers might then choose to follow up their analysis by going on to fit a parametric model which might have an easier interpretation. However, even in these cases, the nonparametric approach can still be useful in an exploratory or diagnostic role. In the future, further methodological research may make it easier to interpret the coefficient function in scalar-on-function regression.

Acknowledgements

This project was supported by Award R03CA171809-01 funded by the National Cancer Institute, by Awards P50DA039838, R21DA024260, and R01DA039901 from the National Institute on Drug Abuse, and by the National Institutes of Health (NIH) grant 1R01CA229542-01 funded by the National Cancer Institute and the Office of Behavioral and Social Science Research (OBSSR). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institute on Drug Abuse, the National Cancer Institute, the National Institutes of Health, or the Office of Behavioral and Social Science Research. R 3.6.1 software, including the gee package (originally written by V. J. Carey and ported by T. Lumley and B. Ripley) and plotrix package (written by J. Lemon) was used for analyses and graphs. John Dziak thanks Dr. Stephanie T. Lanza for discussions which were enormously helpful in beginning this paper. John Dziak also thanks Jessica Dolan for assistance in implementing the paper.

Appendix A: Technical explanation of model-fitting details

The approach used in the `funreg` package for fitting Model (1.1) basically follows that outlined in [19] with back-end calculations handled using the `mgcv` and `splines` packages in R [see 93]. There are two practical challenges which the approach of [19] overcomes. The first challenge is how to manage the sparsity of the $X_i(t)$ functions, observed at only a finite number of points despite being theoretically smooth functions. This is managed by modeling the distribution of the $X_i(t)$ using functional principal components analysis. The second is how to avoid spuriously overfitting the data because of the great flexibility of the model. In classic regression, overparameterization is often addressed either by selecting a subset of variables or by using a ridge or empirical Bayes penalty.

An approach combining these options (basis selection and penalization) is used here.

First, the X_i are viewed as having been measured with some error, so that conceptually one is estimating a latent X_i which may be slightly more smooth and less variable than the observed X_i. One approach would be to smooth each participant's curve separately. However, this may oversmooth and distort the curves, especially for those subjects who have sparser data. It also does not provide insight into the distribution of the X_i's in general. Therefore, [19] combine information from all the X_i's by using functional principal components analysis [3]. It is assumed that each $X_i(t)$ can be expressed as a fixed mean function $\mu_X(t)$ plus a random process $X_i^c(t)$ with mean zero and a finite covariance function $\Sigma_X(s,t) = \text{Cov}(X_i(s), X_i(t)) = \text{Cov}(X_i^c(s), X_i^c(t))$. Suppose that the eigenfunctions ψ of the function $\Sigma_X(s,t)$ have been estimated on a fine regular grid of t. Let the first M_X of these eigenfunctions, in decreasing order by eigenvalue, be denoted as $\psi_1(t), ..., \psi_{M_X}(t)$, and let each subject's curve X_i^c have a loading c_{ij} on each $\psi_j(t)$. In this case, we can approximate the random $X_i^c(t)$ by

$$X_i^c(t) \approx \sum_{j=1}^{M_X} c_{ij}\psi_j(t) \tag{A.1}$$

where $\psi_j(t)$ is the jth eigenfunction in order of the eigenvalues. Goldsmith and co-workers [19] recommend a large M_X (in one of their sample code files $M_X = 30$). Thus, the curves can be summarized by a Karhunen-Loève decomposition, and this can be used to rebuild a smoothed version of each curves on as dense a regular grid as is desired.

In order to perform the Karhunen-Loève decomposition described above, one needs to estimate the covariance function $\Sigma_X(s,t) = \text{Cov}(X_i(s), X_i(t))$ of the X_i curves. This seems very challenging because the X_i curves are possibly sparse and error-contaminated data which are not necessarily measured on a common grid. Specifically, because only a finite and perhaps irregular grid of points on the X_i have been observed, c_{ij} in (A.1) is not known but has to be replaced with a good estimate. Fortunately, empirical Bayes theory provides a way to find a best linear unbiased estimate in such a situation. Specifically, if a random quantity $v_i = \hat{\boldsymbol{\mu}}_{v_i} + \mathbf{bZ_i} + e_i$ is being predicted from random predictors $\mathbf{Z}_i$ in a linear mixed model, where the e_i are independent normal and the $\mathbf{Z}_i$ have some nondegenerate multivariate normal distribution, then the empirical Bayes estimate and best linear unbiased predictor for the subject-specific coefficients is

