

A BAYESIAN TIME-VARYING EFFECT MODEL FOR BEHAVIORAL MHEALTH DATA

BY MATTHEW D. KOSLOVSKY¹, EMILY T. HÉBERT^{2,*}, MICHAEL S. BUSINELLE^{2,†} AND
MARINA VANNUCCI³

¹*Department of Statistics, Colorado State University, matt.koslovsky@colostate.edu*

²*Oklahoma Tobacco Research Center, University of Oklahoma Health Sciences Center, *Emily-Hebert@ouhsc.edu;
†Michael-Businelle@ouhsc.edu*

³*Department of Statistics, Rice University, marina@rice.edu*

The integration of mobile health (mHealth) devices into behavioral health research has fundamentally changed the way researchers and interventionists are able to collect data as well as deploy and evaluate intervention strategies. In these studies, researchers often collect intensive longitudinal data (ILD) using ecological momentary assessment methods which aim to capture psychological, emotional and environmental factors that may relate to a behavioral outcome in near real time. In order to investigate ILD collected in a novel, smartphone-based smoking cessation study, we propose a Bayesian variable selection approach for time-varying effect models, designed to identify dynamic relations between potential risk factors and smoking behaviors in the critical moments around a quit attempt. We use parameter-expansion and data-augmentation techniques to efficiently explore how the underlying structure of these relations varies over time and across subjects. We achieve deeper insights into these relations by introducing nonparametric priors for regression coefficients that cluster similar effects for risk factors while simultaneously determining their inclusion. Results indicate that our approach is well positioned to help researchers effectively evaluate, design and deliver tailored intervention strategies in the critical moments surrounding a quit attempt.

1. Introduction.

1.1. *Scientific background.* The integration of mobile health (mHealth) devices into behavioral health research has fundamentally changed the way researchers and interventionists are able to collect data as well as deploy and evaluate intervention strategies. Leveraging mobile and sensing technologies, just-in-time adaptive interventions (JITAI) or ecological momentary interventions are designed to provide tailored support to participants based on their mood, affect and socioenvironmental context (Heron and Smyth (2010), Nahum-Shani et al. (2017)). In order to deliver theory-based interventions at critical moments, researchers collect intensive longitudinal data using ecological momentary assessment (EMA) methods which aim to capture psychological, emotional and environmental factors that may relate to a behavioral outcome in near real time. In practice, JITAI's effectiveness depends on accurately identifying high-risk situations by the user or by predetermined decision rules to initiate the delivery of intervention components. Decision rules for efficacious interventions rely on a thorough understanding of the factors that characterize a subject's risk for a behavioral outcome, the dynamics of these risk factors' relation with the outcome over time and the knowledge of possible strategies to target a risk factor (Nahum-Shani et al. (2017)).

Received August 2020; revised September 2020.

Key words and phrases. Ecological momentary assessment, mHealth, Pólya-Gamma augmentation, time-varying effect model, variable selection.

In the analysis of this paper, we investigate a behavioral health intervention study that targets smoking cessation. Historically, smoking cessation studies have used health behavior theory (Shiffman et al. (2002), Timms et al. (2013)) or group-level trends of smoking antecedents (Piasecki et al. (2013)) to determine when a JITAI should be triggered. However, this approach is limited since current health behavior models are inadequate for guiding the dynamic and granular nature of JITAIs (Riley et al. (2011), Klasnja et al. (2015)). Additionally, the design of efficacious smoking cessation interventions is challenged by the complexity of smoking behaviors around a quit attempt and misunderstandings of the addiction process (Piasecki et al. (2002)). More recently, smoking behavior researchers have capitalized on the ability of mHealth techniques to collect rich streams of data capturing subjects' experiences close to their occurrence at a high temporal resolution. The structure as well as the complexity of these data provide unique opportunities for the development and implementation of more advanced analytical methods compared to traditional longitudinal data analysis methods used in behavioral research (e.g., mixed models, growth curve models) (Trail et al. (2014)). For example, researchers have applied reinforcement learning (Luckett et al. (2020)) and dynamic systems approaches (Trail et al. (2014), Rivera, Pew and Collins (2007), Timms et al. (2013)) to design and assess optimal treatment strategies using mHealth data. Additionally, Koslovsky et al. (2018), de Haan-Rietdijk et al. (2017) and Berardi et al. (2018) have applied hidden and observed Markov models to study transitions between discrete behavioral states, Shiyko et al. (2012) and Dziak et al. (2015) have used mixture models to identify latent structures and Kürüm et al. (2016) have employed joint modeling techniques to study the complexity of smoking behaviors.

