Published in final edited form as:

Biometrics. 2011 June ; 67(2): 486–494. doi:10.1111/j.1541-0420.2010.01461.x.

A Random Intercepts – Functional Slopes Model for Flexible Assessment of Susceptibility in Longitudinal Designs

Brent A Coull

Department of Biostatistics, Harvard School of Public Health, 655 Huntington Avenue, Boston, MA 02115

Summary

In many biomedical investigations, a primary goal is the identification of subjects that are susceptible to a given exposure or treatment of interest. We focus on methods for addressing this question in longitudinal studies when interest focuses on relating susceptibility to a subject's baseline or mean outcome level. In this context, we propose a random intercepts – functional slopes model that relaxes the assumption of linear association between random coefficients in existing mixed models and yields an estimate of the functional form of this relationship. We propose a penalized spline formulation for the nonparametric function that represents this relationship, and implement a fully Bayesian approach to model fitting. We investigate the frequentist performance of our method via simulation, and apply the model to data on the effects of particulate matter on coronary blood flow from an animal toxicology study. The general principles introduced here apply more broadly to settings in which interest focuses on the relationship between baseline and change over time.

Keywords

Latent variable; Laird-Ware model; Penalized spline; Random intercepts – random slopes model; Semiparametric regression

1. Introduction

A key advantage of longitudinal study designs is their ability to identify subject-specific factors that influence an individual's susceptibility to a time-varying exposure or treatment. In some settings, scientific interest focuses on whether pre-existing disease, as manifested by either baseline level or the mean level of the outcome over the course of the study, is one such factor. Our interest lies in methods that estimate this relationship but do not assume a functional form for the relationship *a priori*.

We develop these methods in the context of analyzing the effects of particulate matter exposure on coronary blood flow in an animal toxicology study. Associations between

*bcoull@hsph.harvard.edu.

7. Supplementary Materials

Web Figures referenced in Sections 3, 4, and 5 are available under the Paper Information link at the Biometrics website <http://www.biometrics.tibs.org>.

increased deaths due to ischemic heart disease and elevated levels of ambient particulate matter are well-established. Recent laboratory studies seek to pinpoint the biologic mechanisms underlying these effects. In a recent experiment conducted at the Harvard School of Public Health (Bartoli et al. 2009), investigators used fluorescent microspheres to repeatedly measure regional myocardial blood flow (hereafter blood flow) in a conscious canine model of myocardial ischemia via coronary occlusion. In layman's terms, investigators implanted laboratory animals with arterial balloons, which could then be inflated to occlude an artery and restrict blood flow into the heart. First assessed at baseline (no occlusion), animals were then exposed under occlusion conditions to filtered air and concentrated ambient particles in a crossover protocol (five hour per day exposures, 3 filtered air, 3 particle exposures). During each exposure occasion, investigators injected colored microspheres into an animal's bloodstream, with these spheres eventually traveling to the heart. Each exposure corresponded to a different microsphere color. After the entire study, multiple pieces of myocardium tissue from each animal were assessed for blood flow under baseline, filtered air and particle-exposed conditions. Blood flow was expressed as the number of spheres with color belonging to that particular exposure, normalized by a reference blood flow rate for each exposure.

One of the primary questions of interest of the study, and the one we focus on here, is whether the effect of exposure on blood flow inside the region of the myocardium affected by the occluded artery, known as the ischemic zone, is greater than that outside this region. Such a determination would suggest that exposure worsens the severity of an adverse cardiac event. Thus, interest focuses on the susceptibility to exposure of a given piece of myocardium tissue as a function of the *degree of ischemia* in that piece induced by occlusion. One complication in this analysis is that an objective determination of whether or not a piece of tissue is in the ischemic zone is not available. Instead, one must make this determination by comparing blood flow under baseline (non-occlusion) and under filtered air occlusion (hereafter, control) conditions. Bartoli et al. (2009) dichotomized this change, and fit a model containing an interaction between this term and exposure.

We seek to estimate the form of the association between the change between control and baseline blood flow and the effect of exposure across pieces of tissue. Figure 1 presents data on this association, plotting the difference in blood flow means under exposed and control conditions as a function of the degree of ischemia, defined as the change from baseline under control conditions.

Figure 1 amounts to plotting estimated intercepts and slopes from a simple linear regression model that expresses blood flow change from baseline as a function of a binary exposure indicator. This scatterplot also displays the uncertainty associated with each estimate, by plotting ± 1 standard error intervals in both the x (the intercepts) and y (the slopes) directions. The uncertainty in the x direction clouds estimation of the functional form of this association and necessitates the development of methods to answer this question of interest.

Mixed effect regression models represent an attractive approach to addressing this question due to the subject-specific interpretation of the random effects. In studies focusing on susceptibility, a linear mixed model with random intercepts and slopes model is a natural

starting point. The bivariate normal assumption in a standard random intercepts and slopes model implies that the conditional model for the slopes given the intercepts is a simple linear model. In cases in which interest focuses on estimating the functional form of this relationship, one does not wish to make this assumption *a priori*, as shrinkage towards linearity induced by the linear mixed model may obscure the true relationship.

In this article we propose a random intercepts – functional slopes model to directly address this scientific question. The model generalizes the standard random intercepts and slopes model by replacing the linear model for the slopes given the intercepts with an unspecified, smooth function that is nonparametrically estimated from the data. The proposed model generalizes existing additive mixed model formulations by allowing one to estimate a smooth function of a latent variable, specifically random intercepts, rather than observed covariates.

Our proposal is related to that of Berry, Carroll, and Ruppert (2002), who considered regression spline models for measurement error problems. These authors considered a univariate outcome with a single covariate, whereas the measurement error in our problem arises out of a longitudinal setting in which the latent model parameters are covariates in a second stage nonparametric regression model. In this sense this work extends that of Wang, Wang, and Wang (2000) and Li, Zhang and Davidian (2004) to the nonparametric regression setting, using the popular Bayesian formulation of penalized splines (Crainiceanu, Ruppert, and Wand 2005; Zhao et al. 2006).