$$\tilde{\mathbf{b}}_i = \text{Cov}\,(\mathbf{b})\,\mathbf{Z}_i^T\,(\text{Cov}\,(\mathbf{v}))^{-1}\,(\mathbf{v}_i - \hat{\boldsymbol{\mu}}_{v_i})$$

where

$$\text{Cov}\,(\mathbf{v}) = \mathbf{Z}_i\text{Cov}\,(\mathbf{b})\,\mathbf{Z}_i^T + (\text{Var}(e))\mathbf{I}$$

[61, 69]. Goldsmith and co-authors [19] are predicting $X_i(t) = \mu_X(t) + \boldsymbol{\psi}(t)\mathbf{c}_i + e_i(t)$, so they plug ψ_i in for $\mathbf{Z}_i$ above, diag(λ) for Cov (**b**), and the measurement

error in X for the variance of the e_i, and thus they can get estimates for the $\mathbf{c}_i$ vectors. They still need an estimate for σ_x^2 (i.e., the measurement error variance of X), so they get it from binning the observations and comparing the smoothed variance between and within bins. They also obtain a $\mu_X(t)$ estimate from a penalized spline fit on all of the X's together. Once the c_{ij}'s have been estimated, the covariance function of X_i and the estimates of $X_i(t)$ at any point are treated as known.

One might ask whether to use the original X_i or the centered $X_i^c(t) = X_i(t) - \mu_X(t)$ as predictors. The Goldsmith et al. (2011) sample code uses $X_i^c(t)$, after estimating $\mu_X(t)$ using a penalized spline, and funreg follows this approach. This centering may make numerical computations more stable. However, for interpreting the estimate of $\hat{\theta}_1(t)$, it matters little whether the X_i are centered or not. This is because

$$\begin{aligned}\int_0^1 X_i^c(t)\hat{\theta}_1(t)dt &= \int_0^1 \left(X_i(t) - \mu_X(t)\right)\hat{\theta}_1(t)dt \\ &= \int_0^1 X_i(t)\hat{\theta}_1(t)dt - \int_0^1 \mu_X(t)\hat{\theta}_1(t)dt\end{aligned}$$

where the second term does not depend on i and therefore gets absorbed into the scalar intercept constant θ_0.

Once the X_i have been estimated for all subjects on each point of a fine and regular grid, it is possible to continue with a regression. However, the goal is not merely to do a regression of y on a grid of hundreds of $X(t)$ values, but to do this in a special way so that the θ_1 estimates for each consecutive pair of bins are close to each other, thus giving $\theta_1(t)$ a smooth form. Specifically, it is assumed that $\theta_1(t)$ is a nonparametric function assumed to be reasonably smooth. It can therefore be approximated by using a parametric regression on some appropriate basis consisting of functions $\boldsymbol{\phi}(t)$ [63]. That is, we assume

$$\theta_1(t) \approx \sum_{m=1}^{M} b_m \phi_j(t) \tag{A.2}$$

for some finite and reasonably small integer M.

Many options are available for the form of the basis functions ϕ. Goldsmith and co-authors [19] point out that they could have reused the eigenfunctions from the decomposition and reconstruction of the $\mathbf{X}$ curves as the basis functions for $\theta_1(t)$ (i.e., set $\phi_m(t) = \psi_m(t)$), but this would rest on a possibly unjustified assumption that the functions which summarize X well also summarize β well. Alternatively, they could also have used B-splines [15], a Fourier basis, or many other choices. Instead, they argue that a penalized truncated power spline basis is reasonable [19, 71]. This basis consists of some simple functions of time (e.g., $1, t, t^2, t^3$), and some knot terms (e.g., $\left((t-\kappa_k)_+\right)^3$) for several prespecified knot locations $\kappa_1, ..., \kappa_K$. As [7] show, the choice of basis for approximating $\theta_1(t)$ is very important for estimation quality, so further research may be warranted here.