Greater insights into the dynamic relation between risk factors and smoking behaviors have been generated by the application of functional data techniques (Trail et al. (2014), Vasilenko et al. (2014), Koslovsky et al. (2018), Tan et al. (2012)). These methods are well suited for high-dimensional data with unbalanced and unequally-spaced observation times, matching the format of data collected with EMAs. They also require little assumptions on the structure of the relations between risk factors and behavioral outcomes. One popular approach uses varying-coefficient models which belong to the class of generalized additive (mixed) models. These semiparametric regression models allow a covariate's corresponding coefficient to vary as a smooth function of other covariates (Hastie and Tibshirani (1993)). For example, Selya et al. (2015) examined how the relation between the number of cigarettes smoked during a smoking event and smoking-related mood changes varies as a function of nicotine dependence. More frequently, penalized splines have been employed in varying-coefficient models to investigate how the effect of a covariate varies as a function of time, leading to time-varying effect models (TVEM) (Tan et al. (2012), Lanza et al. (2014), Koslovsky et al. (2018), Shiyko et al. (2012), Mason et al. (2015), Vasilenko et al. (2014)). These approaches allow researchers to identify the critical moments that a particular risk factor is strongly associated with smoking behaviors, information that can be used to design tailored intervention strategies based on a subject's current risk profile.

1.2. Model overview. While there are various inferential challenges that functional data analysis models can address, in the application of this paper we focus on incorporating three recurring themes in behavioral research to explore the relations between risk factors and smoking behaviors:

1. *Model Assumptions:* Numerous smoking behavior research studies have relied on semi-parametric, spline-based methods to learn the relational structure between risk factors and outcomes (Tan et al. (2012), Vasilenko et al. (2014)).

2. *Variable Selection:* One of the main objectives of intensive longitudinal data analysis is to identify or reaffirm complex relations between risk factors and behavioral outcomes over time (Walls and Schafer (2006)).

3. *Latency*: A common aim in smoking behavior research studies is to identify latent structure in the data, such as groups or clusters of subjects with similar smoking behaviors, over time (McCarthy et al. (2016), Cursio, Mermelstein and Hedeker (2019), Geiser et al. (2013), Dziak et al. (2015), Brook et al. (2008)).

To incorporate and expand upon these features in our analysis, we develop a flexible Bayesian varying-coefficient regression modeling framework for longitudinal binary responses that uses variable selection priors to provide insights into the dynamic relations between risk factors and outcomes. We embed spike-and-slab variable selection priors as mixtures of a point mass at zero (spike) and a diffuse distribution (slab) (George and McCulloch (1993), Brown, Vannucci and Fearn (1998)) and adopt the formulation of Scheipl, Fahrmeir and Kneib (2012) to deconstruct the varying-coefficients' terms, in our case time-varying effects, into a main effect, linear interaction term and nonlinear interaction term. Unlike previous approaches in behavioral health research that use time-varying effect models, our formulation allows us to gain inference on whether a given risk factor is related to the smoking behavior while also learning the type of relation. Additionally, by performing selection on fixed as well as random effects, our method is equipped to identify relations that vary over time and across subjects. For this, we exploit a Pólya-Gamma augmentation scheme that enables efficient sampling without sacrificing interpretability of the regression coefficients as log odds ratios (Polson, Scott and Windle (2013)). Furthermore, we adopt a Bayesian semiparametric approach to model fixed and random effects by replacing the traditional spike-and-slab prior with a nonparametric construction to cluster risk factors that have similar strengths of association.

1.3. *Just-in-time adaptive interventions for smoking abstinence*. Although multiple studies have examined momentary predictors of smoking lapse (Shiffman et al. (2000), Piasecki et al. (2003), Businelle et al. (2014)), JITAIs for smoking cessation are still nascent. Thus far, studies have used participant-labeled GPS coordinates to trigger supportive messages to prevent smoking (Naughton et al. (2016)) or have tailored messages to the duration and intensity of participant's self-reported side effects while taking varenicline (McClure et al. (2016)). Using our proposed approach, we analyze ILD collected in a study investigating the utility of a novel, smartphone-based smoking cessation JITAI (*SmartT*). The *SmartT* intervention (Businelle et al. (2016)) uses a lapse risk estimator to identify moments of heightened risk for lapse and tailors treatment messages in real time based upon the level of imminent smoking lapse risk and currently present lapse triggers. To our knowledge, no other studies have used EMA data to estimate risk for imminent smoking lapse and deliver situation-specific, individually-tailored treatment content prior to lapse.