Section 2 reviews the random intercepts and slopes linear mixed model (LMM) of Laird and Ware (1982), presents the random intercepts – functional slopes extension, considers model identifiability, and presents a Bayesian approach to model fitting. Section 3 presents a simulation study that compares the performance of the proposed approach to *ad-hoc* approaches based on existing models. Section 4 applies the methods to analyze the coronary blood flow data, and Section 5 presents further discussion.

2. Random Intercepts – Functional Slopes Model

Let Y_{it} and E_{it} denote the outcome and exposure of interest, respectively, for subject i , $i = 1, \dots, n$ at time $t = 1, \dots, T_i$. Ignoring other covariates for now, the Laird-Ware linear mixed model (Laird and Ware 1982) with random intercepts and slopes is

$$Y_{it} = b_{0i} + b_{1i}E_{it} + \varepsilon_{it}, \quad (1)$$

where

$$\mathbf{b}_i = (b_{0i}, b_{1i})^\top \stackrel{iid}{\sim} N \left(\begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}, \begin{bmatrix} \sigma_0^2 & \sigma_0\sigma_1\rho \\ \sigma_0\sigma_1\rho & \sigma_1^2 \end{bmatrix} \right) \quad (2)$$

and $\varepsilon_{it} \stackrel{iid}{\sim} N(0, \sigma_\varepsilon^2)$, with $\mathbf{b}_i \perp \varepsilon_{it}$. The bivariate normal assumption (2) for the random effects implies that the random slopes and random intercepts satisfy a linear model, such that model (1) can be re-expressed as

$$Y_{it} = b_{0i} + (\gamma_0 + \gamma_1 b_{0i} + \eta_i) E_{it} + \varepsilon_{it},$$

where $\gamma = (\sigma_1 \rho) / \rho_0$ and $\eta_i \stackrel{iid}{\sim} N(0, \sigma_{1.0}^2)$ with $\sigma_{1.0}^2 = \sigma_1^2 (1 - \rho^2)$. Just as the empirical distribution of the best linear unbiased predictions (BLUPs) of these random effects will be shrunk towards normality (Carlin and Louis 2000), the observed relationship between these BLUPs will be shrunk towards linearity, with the amount of shrinkage depending on the influence of this distributional assumption relative to the data. Thus, this shrinkage will be relatively strong in situations where the estimates obtained from the subject-specific data are relatively imprecise, which occurs when there is a relatively small number of repeated measures, the exposure variance is not too large, the heterogeneity in susceptibility is relatively large, or any combination of the above.

We seek to relax the linearity assumption between subject-specific intercepts and slopes in the Laird and Ware (1982) model by specifying a nonparametric term for this functional form. We propose a random intercepts - functional slopes model, hereafter functional slopes model (FSM), taking form (1), but with

$$b_{0i} \stackrel{iid}{\sim} N(\beta_0, \sigma_0^2), \quad b_{1i} | b_{0i} \stackrel{iid}{\sim} N(f(b_{0i}), \sigma_{1.0}^2), \quad \text{and} \quad \varepsilon_{it} \stackrel{iid}{\sim} N(0, \sigma_\varepsilon^2). \quad (3)$$

Here, $f(b)$ is an unspecified smooth function that is estimated from the data. Letting η_i denote the residual error of b_{1i} around its conditional mean $f(b_{0i})$, we assume $b_{0i} \perp \eta_i \perp \varepsilon_{it}$.

We model the nonparametric function $f(b)$ using a penalized spline model. We consider the first degree model with *canonical form* (Ruppert, Wand, and Carroll 2009),

$$f(b) = \gamma_0 + \gamma_1 b + \sum_{k=1}^K u_k z_k(b), \quad \mathbf{u} \sim N(\mathbf{0}, \sigma_u^2 \mathbf{I}),$$

where $\{z_k(b), k = 1, \dots, K\}$ is a set of spline basis functions and $\mathbf{u} = (u_1, \dots, u_K)^T$. Many good options now exist for the specific form of the spline basis functions. See Section 2.1 of Ruppert, Wand, and Carroll (2009) for a review. In our functional slopes model, we choose a thin-plate spline formulation, such that

$$f(b) = \gamma_0 + \gamma_1 b + \sum_{k=1}^K \tilde{u}_k |b - \kappa_k|^3,$$

where $\kappa_1 < \kappa_2 < \dots < \kappa_K$ are fixed knots (Crainiceanu, Ruppert, and Wand 2005). Let $\mathbf{b}_0 = (b_{01}, \dots, b_{0n})^T$ and $\mathbf{b}_1 = (b_{11}, \dots, b_{1n})^T$. Further, let $\mathbf{B}_{n \times 2}$ be the matrix with i^{th} row $B_i = (1, b_{0i})$, $\mathbf{Z}$ be an $n \times K$ matrix with i^{th} row $[|b_{0i} - \kappa_1|^3 \dots |b_{0i} - \kappa_K|^3]$, $\mathbf{\Omega}$ be a $K \times K$ matrix with $(l, k)^{\text{th}}$ element equal to $|\kappa_l - \kappa_k|^3$, $\boldsymbol{\gamma} = (\gamma_0, \gamma_1)^T$, and $\tilde{\mathbf{u}} = (\tilde{u}_1, \dots, \tilde{u}_K)^T$. The thin-plate spline penalty is induced by assuming a multivariate normal distribution for $\tilde{\mathbf{u}}$, so that we can write

$$\mathbf{f}(\mathbf{b}_0) = \mathbf{B}\boldsymbol{\gamma} + \mathbf{Z}\tilde{\mathbf{u}}$$

where $\tilde{\mathbf{u}} \sim N \left[\mathbf{0}, \sigma_u^2 \mathbf{\Omega}^{-1/2} (\mathbf{\Omega}^{-1/2})^T \right]$. Under the reparameterization $\mathbf{u} = \mathbf{\Omega}^{1/2} \tilde{\mathbf{u}}$ and $\mathbf{Z} = \mathbf{Z}\tilde{\mathbf{\Omega}}^{-1/2}$, the model for $\mathbf{b}_1$ has canonical form

$$\mathbf{b}_1 = \mathbf{B}\boldsymbol{\gamma} + \mathbf{Z}\mathbf{u} + \boldsymbol{\eta}, \text{cov} \begin{pmatrix} \mathbf{u} \\ \boldsymbol{\eta} \end{pmatrix} = \begin{pmatrix} \sigma_u^2 \mathbf{I}_K & \mathbf{0} \\ \mathbf{0} & \sigma_{1.0}^2 \mathbf{I}_n \end{pmatrix}.$$

Our experience suggests that the model based on this choice of basis yields satisfactory mixing of the Markov Chain Monte Carlo algorithm proposed in Section 2.2.