Once the basis functions have been chosen, substituting (A.1) and (A.2) into Model (1.1), we get

$$\begin{aligned} g\left(\mu_{Y_i}\right) &= \alpha + \int_0^1 X(t)\theta_1(t)dt \\ &= \alpha^c + \int_0^1 X^c(t)\beta(t)dt \\ &\approx \alpha^c + \sum_{j=1}^{M} \int_0^1 c_{ij}\psi_j(t)\phi_j(t)\mathbf{b}dt \\ &= \alpha^c + \left(\int_0^1 \psi_j(t)\phi_j(t)\right)\left(\sum_{j=1}^{M} c_{ij}\mathbf{b}dt\right) \\ &= \alpha^c + \mathbf{CJb} \end{aligned} \tag{A.3}$$

where α^c is the adjusted value of α because of the centering, $\mathbf{C}$ the matrix of c_{ij}'s, $\mathbf{J} = \int_0^1 \psi_j(t)\phi(t)$, and $\mathbf{b}$ is simply a vector of constants to be estimated. Thus, the functional linear regression is estimated as a parametric multiple regression where the matrix $\mathbf{CJ}$ is the regression matrix (design matrix).

Even with a reasonably small M, to prevent overfitting one must still keep the $\mathbf{b}$ coefficients from being too large individually. Instead of using ordinary linear regression estimates of the b_i, [19] apply a smoothness penalty, which can be seen as something like an empirical Bayes prior or a ridge regression penalty. The strength of this penalty function can be chosen automatically using an approach based on restricted maximum likelihood [the REML approach to P-splines; see 71]. Thus [19] call their approach penalized functional regression. The model-based estimates for the covariance matrix of $\mathbf{b}$ in the intermediate regression (A.3), are then transformed using Cramèr's delta method (Taylor linearization) into the metric of $\theta_1(t)$ for each t of interest to obtain pointwise confidence intervals for $\theta_1(t)$. [19] admit that one disadvantage of this approach is that it only takes into account the uncertainty in regressing y on the reconstructed X_i's, and not the uncertainty about the reconstructed X_i's themselves (i.e., the uncertainty in the estimates of the c_{ij} loadings, as well as the choice of M_X). In addition, it ignores possible bias and uncertainty due to the necessity of using a smoothing penalty, and it only provides pointwise coverage so it does not directly allow inferences on differences in the function between multiple times. For these reasons, resampling techniques like bootstrapping might give more realistic confidence intervals.

References

[1] Andersen, S. L. and Teicher, M. H. (2008). Stress, sensitive periods and maturational events in adolescent depression. *Trends in Neurosciences*, **31**, 183–191.

[2] Andersen, S. L., Tomada, A., Vincow, E. S., Valente, E., Polcari, A., and Teicher, M. H. (2008). Preliminary evidence for sensitive periods in the effect of childhood sexual abuse on regional brain development. *Journal of Neuropsychiatry and Clinical Neurosciences*, **20**, 292–301.

[3] Ash, R. B. and Gardner, M. F. (1975). *Topics in Stochastic Processes.* New York: Academic Press. MR0448463

[4] Ben-Zeev, D., Scherer, E. A., Wang, R., Xie, H., and Campbell, A. T. (2015). Next-generation psychiatric assessment: Using smartphone sensors to monitor behavior and mental health. *Psychiatric Rehabilitation Journal*, **38**, 218–226.

[5] Borland, R., Yong, H.-H., O'Connor, R. J., Hyland, A., and Thompson, M. E. (2010). The reliability and predictive validity of the Heaviness of Smoking Index and its two components: Findings from the International Tobacco Control Four Country study. *Nicotine & Tobacco Research*, **12**, S45–S50.

[6] Braveman, P., Acker, J., Arkin, E., Bussel, J., Wehr, K., and Proctor, D. (2018). *Early Childhood Is Critical to Health Equity.* Princeton, NJ: Robert Wood Johnson Foundation.