In this study, adult smokers ($N = 81$) recruited from a smoking cessation research clinic were randomized to the *SmartT* intervention, the National Cancer Institute's QuitGuide (*NCI QuitGuide*) or weekly counseling sessions (*usual care*) and followed over a five-week period spanning one week prior to a scheduled quit attempt to four weeks after. At the beginning of the assessment period, baseline measures were collected, and subjects were shown how to complete EMAs on a study-provided smartphone. Throughout the assessment period, subjects completed daily diaries and received four random EMAs from the smartphone to complete each day. For each EMA, subjects were prompted on their recent smoking behaviors, alcohol consumption as well as various questions regarding their current psychological, social and environmental factors that may contribute to an increased risk of smoking behaviors.

Findings indicate that our approach is well positioned to help researchers evaluate, design and deliver tailored intervention strategies in the critical moments surrounding a quit attempt. In particular, results confirm previously identified temporal relations between smoking behaviors around a quit attempt and risk factors. They also indicate that subjects differ

in how they respond to different risk factors over time. Furthermore, we identify clusters of active risk factors that can help researchers prioritize intervention strategies based on their relative strength of association at a given moment. Importantly, our approach generates these insights with minimal assumptions regarding which risk factors were related to smoking in the presence of others, the structural form of the relation for active terms or the parametric form of regression coefficients.

The rest of the paper is organized as follows. In Section 2 we present our modeling approach and describe prior constructions. In Section 3 we investigate the relations between risk factors and smoking behaviors in the critical moments surrounding a scheduled quit attempt using mHealth data. In Section 4 we conduct a simulation study investigating the variable selection and clustering performance of our proposed method on simulated data. In Section 5 we evaluate prior sensitivity of our model. In Section 6 we provide concluding remarks.

2. Methods. The objective of our analysis is to identify relations between a set of risk factors (i.e., baseline and EMA items) and a binary outcome (i.e., momentary smoking) repeatedly collected over time. For this, we employ a Bayesian variable selection framework that allows a flexible structure for the unknown relations. We achieve this by performing selection not only on main effects but, additionally, on linear and nonlinear interaction terms as well as random effects. In this work we refer to fixed and random effects in the context of hierarchical or multilevel models, where fixed effects are constant across subjects and random effects differ at the subject level. We chose this terminology based on its familiarity within both frequentist and Bayesian paradigms but point out that the fixed or population-level effects are treated as random variables in our model and thus follow a probability distribution.

2.1. *A varying-coefficient model for intensive longitudinal data collected with EMAs.* Let $y_{ij} \in \{0, 1\}$ represent momentary smoking for subject $i = 1, \dots, N$ and \mathbf{x}_{ij} and \mathbf{z}_{ij} represent P - and D -dimensional vectors of risk factors collected on each subject at time $j = 1, \dots, n_i$, respectively. To maintain temporality in our particular application (see Section 3 for more details), we model the relation between momentary smoking by the next assessment and current, potential risk factors as a varying-coefficient model of the type

$$(2.1) \quad \text{logit}(P(y_{i,j+1} = 1 | \mathbf{x}_{ij}, \mathbf{z}_{ij}, u_{ij})) = \sum_{p=1}^P f_p(u_{ij})x_{ijp} + \boldsymbol{\alpha}'_i \mathbf{z}_{ij},$$

where $f_p(u)$ are smooth functions of a scalar covariate u and $\boldsymbol{\alpha}_i$ represents subject specific random effects. Similar temporal assumptions have been made previously in smoking behavior research studies (Bolman et al. (2018), Minami et al. (2014), Shiffman et al. (1996, 2013), Shiyko et al. (2014)). Note that, in general, researchers may use the framework of (2.1) to model the relation between a binary outcome and potential risk factors collected concurrently, in addition to lagged trends, as is typical in longitudinal studies (Fitzmaurice, Laird and Ware (2012)). With this formulation, we include varying-coefficient terms for each of the P risk factors based on u . However, in general, we can specify varying-coefficient terms that depend on $u' \neq u$, and, thus, the number of varying-coefficient terms in the full model is not strictly P . If u is chosen to represent time, then this model is commonly referred to as a *time-varying effect model* in smoking behavior research (Tan et al. (2012), Vasilenko et al. (2014), Dziak et al. (2015), Koslovsky et al. (2018)). Note that \mathbf{z}_{ij} is typically a subset of \mathbf{x}_{ij} (Kinney and Dunson (2007), Cheng et al. (2010), Hui, Müller and Welsh (2017)) and that incorporating a 1 in \mathbf{x}_{ij} and \mathbf{z}_{ij} allows for an intercept term that varies as a function of u and a random intercept term, respectively. Additionally, this formulation can handle time-invariant risk factors, such as baseline items, by fixing x_{ijp} (z_{ijd}) to x_{ip} (z_{id}) for all observations j .