An advantage of the Laird-Ware model is interpretability of β_0 and β_1 , which represent the mean baseline and exposure slope, respectively, in the study population (i. e. μ_0 and μ_1). In the functional slopes model, β_0 again represents the mean baseline level, μ_0 . However, the mean exposure slope is $\mu_1 = \int f(b)\phi(b)db$, where $\phi(b)$ is the standard normal density. While this latter quantity is not available in closed form for general f , an estimate of this mean susceptibility value is available as a by-product of the Monte Carlo algorithm proposed in Section 2.2 for simulating from the joint posterior of the model parameters.

In standard applications of penalized spline models, the nonparametric term $f(\cdot)$ is a function of an observed covariate, and so knots are typically chosen to be spread throughout the range of the covariate (Ruppert, Wand, and Carroll 2003). An issue that arises in our functional slopes formulation is that the argument of $f(\cdot)$, $\mathbf{b}_0$, is unobserved. To address this issue, we first fit the Laird-Ware random intercepts and slopes model to the data, obtain the BLUPs of the random intercepts from this preliminary fit, and use the approach of Ruppert et al. (2003) treating these predictions as fixed. This approach is similar to the strategy advocated by Berry et al. (2002), who place knots within the estimated range of a latent variable. We investigate the effectiveness of this strategy in a simulation study described in Section 3.

Model (1) and (3) can be easily extended to accommodate systematic subject-specific factors that are associated with the baseline level of the outcome. The general form of the functional slopes model is

$$Y_{it} = b_{0i} + b_{1i}E_{it} + \mathbf{W}_{it}\boldsymbol{\alpha} + \varepsilon_{it}, \quad (4)$$

where $b_{0i} \sim N(\mathbf{X}_i \boldsymbol{\beta}, \sigma_0^2)$, $b_{1i} | b_{0i} \sim N(f(b_{0i}), \sigma_{1.0}^2)$, with $\mathbf{X}_i$ a $p \times 1$ vector containing between-subject covariates that remain constant throughout the duration of the study and $\mathbf{W}_{it}$ a $q \times 1$ vector of time-varying covariates. This formulation assumes that between-subject factors relate to a subject's overall outcome level, which is in turn associated with that subject's susceptibility to exposure. This formulation yields a natural partitioning of covariates into between- and within-subject covariates, as recommended by Neuhaus and McCulloch (2006) and others.

2.1 Model Identifiability

An important question concerns identifiability of model (4). This issue amounts to a problem of identifiability in a nonparametric regression measurement error setting (Carroll et al. 2004). Consider fixed effects estimates $\{b_{0i}\}$ and $\{b_{1i}\}$ of the subject-specific intercepts and slopes, respectively. Consider estimation of the parameters in the model $b_{1i} = f(b_{0i}) + w_i$. Instead of having in hand the covariate values $\{b_{0i}\}$, we have versions $\{b_{0i}\}$ measured with error. Specifically, we have $b_{0i} = b_{0i} + v_i$, where v_i is estimation error. Berry et al. (2002) considered this measurement error framework when one has independent replicates of the covariate measured with error (b_{0i} in this case), and noted that the model was identifiable in this case. Carroll et al. (2004) showed that the model is identifiable if, instead of independent replicates, an instrumental variable for the true values of the covariates is available. That is, the model is identifiable if there exists a variable S such that $S_i = a_0 + a_1 b_{0i} + v_i$, where v_i is independent of (b_{0i}, v_i, w_i) , and $\text{Cov}[b_{0i}, f(b_{0i})] \neq 0$.

In the functional slopes model, we do not have either replicates or an instrument for the random intercepts, but the model is identifiable due to the form of the variance of v_i . That is, in the functional slopes setting, the error structure takes the form

$$\text{Var}(v_i) = \sigma_\varepsilon^2 (D_i^\top D_i)^{-1},$$

where D_i is the $n_i \times 2$ covariate matrix from the subject-specific regression $Y_{it} = \beta_{0i} + \beta_{1i} E_{it} + \varepsilon_{it}$. Because σ_ε^2 is estimable, the model is identifiable. See also Wang et al. (2000) and Wall and Amemiya (2000) for discussions of this important difference between this longitudinal setting, in which the error component is identified, and the general measurement error setting.

2.2 Estimation

We take a fully Bayesian approach to model fitting, using Monte Carlo Markov chain (MCMC) methods (Gelfand and Smith 1990) to sample from the joint posterior distribution of all model parameters. Specifically, let $[A]$ and $[A|B]$ represent densities and conditional densities, respectively, and let $N = \sum_{i=1}^n T_i$ denote the total number of observations taken on the n subjects in the study. Define $\mathbf{Y}$ as an $N \times 1$ vector containing all responses, $\mathbf{X}$ as an $n \times p$ matrix with rows $\mathbf{X}_i$, $\mathbf{W}$ as an $N \times q$ matrix with rows $\mathbf{W}_{it}$, and $\mathbf{E}$ as an $N \times 1$ vector of all exposures. Further, for notational convenience, let $\mathbf{Z}_0$ be an $N \times n$ matrix whose i^{th} column contains 1's in those rows corresponding to subject i and 0's otherwise. Let $\mathbf{Z}_1$ be an