[7] Cai, T. T. and Yuan, M. (2012). Minimax and adaptive prediction for functional linear regression. *Journal of the American Statistican Association*, **107**, 1201–1216. MR3010906

[8] Cardot, H., Ferraty, F., Mas, A., and Sarda, P. (2003). Testing hypotheses in the functional linear model. *Scandinavian Journal of Statistics*, **30**, 241–225. MR1965105

[9] Chow, S.-M., Witkiewitz, K., Grasman, R. P. P. P., and Maisto, S. A. (2015). The cusp catastrophe model as cross-sectional and longitudinal mixture structural equation models. *Psychological Methods*, **20**, 142–164.

[10] Cofta-Woerpel, L., McClure, J. B., Li, Y., Urbauer, D., Cinciripini, P. M., and Wetter, D. W. (2011). Early cessation success or failure among women attempting to quit smoking: Trajectories and volatility of urge and negative mood during the first postcessation week. *Journal of Abnormal Psychology*, **120**, 596–606.

[11] Compton, W. M., Jones, C. M., Baldwin, G. T., Harding, F. M., Blanco, C., and Wargo, E. M. (2019). Targeting youth to prevent later substance use disorder: An underutilized response to the US opioid crisis. *AJPH Perspectives*, **109**(S3), S185–S189.

[12] Crainiceanu, C., Reiss, P., Goldsmith, J., Huang, L., Huo, L., Scheipl, F. (2014). refund: Regression with Functional Data. R package version 0.1-11. Accessed at `cran.r-project.org`.

[13] Dziak, J. J., Li, R., Tan, X., Shiffman, S., and Shiyko, M. P. (2015). Modeling intensive longitudinal data with mixtures of nonparametric trajectories and time-varying effects. *Psychological Methods*, **20**, 444–469.

[14] Dziak, J. J. and Shiyko, M. P. (2016). funreg: Functional Regression for Irregularly Timed Data. R package version 1.2. Accessed at `http://CRAN.R-project.org/package=funreg`.

[15] Eilers, P. H. C., and Marx, B. D. (1996). Flexible smoothing with B-

splines and penalties (with comments and rejoinder). *Statistical Science*, **11**, 89–121. MR1435485

[16] Escabias, M., Aguilera, A. M., and Vanderrama, M. J. (2004). Principal component estimation of functional logistic regression: Discussion of two different approaches. *Nonparametric Statistics*, **16**, 365–384. MR2073031

[17] Fish, J. N., Rice, C. E., Lanza, S. T., and Russell, S. T. (2018). Is young adulthood a critical period for suicidal behavior among sexual minorities? Results from a US national sample. *Prevention Science*, in press.

[18] GBD 2015 Tobacco Collaborators (2017). Smoking prevalence and attributable disease burden in 195 countries and territories, 1990-2015: a systematic analysis from the Global Burden of Disease Study 2015. *Lancet*, **389**, 1885–1906.

[19] Goldsmith, J., Bobb, J., Crainiceanu, C. M., Caffo, B., and Reich, D. (2011a). Penalized functional regression. *Journal of Computational and Graphical Statistics*, **20**, 830–851. MR2878950

[20] Goldsmith, J., Crainiceanu, C. M., Caffo, B. S., and Reich, D. S. (2011b). Penalized functional regression analysis of white-matter tract profiles in multiple sclerosis. *Neuroimage*, **57**, 431–439.

[21] Goldsmith, J., Huang, L., and Crainiceanu, C. M. (2014). Smooth scalar-on-image regression via spatial Bayesian variable selection. *Journal of Computational and Graphical Statistics*, **23**, 46–64. MR3173760

[22] Goldsmith, J., Scheipl, F., Huang, L., Wrobel, J., Gellar, J., Harezlak, J., McLean, M. W., Swihart, B., Xiao, L., Crainiceanu, C. M., and Reiss, P. T. (2016). refund: Regression with Functional Data. R package version 0.1-16. Accessed http://CRAN.R-project.org/package=refund.

[23] Hastie, T., and Tibshirani, R. (1993). Varying-coefficient models. *Journal of the Royal Statistical Society, Series B*, **55**, 757–796. MR1229881

[24] Heatherton, T. F., Kozlowski, L. T., Frecker, R. C., Rickert, W., and Robinson, J. (1989). Measuring the heaviness of smoking: using self-reported time to the first cigarette of the day and number of cigarettes smoked per day. *British Journal of Addiction*, **84**, 791–800.