We approximate the smooth functions with spline basis functions. Specifically,

$$(2.2) \quad f_p(u_{ij}) = \mathbf{U}'_{ij} \boldsymbol{\phi}_p,$$

where \mathbf{U}_{ij} is a spline basis function for u_{ij} and $\boldsymbol{\phi}_p$ is a r_p -dimensional vector of corresponding spline coefficients. For simplicity, the splines are constructed with an equal number of equally spaced knots that depend on the minimum and maximum values of \mathbf{u} .

2.2. Penalized priors for the spline coefficients. Using a combination of variable selection and shrinkage priors, our approach generates insights on the underlying structure of the smooth functions by reconstructing them as the summation of main effect, linear interaction and nonlinear interaction components. Formally, we rewrite equation (2.2) as

$$(2.3) \quad f_p(u_{ij}) = \beta_p^* \mathbf{U}'_{ij} \boldsymbol{\xi}_p + \beta_p^\circ u_{ij} + \beta_{0p},$$

where the constant term β_{0p} captures the main effect of \mathbf{x}_p , β_p° represents the effect of the linear interaction between \mathbf{u} and \mathbf{x}_p and $\beta_p^* \boldsymbol{\xi}_p$ is a parameter-expanded vector of coefficients corresponding to the nonlinear interaction term.

To derive the nonlinear component in equation (2.3), we start by penalizing the spline functions in equation (2.2) with a second-order Gaussian random walk prior following

$$(2.4) \quad \mathbf{U} \boldsymbol{\phi}_p | s^2 \sim N(\mathbf{0}, s^2 \mathbf{U} \mathbf{P}^{-1} \mathbf{U}'),$$

where \mathbf{U} is a $\sum_{i=1}^N (n_i - 1) \times r_p$ -dimensional matrix with each row corresponding to \mathbf{U}'_{ij} for the i th subject at the j th assessment, s^2 controls the amount of smoothness and \mathbf{P} is the appropriate penalty matrix (Lang and Brezger (2004)). Next, we take the spectral decomposition of $\mathbf{U} \mathbf{P}^{-1} \mathbf{U}' = [\mathbf{U}_+ \ \mathbf{U}_\circ] \begin{bmatrix} \mathbf{V}_+ & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{U}_+ \\ \mathbf{U}_\circ \end{bmatrix}$, where \mathbf{U}_+ is a matrix of eigenvectors with corresponding positive eigenvalues along the diagonal of matrix \mathbf{V}_+ , and \mathbf{U}_\circ are the eigenvectors associated with the zero eigenvalues. Now, we can redefine the smooth functions in equation (2.2) as the sum of nonlinear (penalized) interaction, linear (nonpenalized) interaction and main effect terms as presented in equation (2.3), where the penalized term is written as $\mathbf{U}^* \boldsymbol{\varphi}_p^*$ with $\mathbf{U}^* = \mathbf{U}_+ \mathbf{V}_+^{1/2}$. By assuming independent normal priors for $\boldsymbol{\varphi}_p^*$, a proper prior for the penalized terms that is proportional to equation (2.4) can be obtained.

We take two additional measures to enhance the computational efficiency of the resulting MCMC algorithm. First, only eigenvalues/vectors that explain a majority of the variability in equation (2.4) are used to construct \mathbf{U}^* . Additionally, we apply a parameter-expansion technique for the penalized terms in $f_p(\cdot)$, setting $\boldsymbol{\varphi}_p^* = \beta_p^* \boldsymbol{\xi}_p$, where β_p^* is a scalar and $\boldsymbol{\xi}_p$ is a vector with the same dimension as $\boldsymbol{\varphi}_p^*$. This technique enables us to perform selection on the penalized terms as a group rather than determining their inclusion separately. By rescaling β_p^* and $\boldsymbol{\xi}_p$ at each MCMC iteration such that $|\boldsymbol{\xi}_p|$ has mean equal to one, $\boldsymbol{\xi}_p$ maintains the shape of the smooth function, and β_p^* represents the term's strength of association while preserving identifiability, similar to Scheipl, Fahrmeir and Kneib (2012).