$N \times n$ matrix defined similarly, except the i^{th} column contains $\mathbf{E}_i = (E_{i1}, \dots, E_{iT_i})^T$ in those rows corresponding to subject i . Let $\mathbf{C}_b = [\mathbf{B}|\mathbf{Z}]$, with the b subscript denoting the dependence of this matrix on $\mathbf{b}_0$, $\mathbf{C}_{bi}$ denote the i^{th} row of $\mathbf{C}_b$, and let $\boldsymbol{\omega} = (\boldsymbol{\gamma}, \mathbf{u})^T$. Consider the general model (4), and let $\boldsymbol{\theta} = (\mathbf{b}_0, \mathbf{b}_1, \boldsymbol{\omega}, \boldsymbol{\beta}, \boldsymbol{\alpha}, \sigma_0^2, \sigma_{1.0}^2, \sigma_u^2, \sigma_\varepsilon^2)^T$. The joint posterior density of the parameters given the data $\mathbf{y}$, $\mathbf{X}$, $\mathbf{E}$, and $\mathbf{W}$ is

$$[\boldsymbol{\theta}|\mathbf{y}, \mathbf{X}, \mathbf{E}, \mathbf{W}] \propto [\mathbf{y}|\mathbf{X}, \mathbf{E}, \mathbf{W}, \mathbf{b}_0, \mathbf{b}_1, \sigma_\varepsilon^2] [\mathbf{b}_1|\mathbf{b}_0, \boldsymbol{\omega}, \sigma_{1.0}^2] [\mathbf{b}_0|\sigma_0^2] [\mathbf{u}|\sigma_u^2] \times [\boldsymbol{\beta}|\boldsymbol{\gamma}] [\boldsymbol{\alpha}] [\sigma_0^2] [\sigma_u^2] [\sigma_{1.0}^2] [\sigma_\varepsilon^2].$$

To implement our MCMC approach, we use flat, non-informative priors for the elements of $\boldsymbol{\beta}$, $\boldsymbol{\gamma}$, and $\boldsymbol{\alpha}$, and inverse gamma priors

$\sigma_0^2 \sim IG(A_0, B_0)$, $\sigma_u^2 \sim IG(A_u, B_u)$, $\sigma_{1.0}^2 \sim IG(A_{1.0}, B_{1.0})$, and $\sigma_\varepsilon^2 \sim IG(A_\varepsilon, B_\varepsilon)$, for the variance components. Here, IG denotes the inverse-gamma distribution defined as

$$f(x|A, B) = \frac{1}{\Gamma(A)B^A x^{A+1}} \exp\left(-\frac{1}{Bx}\right) I_{(0, \infty)}(x).$$

In the simulations and data analysis, we fixed the hyperparameters at $A_\varepsilon = A_0 = A_{1.0} = A_u = .1$ and $B_\varepsilon = B_0 = B_{1.0} = B_u = 10$. These choices represent relatively vague priors, and simulations presented in the next section show that they yield an algorithm that is able to improve upon the LMM fit for a variety of nonlinear functional forms for the intercept-slope association. In our simulation study, we evaluated the sensitivity of results to these choices, re-running all simulation scenarios using the variance component priors proposed by Berry et al. (2002) for a general nonparametric regression problem with measurement error.

Under model (4), the joint posterior of $\boldsymbol{\theta}$ given the data is proportional to

$$\begin{aligned} & \exp\left\{-\frac{1}{2\sigma_\varepsilon^2}\|\mathbf{Y} - (\mathbf{Z}_0\mathbf{b}_0 + \mathbf{Z}_1\mathbf{b}_1 + \mathbf{W}\boldsymbol{\alpha})\|^2 - \frac{1}{2\sigma_{1.0}^2}\|\mathbf{b}_1 - \mathbf{C}_b\boldsymbol{\omega}\|^2\right\} \\ & \times \exp\left\{-\frac{1}{2\sigma_0^2}\|\mathbf{b}_0 - \mathbf{X}\boldsymbol{\beta}\|^2 - \frac{1}{2\sigma_u^2}\|\mathbf{u}\|^2 - \frac{1}{B_\varepsilon\sigma_\varepsilon^2} - \frac{1}{B_u\sigma_u^2} - \frac{1}{B_0\sigma_0^2} - \frac{1}{B_{1.0}\sigma_{1.0}^2}\right\} \quad (5) \\ & \times \sigma_\varepsilon^{-2(N/2+A_\varepsilon+1)} \sigma_u^{-2(K/2+A_u+1)} \times \sigma_0^{-2(n/2+A_0+1)} \sigma_{1.0}^{-2(n/2+A_{1.0}+1)}. \end{aligned}$$

It is possible to set up a Gibbs sampler to sample from (5) using the software package WinBUGS (Spiegelhalter, Thomas, and Best 2000). This package offers general MCMC sampling strategies that tend to work well in a wide variety of problems. The major advantage of this implementation is that it is extremely easy to program. Experience, however, shows that these standard strategies tend to mix slowly for our functional slopes formulation and that one can significantly improve upon this performance using a Metropolis-Hastings-within-Gibbs algorithm tailored specifically for this problem. Letting $\boldsymbol{\theta}_0 = \boldsymbol{\theta}\mathbf{b}_0$, $\lambda_u = \sigma_{1.0}^2/\sigma_u^2$, $\lambda_1 = \sigma_\varepsilon^2/\sigma_{1.0}^2$, and $\mathbf{D} = \text{Diag}[0, 0, \mathbf{1}_K]$, we sample the model parameters as follows:

$$1. \quad \boldsymbol{\beta}\mathbf{b}_0, \sigma_0^2, \mathbf{X} \sim N\left[\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{b}_0, \sigma_0^2\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\right]$$