[25] Hedström, A. K., Olsson, T., Alfredsson, L. (2016). Smoking is a major preventable risk factor for multiple sclerosis. *Multiple Sclerosis*, **22**, 1021–1026.

[26] Heinonen, K., Räikkönen, K., Pesonen, A.-K., Kajantie E., Andersson, S., Eriksson, J. G., Niemelä, A., Vartia, T., Peltola, J., and Lano, A. (2008). Prenatal and postnatal growth and cognitive abilities at 56 months of age: A longitudinal study of infants born at term. *Pediatrics*, **121**, e1325–e1333.

[27] Hendricks, P. S., Ditre, J. W., Drobes, D. J. and Brandon, T. H. (2006). The early time course of smoking withdrawal effects. *Psychopharmacology*, **187**, 385–396.

[28] Hicks, J. L., Althoff, T., Sosic, R., Kuhar, P., Bostjancic, B., King, A. C., Leskovec, J., and Delp, S. L. (2019). Best practices for

analyzing large-scale health data from wearables and smartphone apps. *npj Digital Medicine*, **2**, article 45.

[29] Ivanescu, A. E., Crainiceanu, C. M., and Checkley, W. (2017). Dynamic child growth prediction: A comparative methods approach. *Statistical Modelling*, **17**(6), 468–493. MR3720942

[30] James, G. (2002). Generalized linear models with functional predictor variables. *Journal of the Royal Statistical Society, Series B*, **64**, 411–432. MR1924298

[31] James, G. M., and Hastie, T. J. (2001). Functional linear discriminant analysis for irregularly sampled curves. *Journal of the Royal Statistical Society, Series B*, **63**, 533–550. MR1858401

[32] James, G. M., Wang, J., and Zhu, J. (2009). Functional linear regression that's interpretable. *Annals of Statistics*, **37**, 2083–2108. MR2543686

[33] Kalkhoran, S., Benowitz, N. L., Rigotti, N. A. (2018). Prevention and treatment of tobacco use: JACC health promotion series. *Journal of the American College of Cardiology*, **72**, 1030–1045.

[34] Kamarck, T. W., Muldoon, M. F., Shiffman, S. and Sutton-Tyrrell, K. (2007). Experiences of demand and control during daily life are predictors of carotid atherosclerotic progression among healthy men. *Health Psychology*, **26**, 324–332.

[35] Kaye, A. P., Kwan, A. C., Ressler, K. J., and Krystal, J. H. (2019). A computational model for learning from repeated trauma. *bioRxiv*, https://doi.org/10.1101/659425.

[36] Khoury, J., Gonzalez, A., Levitan, R. D., Pruessner, J. C., Chopra, K., Santo Basile, V., Masellis, M., Goodwill, A., and Atkinson, L. (2015). Summary cortisol reactivity indicators: Interrelations and meaning. *Neurobiology of Stress*, **2**, 34–43.

[37] Knudsen, E. I. (2004). Sensitive periods in the development of the brain and behavior. *Journal of Cognitive Neuroscience*, **16**, 1412–1425.

[38] Kong, D., Staicu, A.-M., and Maity, A. (2016). Classical testing in functional linear models. *Journal of Nonparametric Statistics*, **28**, 813–838. MR3555459

[39] Kozlowski, L. T., Porter, C. Q., Orleans, C. T., Pope, M. A., and Heatherton, T. (1994). Predicting smoking cessation with self-reported measures of nicotine dependence: FTQ, FTND, and HSI. *Drug and Alcohol Dependence*, **34**, 211–216.

[40] Kuehl, R. O. (2000). *Design of Experiments: Statistical Principles of Research Design and Analysis (2nd ed.)*. Pacific Grove, CA: Duxbury Thomson.

[41] Laber, E. B., and Staicu, A.-M. (2017). Functional feature construction for individualized treatment regimes. *Journal of the American Statistical Association*, in press. MR3862352

[42] Liang, K.-Y., and Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika*, **73**, 13–22. MR0836430

[43] Lindor, K. D., Gershwin, M. E., Poupon, R., Kaplan, M., Bergasa, N. V., Heathcote, E. J. (2009). Primary biliary cirrhosis. *Hepatology*,

2009, 291–308.