For variable selection we impose spike-and-slab prior distributions on the $3 * P = T$ -dimensional vector $\boldsymbol{\beta} = (\beta_1^*, \beta_1^\circ, \beta_{01}, \dots, \beta_P^*, \beta_P^\circ, \beta_{0P})'$. In general, the spike-and-slab prior distribution is composed of a mixture of a Dirac delta function at zero, $\delta_0(\cdot)$, and a known distribution, $\mathcal{S}(\cdot)$, such as a normal with mean zero and diffuse variance (George and McCulloch (1993), Brown, Vannucci and Fearn (1998)). A latent indicator variable, v_t , representing a risk factor's inclusion or exclusion in the model, determines whether the risk factor's regression coefficient is set to zero (spike) or free to be estimated in the model (slab). Specifically for a given coefficient β_t , we assume

$$(2.5) \quad \beta_t | v_t \sim v_t \cdot \mathcal{S}(\beta_t) + (1 - v_t) \delta_0(\beta_t).$$

To complete the prior specification for this portion of the model, we assume that the slab component, $\mathcal{S}(\beta_t)$, follows a $N(0, \tau^2)$ with variance τ^2 and that the inclusion indicators are distributed as $v_t|\theta_t \sim \text{Bernoulli}(\theta_t)$ with prior probability of inclusion $\theta_t \sim \text{Beta}(a_{v_t}, b_{v_t})$. Integrating out θ_t , we obtain $v_t \sim \text{Beta-Binomial}(a_{v_t}, b_{v_t})$, where hyperparameters a_{v_t} and b_{v_t} are set to control the sparsity in the model. Lastly, each element of ξ_p , ξ_{pr} , is assumed to follow a $N(\mu_{pr}, 1)$ with mean $\mu_{pr} = \pm 1$ with equal probability. Placing a majority of the prior mass for each ξ_{pr} around ± 1 is motivated by the role it plays in the expansion of φ_p^* , as described above.

2.3. Prior specification for the random effects. We perform selection on the random effects, α_i , using the modified Cholesky decomposition approach of [Chen and Dunson \(2003\)](#). Specifically, we reparameterize the random effects

$$(2.6) \quad \alpha_i = \mathbf{K}\mathbf{\Gamma}\xi_i,$$

where \mathbf{K} a positive diagonal matrix with elements $\kappa = (\kappa_1, \dots, \kappa_D)'$ and $\mathbf{\Gamma}$ a lower triangle matrix with diagonal elements set to one and free elements otherwise. To perform variable selection, we set the prior for κ to follow a similar spike-and-slab prior distribution, as in [Section 2.2](#), where the slab distribution $\mathcal{S}(\kappa_d) = \text{FN}(m_0, v_0)$. Here, FN represents a folded normal distribution defined as

$$\begin{aligned} \text{FN}(m_0, v_0) &= (2\pi v_0)^{-1/2} \exp(-(\kappa_d - m_0)^2 / (2v_0)) \\ &\quad + (2\pi v_0)^{-1/2} \exp(-(\kappa_d + m_0)^2 / (2v_0)), \end{aligned}$$

where $m_0 \in \mathbb{R}$ and $v_0 > 0$ are location and scale parameters, respectively. Note that we forgo the parameter-expansion approach of [Kinney and Dunson \(2007\)](#), which introduces a redundant multiplicative parameter in the implied random effect covariance matrix, in favor of a model that enables meaningful inference for κ and, ultimately, their cluster assignments. Similar to [Section 2.2](#), we let the corresponding inclusion indicators λ_d follow a Beta-Binomial($a_{\lambda_d}, b_{\lambda_d}$) to induce sparsity on the random effect terms. Lastly, we assume the $D(D-1)/2$ -dimensional vector of free elements in $\mathbf{\Gamma}$ follow $N(\boldsymbol{\gamma}_0, V_\gamma) \cdot I(\boldsymbol{\gamma} \in \mathcal{Z})$, where I represents an indicator function and \mathcal{Z} represents the parameters with corresponding random effects included in the model. For example, if the d^{th} random effect is included (i.e., $\lambda_d = 1$), then $\gamma_{d1}, \dots, \gamma_{d,d-1}$ and $\gamma_{d+1,d}, \dots, \gamma_{D,d} \in \mathcal{Z}$. Lastly, we assume $\xi_i \sim N(\mathbf{0}, \mathbf{I})$.