2. $\alpha|\mathbf{b}_0, \mathbf{b}_1, \mathbf{y}, \mathbf{E}, \mathbf{W}, \sigma_\varepsilon^2 \sim N \left[\left(\mathbf{W}^\top \mathbf{W} \right)^{-1} \mathbf{W}^\top [\mathbf{y} - \mathbf{Z}_0 \mathbf{b}_0 - \mathbf{Z}_1 \mathbf{b}_1], \sigma_\varepsilon^2 \left(\mathbf{W}^\top \mathbf{W} \right)^{-1} \right]$
3. $\omega|\mathbf{b}_0, \mathbf{b}_1, \sigma_{1.0}^2, \sigma_u^2 \sim N \left\{ \left[\mathbf{C}_b^\top \mathbf{C}_b + \lambda_u \mathbf{D} \right]^{-1} \mathbf{C}_b^\top \mathbf{b}_1, \sigma_{1.0}^2 \left[\mathbf{C}_b^\top \mathbf{C}_b + \lambda_u \mathbf{D} \right]^{-1} \right\}$
4. Sample from $\mathbf{b}_0|\boldsymbol{\theta}_0, \mathbf{y}, \mathbf{E}, \mathbf{X}, \mathbf{W}$ via Metropolis-Hastings; see below.
5. $\mathbf{b}_1|\mathbf{b}_0, \omega, \alpha,$
 $\sigma_\varepsilon^2, \sigma_{1.0}^2, \mathbf{y}, \mathbf{E}, \mathbf{W} \sim N \left[\left(\mathbf{Z}_1^\top \mathbf{Z}_1 + \lambda_1 \mathbf{I}_n \right)^{-1} \left[\mathbf{Z}_1^\top (\mathbf{y} - \mathbf{Z}_0 \mathbf{b}_0 + \mathbf{W} \alpha) + \lambda_1 \mathbf{C}_b \omega \right], \sigma_\varepsilon^2 \left(\mathbf{Z}_1^\top \mathbf{Z}_1 + \lambda_1 \mathbf{I}_n \right)^{-1} \right]$
6. $\sigma_\varepsilon^2|\mathbf{b}_0, \mathbf{b}_1, \alpha, \mathbf{y}, \mathbf{E}, \mathbf{W} \sim IG(A_\varepsilon + N/2, \{1/B_\varepsilon + \|\mathbf{y} - (\mathbf{Z}_0 \mathbf{b}_0 + \mathbf{Z}_1 \mathbf{b}_1 + \mathbf{W} \alpha)\|\}^{-1})$
7. $\sigma_0^2|\beta, \mathbf{b}_0, \mathbf{X} \sim IG(A_0 + n/2, \{1/B_0 + (1/2) \|\mathbf{b}_0 - \mathbf{X} \beta\|\}^{-1})$
8. $\sigma_{1.0}^2|\mathbf{b}_0, \mathbf{b}_1, \omega \sim IG(A_{1.0} + n/2, \{1/B_{1.0} + (1/2) \|\mathbf{b}_1 - \mathbf{C}_b \omega\|\}^{-1})$
9. $\sigma_u^2|\omega \sim IG(A_u + K/2, \{1/B_u + (1/2) \|\mathbf{u}\|\}^{-1})$

Because the full conditionals for the b_{0i} 's do not have closed form, we use a random walk Metropolis-Hastings (Chib and Greenberg 1993) step to sample these random intercepts. We generate a candidate value from a normal distribution with mean equal to the current value of b_{0i} and variance equal to a scaling factor τ times the variance estimate of b_{0i} from the analogous LMM. Because it is typically more efficient to overestimate the posterior variance of b_{0i} when generating candidate values, in our simulations and data analysis we set $\tau = 4$. We then use the standard Metropolis-Hastings rejection step to decide whether to accept the candidate value in the current iteration of the chain.

The resulting Gibbs sampler samples from each of these full conditional distributions in the order presented. For both the simulations and data analysis, we initialized each chain using the parameter estimates from the analogous LMM fit. We used 1000 iterations as burn-in, and kept 500 iterations for posterior sample processing. We assessed the mixing of the proposed Metropolis-within-Gibbs algorithm for a random sample of simulated data sets as well as for the blood flow data, and observed proper mixing in all cases. As a second check, we re-ran a sample of the simulated datasets using twice the number of Gibbs samples and observed no change in parameter estimates.

We note that one might also consider integrating out the random slopes to reduce the dimensionality of the model, which could potentially improve the mixing properties of the resulting Markov chain. We investigated this option, but this alternative approach did not significantly improve upon the original Gibbs sampler outlined above, primarily because the mixing properties of this original chain perform quite well without any modification. R functions for implementing this sampler are available from the journal's website.

The functional slopes model is a natural extension of the Laird-Ware LMM in that it reduces to the LMM in the special case when f is linear. In the penalized spline formulation, this corresponds to $\sigma_u^2 = 0$. Thus, one can evaluate the evidence of a nonlinear association, and hence lack of fit of the LMM, by evaluating the evidence that $\sigma_u^2 \neq 0$. Because formal

assessment of this condition is an open problem even in standard Bayesian penalized spline models, we follow the recommendations of Ruppert, Wand, and Carroll and assess this question using the posterior distribution of σ_u^2 .

3. Simulation Study

We performed simulations to compare the ability of the proposed functional slopes model to accurately estimate the functional relationship between subject-specific intercepts and slopes, as well as the loss of efficiency of our approach when the random effects are truly multivariate normal. Because we are not aware of any existing methods that yield a direct nonparametric estimate of this relationship, we compared the performance of the proposed approach to two simpler ad-hoc approaches that yield smoothed estimates of this relationship, and are hence competitors to the proposed approach. The first competitor, which we denote GAM-LMM, is a two-stage approach that fits a linear mixed model with random intercepts and slopes in a first stage, and fits a penalized regression spline model to the resulting BLUPs with smoothing parameter selected via generalized cross-validation. The second competitor, which we denote GAM-FIXED, is similar to the first, except the subject-specific effects in the first stage are treated as fixed effects, which avoids smoothing the resulting estimates towards a bivariate normal prior for the random effects. We fit the second stage GAMs in these two approaches using the gam function from the mgcv package in R.