[44] Lindquist, M. A., and McKeague, I. W. (2009). Logistic regression with Brownian-like predictors. *Journal of the American Statistical Association*, **104**, 1575–1585. MR2597002

[45] Lupien, S. J., McEwen, B. S., Gunnar, M. R., and Heim, C. (2009). Effects of stress throughout the lifespan on the brain, behaviour and cognition. *Nature Reviews Neuroscience*, **10**, 434–445.

[46] N. Maruyama, F. Takahashi, and M. Takeuchi (2009). Prediction of an outcome using trajectories estimated from a linear mixed model. *Journal of Biopharmaceutical Statistics*, **19**, 779–790. MR2751087

[47] McCarthy, D. E., Piasecki, T. M., Fiore, M. C., and Baker, T. B. (2006). Life before and after quitting smoking: an electronic diary study. *Journal of Abnormal Psychology*, **115**, 454–466.

[48] McCullagh, P. and Nelder, J. (1989). *Generalized Linear Models (2nd ed.)*. Boca Raton: Chapman and Hall/CRC. MR3223057

[49] McVicar, D., Moschion, J., and Ours, J. C. (2019). Early illicit drug use and the age of onset of homelessness. *Journal of the Royal Statistical Society, A*, **182**, 345–372. MR3902647

[50] Müller, H.-G. and Stadtmüller, U. (2005). Generalized functional linear models. *Annals of Statistics*, **33**, 774–805. MR2163159

[51] National Academies of Sciences, Engineering, and Medicine (2019). *Vibrant and Healthy Kids: Aligning Science, Practice, and Policy to Advance Health Equity*. Washington, DC: The National Academies Press.

[52] National Institute on Drug Abuse (2003). *Preventing Drug Use Among Children and Adolescents: A Research-Based Guide for Parents, Educators, and Community Leaders (2nd ed.)*. Available online at https://www.drugabuse.gov/sites/default/files/preventingdruguse_2.pdf.

[53] Neely, K. A., Planetta, P. J., Prodoehl, J., Corcos, D. M., Comella, C. L., Goetz, C. G., Shannon, K. L., and Vaillancourt, D. E. (2013). Force control deficits in individuals with Parkinson's disease, multiple systems atrophy, and progressive supranuclear palsy. *PLOS ONE*, **8**, e58403.

[54] Nguyen, H., and Loughran, T. A. (2018). On the measurement and identification of turning points in criminology. *Annual Review of Criminology*, **1**, 335–358.

[55] Njagi, E. J., Rizopoulos, D., Molenberghs, G., Dendale, P., and Willekens, K. (2013). A joint survival-longitudinal modelling approach for the dynamic prediction of rehospitalization in telemonitored chronic heart failure patients. *Statistical Modeling*, **13**, 179–198. MR3179523

[56] Orben, A., and Przybylski, A. K. (2019). The association between adolescent well-being and digital technology use. *Nature Human Behavior*, https://doi.org/10.1038/s41562-018-0506-1.

[57] Pechtel, P., Lyons-Ruth, K., Anderson, C. M., and Teicher, M. H. (2014). Sensitive periods of amygdala development: The role of maltreatment in preadolescence. *Neuroimage*, **97**, 236–244.

[58] Piasecki, T. M., Niaura, R., Shadel, W. G., Abrams, D., Gold-

stein, M., Fiore, M. C., Baker, T. B. (2000). Smoking withdrawal dynamics in unaided quitters. *Journal of Abnormal Psychology*, **109**, 74–86.

[59] Piasecki, T. M. (2006). Relapse to smoking. *Clinical Psychology Review*, **26**, 196–215.

[60] R Core Team (2019). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Accessed at http://www.R-project.org/.