For each scenario, we generated 100 simulated data sets, with $T_i = T = 5$, $E_{it} \sim N(0, 0.6)$, $b_{0i} \sim N(0, 0.8)$, $\eta_i \sim N(0, 0.1)$, and $\varepsilon_{it} \sim N(0, 0.3)$. We chose four functional forms for f that represent realistic functional relationships between overall level and susceptibility in a variety of biomedical contexts (Figure 2): (a) Symmetric: $f(b) = \phi(b/0.6)$, where ϕ is a normal pdf, (b) Multi-modal $f(b) = 2.5\phi(|x/3| - 0.35)/0.15$, (c) Sinusoidal: $f(b) = \sin(b\pi/1.8)$, and (d) Linear: $f(b) = 0.5b$. Scenario (a) represents a setting in which individuals with moderate levels have high susceptibility, whereas individuals with extreme levels have low susceptibility. Scenario (b) represents a setting with high susceptibility among individuals with moderately high or low levels, and low susceptibility among individuals with extreme or average levels. Scenario (c) implies different biologic mechanisms for individuals with moderately high and low levels, with exposure among subjects with low levels driving the response lower, exposure among subjects with high levels driving the response higher, and no susceptibility among individuals with either extreme or average levels. Web Figures 1–4 in the supplementary material show a generated set of random intercepts and slopes for scenarios (a)–(d), respectively. These figures also show plots of individual profiles from one simulated dataset from a given scenario, with these plots generated assuming a binary exposure variable and $n = 75$ for clarity of presentation. We also ran two additional simulations, using the estimated parameters obtained from the Bayesian functional slopes and LMM fits of the ischemia data described in Section 1. To match the application as closely as possible, these last two simulations used $n = 109$, the number of tissue pieces in the application.

For each method, we computed the mean squared bias (Bias²) and mean squared error (MSE) of the f estimators over a grid defined by 100 evenly-spaced points in the interval

from $[-2\sigma_0, 2\sigma_0]$, or approximately 95% of the range of b_{0i} . These results, shown in Table 1, suggest that for the scenarios in which the true functional form between subject-specific level and susceptibility is nonlinear, the proposed functional slopes model significantly outperforms the two simpler approaches, both in terms of MSE and bias. Depending on the true f , relative to the GAM-LMM approach the functional slopes model offers an improvement in MSE of between 10%–35% for $n = 100$ and 50%–60% for $n = 200$, and even more for squared bias. In general, for scenarios (a)–(c), the parametric distributional assumption in the first stage of the GAM-LMM approach results in the largest bias among the three approaches, but a variance smaller than its fixed counterpart and hence smaller MSE. Web Figure 5 in the supplementary material illustrates a typical result for a single randomly selected dataset generated under scenario (a). The procedure based on the LMM BLUPs shrinks the estimate towards linearity, the GAM based on the fixed effect estimates does somewhat better, and the functional slopes formulation improves upon this latter estimate by correcting for the measurement error associated with the estimated intercepts. For the application-specific simulation, the gains obtained by the functional slopes model are similar, with MSE values equal to 1.70, 3.32, and 3.62 and Bias² values equal to 0.33, 3.09, and 3.12 for the functional slopes, GAM-LMM, and GAM-Fixed approaches, respectively

Upon re-running the simulations using the variance component priors proposed by Berry et al. (2002) for a general nonparametric regression problem with measurement error, we observed the same ordering among the three methods for all scenarios. Taken together, these simulation results suggest that the functional slopes model is preferable to existing subject-specific models when scientific interest focuses on the functional form of the association between level and susceptibility and this association is nonlinear.

4. Data Analysis

Here we analyze the cardiac blood flow data first considered by Bartoli et al. (2009) and described in Section 1. The costs of these microsphere experiments are prohibitive, so the methodology was applied to a small number of animals. In this analysis we use data from $n = 109$ tissue pieces from two laboratory animals. We consider the longitudinal responses defined by the piece-specific values

$$y_{it} = (\text{Occlusion Blood Flow} - \text{Baseline Blood Flow})_{it}, \quad i = 1, \dots, 109, \quad t = 1, \dots, 6.$$

In preliminary analyses, we inspected scatterplots of the fixed-estimates of the intercepts and slopes from the standard model

$$Y_{it} = b_{0i} + b_{1i}E_{it} + \varepsilon_{it},$$

constructed for each animal separately. Observing little difference in the form of this relationship in the two animals, we fit a model that allowed the mean of the intercept to

differ by animal but for the intercept-slope relationship to be constant across animal. That is, we fit the functional slopes model

$$Y_{it} = b_{0i} + [f(b_{0i}) + \eta_i] E_{it} + \varepsilon_{it}, \quad (6)$$

where $b_{0i} \sim N(\beta_0 + \beta_1 \text{Animal}_i, \sigma_0^2)$. Here, Animal_i is an animal indicator that accounts for the fact that multiple tissue pieces are taken from the same animal, E_{it} is a binary indicator of particle (vs. filtered air) exposure during occlusion, and

$$\eta_i \stackrel{iid}{\sim} N(0, \sigma_{\eta}^2), \quad \varepsilon_{it} \stackrel{iid}{\sim} N(0, \sigma_{\varepsilon}^2), \quad b_{0i} \perp \eta_i \perp \varepsilon_{it}.$$

In this formulation, b_{0i} represents the difference between the average occlusion blood flow under filtered air conditions and baseline blood flow, which as described in Section 1 is thought to reflect the degree of ischemia for a given piece. In this framework, the analysis presented in Bartoli et al. (2009) corresponds to estimation of f under the assumption that this function is a step function, with the step defined by dichotomizing the sufficient statistic for b_{0i} , $s_i = (1/\mathcal{F}_i) \sum_{t \in \mathcal{F}_i} y_{it}$ at 0, where $\mathcal{F}_i$ denotes the set of outcomes y_{it} corresponding to filtered air exposure for subject i . Model (6) relaxes this assumption and estimates f nonparametrically.

The first panel of Figure 3 shows the posterior mean of f , and the corresponding 95% credible interval for this estimate, from the fit of model (6) overlaid on the fixed effect estimates presented in Figure 1. This estimate suggests that the effects of exposure occur in those pieces that are most ischemic under control conditions, with pieces that are not severely ischemic showing little effect of exposure. This result suggests that pollution exposure may not only make the myocardial region affected by coronary artery disease larger, but may also further restrict flow in the already affected region. This finding suggests that subjects with existing heart disease may be a subpopulation particularly susceptible to particle exposure. A 95% credible interval for the variance component σ_u^2 , which represents the smoothing parameter for f , is (0.06, 0.55). Being bounded away from zero, this interval provides some evidence that the functional slopes model fits the data better than the standard linear mixed model.

For comparison, we also applied the *ad-hoc* two-stage approaches that fit a penalized spline model to estimates of the slopes, either fixed effects or BLUPs from a linear mixed model, as a function of the corresponding estimates of the intercepts. Because there were missing data for some pieces, we fit weighted penalized spline models, using as weights the inverse variance estimates of the slope estimates from the first stage. Shown in the second and third panels of Figure 3, the results illustrate those patterns observed in the simulation study. The two-stage approaches do not account for the measurement error associated with the estimated slopes, presumably leading to biased estimate of f . It appears that the naive approaches oversmooth the functional form for f , which is in line with what is typically observed in other nonparametric regression measurement error problems (Berry et al. 2002).

Just as importantly, the unified functional slopes model appears to adequately propagate the uncertainty associated with the first-stage estimates forward in the estimation of the intercept-slope association, whereas the two-stage estimates yield unrealistically narrow confidence bands. In terms of the conclusions of the study, using the functional slopes model one would conclude that there is no effect of particle exposure on those pieces with little to no pre-existing ischemia, whereas the naive two-stage approaches suggest exposure increases flow in such pieces. This latter conclusion is less biologically plausible than the former.

Finally, we compared the estimates of μ_1 , the mean of the random slopes, and the predicted random effects from the functional slopes and the linear mixed models. The estimates of μ_1 were not appreciably different, with the standard LMM yielding $\mu_1 = 0.03$ (S.E. = 0.02) and the functional slopes model yielding $\mu_1 = 0.02$ (posterior S.D. = 0.04). This suggests that on average, for these two animals, there was no exposure effect averaged over all tissue pieces. Rather, exposure effects seem to appear only in ischemic tissue. Web Figure 6 in the supplementary material shows scatterplots of the predicted intercepts and predicted slopes from the two models. The plots show that the predicted intercepts from the two models do not appear to systematically differ from one another. Compared to the functional slopes model, the LMM shrinks the extreme negative slopes back towards the overall mean slightly more aggressively. This over-shrinkage is to be expected when the true distribution of the slopes is skewed (Coull et al. 2001).

5. Discussion

To our knowledge, there are no good methods available when scientific interest focuses on *direct* estimation of a potentially nonlinear association between baseline and susceptibility in a mixed model setting. We argue that simple approaches, such as dichotomizing the average response and existing mixed model formulations, can place somewhat strong assumptions on this relationship that can obscure the true relationship. Recently proposed methods that place nonparametric assumptions on the multivariate distribution of the random effects do not yield a direct estimate of this relationship. The proposed functional slopes model can avoid these issues since it does not shrink this relationship towards linearity and properly accounts for all sources of uncertainty in the estimation process.

Although we developed the methods with the primary goal of relating susceptibility to an environmental exposure to overall response level, another consequence of our chosen approach is the relaxation of the multivariate normal assumption for set of random effects. For instance, Figure 4 shows the contour plot of the bivariate density for the intercepts and slopes, as well as the marginal density of the slopes, implied by the estimate of f in the coronary blood flow example. Analogous plots for the three simulation scenarios corresponding to a nonlinear f appear in Web Figures 7–9 in the supplementary material. Taken together, these plots indicate that the model can accommodate a wide range of marginal distributions for the slopes, including multi-modal and skewed distributions. These plots also demonstrate that a nonlinear form for the intercept-slope association does not necessarily imply a strongly non-normal distribution for the slopes. However, because nonparametric estimation of the random effects distribution is not the primary motivation of

the method, we do not consider the performance of the method in this regard here. Indeed, there already exist many methods for both testing and relaxing the normality assumptions in linear mixed models (for instance, Zhang and Davidian 2001), and more research is necessary to determine whether this approach offers any advantages over existing nonparametric approaches.

Although the application we considered involves a binary exposure, the general principles introduced here apply more broadly to settings in which interest focuses on the relationship between baseline and change over time, or between baseline and the effect of some other time-varying continuous covariate. However, in this more general setting, additional subtleties arise. For instance, in this case one should take care in centering the time-varying covariate (Edland 2000). Moreover, incorrectly assuming a linear form for the effect of time can spuriously induce correlation between intercepts and slopes (Rabinowitz and Shea 1997). Thus, in the general case, it is important to diagnose any lack of fit of the underlying linear mixed effects model before extending it to include functional slopes. The diagnostics proposed by Rabinowitz and Shea (1997) are likely to be useful for this purpose. Moreover, because the application considered here was a randomized toxicology study, there were no additional time-varying confounders to consider. In the general case, it would be important to check whether such effects should be modeled as fixed effects constant across subjects or subject-specific random effects (Kinney and Dunson 2007).

We focused on the situation in which the random slopes mean is a nonlinear function of random intercepts. The methodology should transfer easily if interest focuses on estimation of a nonlinear relationship between another pair of random effects. Similarly, extension to the case where more than one set of random effects depends nonlinearly on the intercepts is also straightforward.

There are other potentially interesting directions for future research. A natural direction is the extension of the proposed functional slopes framework to the generalized case for non-normal responses. Interesting questions include how well can one estimate the form of the relationship between level and susceptibility when one has binary outcomes, which contain much less information than continuous endpoints. Model fitting is also likely to be more challenging, as the development of MCMC algorithms with good mixing properties in the generalized additive framework is delicate (Zhao et al. 2006). Latent Gaussian model formulations (Albert and Chib 1993) and auxiliary mixture sampling strategies (Frühwirth-Schnatter et al. 2008) may be useful in this setting.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This research was supported by US National Institute of Environmental Health Sciences grants ES012044 and ES012972, and US Environmental Protection Agency grant R832416. The author thanks the Associate Editor and three referees for insightful comments that greatly improved the manuscript. The author also thanks John Godleski for use of the microsphere data, Chris Paciorek and Eric Tchetgen for helpful discussions, and Louise Ryan for thoughtful comments on a previous draft of the manuscript.