[61] Rabe-Hesketh, S., and Skrondal, A. (2008). Generalized linear mixed-effects models. In Fitzmaurice, G., Davidian, M., Verbeke, G., and Molenberghs, G., *Longitudinal Data Analysis*, pp. 79–106. Boca Raton: Chapman & Hall. MR1500114

[62] Ramsay, J. O. and Dalzell, C. J. (1991). Some tools for functional data analysis. *Journal of the Royal Statistical Society, Series B*, **53**, 539–572. MR1125714

[63] Ramsay, J. O. and Silverman, B. W. (2005). *Functional Data Analysis (2nd ed.)*. Springer: New York. MR2168993

[64] Ramsay, J. O., Wickham, H., Graves, S., and Hooker, G. (2014). fda: Functional Data Analysis. R package version 2.4.4. Accessed at http://CRAN.R-project.org/package=fda.

[65] Ramsey, F., and Schafer, D. (2013). *The Statistical Sleuth: A Course in Methods of Data Analysis (3nd ed.)*. Boston: Brooks/Cole.

[66] Ratcliffe, S. J., Heller, G. Z. and Leader, L. R. (2002). Functional data analysis with application to periodically stimulated foetal heart rate data. II: Functional logistic regression. *Statistics in Medicine*, **21**, 1115–1127.

[67] Reiss, P. T., Goldsmith, J., Shang, H. L., and Ogden, R. T. (2017). Methods for scalar-on-function regression. *International Statistical Review*, **85**, 228–249. MR3686566

[68] Rizopoulos, D. (2011). Dynamic predictions and prospective accuracy in joint models for longitudinal and time-to-event data. *Biometrics*, **67**, 819–829. MR2829256

[69] Robinson, G. K. (1991). That BLUP is a good thing: The estimation of random effects. *Statistical Science*, **6**, 15–51. MR1108815

[70] Roque, N. A. and Ram, N. (2019). tsfeaturex: An R package for automating time series feature extraction. *Journal of Open Source Software*, https://doi.org/10.21105/joss.01279.

[71] Ruppert, D., Wand, M. P., and Carroll, R. J. (2003). *Semiparametric Regression*. Cambridge: Cambridge University Press. MR1998720

[72] Sang, P., Wang, L., & Cao, J. (2018). Estimation of sparse functional additive models with adaptive group LASSO. *Statistica Sinica*, in press, http://www3.stat.sinica.edu.tw/ss_newpaper/SS-2017-0491_na.pdf.

[73] Shapiro, J. M., Smith, H., and Schaffner, F. (1979). Serum bilirubin: a prognostic factor in primary biliary cirrhosis. *Gut*, **20**, 138–140.

[74] Shiffman, S., Gwaltney, C. J., Balabanis, M. H., Liu, K. S., Paty,

J. A., Kassel, J. D., Hickcox, M., and Gnys, M. (2002). Immediate antecedents of cigarette smoking: An analysis from ecological momentary assessment. *Journal of Abnormal Psychology*, **111**, 531–545.

[75] Shiffman, S. (2007). Use of more nicotine lozenges leads to better success in quitting smoking. *Addiction*, **102**, 809–814.

[76] Shiffman, S. (2009). Ecological momentary assessment (EMA) in studies of substance use. *Psychological Assessment*, **21**, 486–497.

[77] Shiffman, S., Engberg, J. B., Paty, J. A., Perz, W. G., Gnys, M., Kassel, J. D., and Hickcox, M. (1997). A day at a time: predicting smoking lapse from daily urge. *Journal of Abnormal Psychology*, **106**, 104–116.

[78] Shiffman, S., Hickcox, M., Paty, J. A., Gnys, M., Kassel, J. D., and Richards, T. J. (1996). Progression from a smoking lapse to relapse: prediction from abstinence violation effects, nicotine dependence, and lapse characteristics. *Journal of Consulting and Clinical Psychology*, **64**, 993–1002.

[79] Shiffman, S., Paty, J. A., Gnys, M., Kassel, J. A., and Hickcox, M. (1996). First lapses to smoking: Within-subjects analysis of real-time reports. *Journal of Consulting and Clinical Psychology*, **64**, 366–379.

[80] Singh, R., Quinn, J. D., Reed, P. M., Keller, K. (2018). Skill (or lack thereof) of data-model fusion techniques to provide an early warning signal for an approaching tipping point. *PLoS ONE*, **13**, e0191768.