References

- Albert JH, Chib S. Bayesian analysis of binary and polychotomous response data. *Journal of the American Statistical Association*. 1993; 88:669–679.
- Bartoli CR, Akiyama I, Wellenius GA, Coull BA, Diaz EA, Lawrence J, Okabe K, Verrier RL, Godleski JJ. Acute inhalation of particulate air pollution causes inappropriate coronary vasoconstriction during ongoing myocardial ischemia. *Environmental Health Perspectives*. 2009; 117:333–337. [PubMed: 19337504]
- Berry SM, Carroll RJ, Ruppert D. Bayesian smoothing and regression splines for measurement error problems. *Journal of the American Statistical Association*. 2002; 97:160–169.
- Carlin, BP.; Louis, RA. *Bayes and Empirical Bayes Methods for Data Analysis*. 2. Boca Raton: Chapman and Hall/CRC Press; 2000.
- Carroll RJ, Ruppert D, Crainiceanu CM, Tosteson TD, Karagas MR. Nonlinear and nonparametric regression and instrumental variables. *Journal of the American Statistical Association*. 2004; 99:736–750.
- Chib S, Greenberg E. Understanding the Metropolis-Hastings algorithm. *The American Statistician*. 1995; 49:327–335.
- Coull BA, Hobert JP, Ryan LM, Holmes LB. Crossed random effect models for multiple outcomes in a study of teratogenesis. *Journal of the American Statistical Association*. 2001; 96:1194–1204.
- Crainiceanu CM, Ruppert D, Wand MP. Bayesian analysis for penalized spline regression using WinBUGS. *Journal of Statistical Software*. 2005; 14:1548–766.
- Edland SD. Blomqvist revisited: how and when to test the relationship between level and longitudinal rate of change. *Statistics in Medicine*. 2000; 19:1441–1452. [PubMed: 10844709]
- Frühwirth-Schnatter S, Frühwirth R, Held L, Rue H. Improved Auxiliary Mixture Sampling for Hierarchical Models of Non-Gaussian Data. *Statistics and Computing*. 2008; 18:1007–1022. [PubMed: 1891094]
- Gelfand AE, Smith AFM. Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association*. 1990; 85:398–409.
- Kinney SK, Dunson DB. Fixed and random effects selection in linear and logistic models. *Biometrics*. 2007; 63:690–698. [PubMed: 17403104]
- Laird NM, Ware JH. Random effects models for longitudinal data. *Biometrics*. 1982; 38:963–974. [PubMed: 7168798]
- Li E, Zhang D, Davidian M. Conditional estimation for generalized linear models when covariates are subject-specific parameters in a mixed model for longitudinal measurements. *Biometrics*. 2004; 60:1–7. [PubMed: 15032767]
- Neuhaus JM, McCulloch CE. Separating between- and within-cluster covariate effects by using conditional and partitioning methods. *Journal of the Royal Statistical Society, Series B*. 2006; 68:859–872.
- Rabinowitz D, Shea S. Random effects analysis of children's blood pressure data. *Statistical Science*. 1997; 12:185–194.
- Ruppert, D.; Wand, MP.; Carroll, RJ. *Semiparametric Regression*. Cambridge, UK: Cambridge University Press; 2003.
- Ruppert D, Wand MP, Carroll RJ. Semiparametric regression during 2003–2007. *Electronic Journal of Statistics*. 2009; 3:1193–1256. [PubMed: 20305800]
- Spiegelhalter, D.; Thomas, A.; Best, N. MRC Biostatistics Unit. Institute of Public Health; Cambridge: 2000. WinBUGS Version 1.3. User's Manual. <http://www.mrc-bsu.cam.ac.uk/bugs>
- Wall MM, Amemiya Y. Estimation for polynomial structural equation models. *Journal of the American Statistical Association*. 2000; 95:929–940.
- Wang CY, Wang N, Wang S. Regression analysis when covariates are regression parameters of a random effects model for observed longitudinal measurements. *Biometrics*. 2000; 56:487–495. [PubMed: 10877308]
- Zhang D, Davidian M. Linear mixed models with flexible distributions of random effects for longitudinal data. *Biometrics*. 2001; 57:795–802. [PubMed: 11550930]

Zhao Y, Staudenmayer J, Coull BA, Wand MP. Towards general design Bayesian generalized linear mixed models. *Statistical Science*. 2006; 21:35–51.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Figure 1. Difference in blood flow means under exposed and control conditions as a function of flow change under control conditions, with ± 1 standard error intervals plotted for both sets of estimates.

Figure 2.
Simulation scenarios for the intercept-slope association.

Figure 3. Results of blood flow analysis. Left panel shows functional slopes estimate. Middle and right panels show estimates from the weighted GAM-Fixed and GAM-LME analyses, respectively. Points represent fixed effect estimates of the intercepts and slopes.

Figure 4. Estimated bivariate density for the intercepts and slopes (left) and marginal density of the slopes (right) obtained from the fit of the functional slopes model to the coronary blood flow data.

Table 1

Simulated MSE and Bias² for f estimators ($\times 10^{-2}$)

Function	Method	$n = 100$		$n = 200$	
		MSE	Bias ²	MSE	Bias ²
M-Modal	FSM	6.25	1.12	2.86	0.86
	GAM-LMM	6.93	5.31	5.67	4.88
	GAM-Fixed	7.11	3.24	4.24	2.41
Norm	FSM	3.91	0.14	1.76	0.06
	GAM-LMM	4.66	3.49	3.52	2.94
	GAM-Fixed	3.97	1.11	2.03	0.68
Sine	FSM	5.52	0.61	2.47	0.25
	GAM-LMM	8.46	5.64	6.33	4.87
	GAM-Fixed	6.71	2.87	3.74	1.97
Linear	FSM	4.64	0.03	2.00	0.07
	GAM-LMM	0.94	0.00	0.50	0.00
	GAM-Fixed	2.31	0.21	1.35	0.32

Lowest value per scenario in bold.
Simulation SE ≈ 0.02 .