[81] Sørensen, H., Goldsmith, J., and Sangalli, L. M. (2013). An introduction with medical applications to functional data analysis. *Statistics in Medicine*, **32**, 5222–5240. MR3141370

[82] Shiyko, M. and Lanza, S. T., and Tan, X. and Li, R. and Shiffman, S. (2011). Using the time-varying effect model (TVEM) to examine dynamic associations between negative affect and self confidence on smoking urges: Differences between successful quitters and relapsers. *Prevention Science*, **13**, 288–299.

[83] Steidtmann, D., Manber, R., Blasey, C., Markowitz, J. C., Klein, D. N., Rothbaum, B. O., Thase, M. E., Kocsis, J. H., & Arnow, B. A. (2013). Detecting critical decision points in psychotherapy and psychotherapy + medication for chronic depression. *Journal of Consulting and Clinical Psychology*, **81**, 783–792.

[84] Stone, A. A. and Shiffman, S. (1994). Ecological momentary assessment (EMA) in behavorial medicine. *Annals of Behavioral Medicine*, **16**, 199–202.

[85] Tan, X., Shiyko, M. P., Li, R., Li, Y., and Dierker, L. (2012). A time-varying effect model for intensive longitudinal data. *Psychological Methods*, **17**, 61–77.

[86] Tibshirani, R. (1996). Regression shrinkage and selection via the Lasso. *Journal of the Royal Statistical Society, Series B*, **58**, 267–288. MR1379242

[87] Trail, J. B., Collins, L. M., Rivera, D. E., Li, R., Piper, M. E., and Baker, T. B. (2014). Functional data analysis for dynamical system identification of behavioral processes. *Psychological Methods*, **19**, 175–187.

[88] van Houwelingen, H. C. (2006). Dynamic prediction by landmarking in event history analysis. *Scandinavian Journal of Statistics*, **34**, 70–85. MR2325243

[89] van Zundert, R. M. P., Boogerd, E. A., Vermulst, A. A., and Engels, R. C. (2009). Nicotine withdrawal symptoms following a quit attempt: An ecological momentary assessment study among adolescents. *Nicotine and Tobacco Research*, **11**, 722–729.

[90] Vinci, C., Li, L., Wu, C., Lam, C. Y., Guo, L., Correa-Fernández, V., Spears, C. A., Hoover, D. S., Etcheverry, P. E., and Wetter, D. W. (2017). The association of positive emotion and first smoking lapse: An ecological momentary assessment study. *Health Psychology*, **36**, 1038–1046.

[91] Walls, T. A., and Schafer, J. L. (2006). *Models for Intensive Longitudinal Data*. Oxford: Oxford University Press. Wand, M. (2013). MR2334216

[92] Wang, J.-L., Chiou, J.-M., and Müller, H.-G. (2016). Review of functional data analysis. *Annual Review of Statistics and its Application*, **3**, 257–295.

[93] Wood, S.N. (2017). *Generalized Additive Models: An Introduction with R (2nd ed.)*. New York: Chapman and Hall/CRC. MR2206355

[94] Worley, M. J.. Heinzerling, K. G., Shoptaw, S., and Ling, W. (2015). Pain volatility and prescription opioid addiction treatment outcomes in patients with chronic pain. *Experimental and Clinical Psychopharmacology*, **23**(6), 428–435.

[95] Wrobel, D., Zipunnikov, V., Schrack, J., and Goldsmith, J. (2018). Registration for exponential family functional data. *Biometrics*, in press. MR3953706

[96] Yen, J. D. L., Thomson, J. R., Paganin, D. M., Keith, J. M., and MacNally, R. (2015). Function regression in ecology and evolution: FREE. *Methods in Ecology and Evolution*, **6**, 17–26.

[97] Yuen, H. P., and Mackinnon, A. (2016). Performance of joint modelling of time-to-event data with time-dependent predictors: an assessment based on transition to psychosis data. *PeerJ*, **4**, e2582. eCollection 2016.

[98] Zhang, Y., Zhou, J., Niu, F., Donowitz, J. R., Haque, R., Petri, W. A. Jr., and Ma, J. Z. (2017). Characterizing early child growth patterns of height-for-age in an urban slum cohort of Bangladesh with functional principal component analysis. *BMC Pediatrics*, **17**, 84.