2.4. Spiked nonparametric priors. To complete our approach, we investigate nonparametric prior constructions for the spike-and-slab components of the reparameterized fixed and random effects by assuming that the slab component follows a Dirichlet process (DP). These priors are commonly referred to as spiked DP (SDP) priors ([Canale et al. \(2017\)](#), [Kim, Dahl and Vannucci \(2009\)](#), [Savitsky and Vannucci \(2010\)](#), [Dunson, Herring and Engel \(2008\)](#)). In the context of our model, SDP priors allow us to simultaneously select influential risk factors while clustering effects with similar relations to the smoking outcome. The formulation we use here is sometimes refers to as an ‘‘outer’’ SDP prior, since the point mass at zero is outside of the base distribution of the DP. Alternatively, the ‘‘inner’’ construction places the spike-and-slab prior inside the DP, serving as the base distribution. The inner formulation provides the opportunity for coefficients to cluster at zero but does not force a point mass at zero explicitly. As such, the likelihood that a coefficient is assigned to the trivial cluster grows with the number of coefficients excluded from the model. Alternatively, the outer formulation is a more informative prior, since it explicitly assigns a point mass at zero, and, in addition, carries less computational demands since it does not require auxiliary variables for MCMC sampling ([Neal \(2000\)](#), [Savitsky and Vannucci \(2010\)](#)). We refer readers to [Canale](#)

et al. (2017) for a detailed explanation of the structural differences between the two prior formulations.

First, we assume the regression coefficients associated with the main effects and linear interaction terms follow a SDP to provide insights on risk factors that share underlying linear trends with momentary smoking by the next assessment over the course of the study. Specifically, we assume the slab component in equation (2.5) is a Dirichlet process prior $H \sim \text{DP}(\vartheta, H_0)$ with base distribution $H_0 = N(0, \tau^2)$ and concentration parameter ϑ . Furthermore, we assume a hyperprior, $\vartheta \sim G(a_\vartheta, b_\vartheta)$, with $a_\vartheta, b_\vartheta > 0$. For the nonlinear interaction terms, we avoid the SDP since it would produce uninterpretable cluster assignments due to the parameter-expansion approach taken to improve selection performance. For example, similar values for β_i^* and $\beta_{i'}^*$ may correspond to vastly different φ_i^* and $\varphi_{i'}^*$, depending on their respective ξ and spline basis functions. Similarly, placing a DP prior on the individual components in ξ , or even φ , would not provide interpretable results on the overall nonlinear effect. We take a similar approach for the random effects. Here, we assume the slab components for the diagonal elements of \mathbf{K} , $\mathcal{S}(\kappa_d) = W$, $W \sim \text{DP}(\mathcal{A}, W_0)$, where $W_0 \sim \text{FN}(m_0, v_0)$ and \mathcal{A} is the concentration parameter of the DP. To complete the prior assumptions for the random effects portion of the model, let $\mathcal{A} \sim G(a_{\mathcal{A}}, b_{\mathcal{A}})$, where $a_{\mathcal{A}}, b_{\mathcal{A}} > 0$ are shape and rate parameters, respectively.

There is evidence that relaxing parametric assumptions for random effects using DP priors may cause inferential challenges, as the mean of the random effects are nonzero almost surely (Li, Müller and Lin (2011), Yang (2012), Cai and Bandyopadhyay (2017)). Our approach differs in that we do not directly replace the typical normal assumption for random effects with a nonparametric prior. Instead, we place a nonparametric prior on the covariance decomposition components, \mathbf{K} , while letting ζ_i follow a normal distribution centered at zero. As such, our approach avoids any identifiability issues with the fixed effects while still relaxing the parametric assumption on the reparameterized random effects, $\mathbf{K}\mathbf{\Gamma}\zeta_i$. It is important to note that by doing this we are adopting a Bayesian semiparametric modeling structure, since the random effects are linear combinations of spiked Dirichlet process and normal random variables (Müller, Quintana and Rosner (2007)).

2.5. *Posterior inference.* For posterior inference we implement a Metropolis–Hastings within Gibbs algorithm. The full joint model is defined as

$$f(\mathbf{y}|\boldsymbol{\varrho}, \boldsymbol{\omega}, \mathbf{x}, \mathbf{u}, \mathbf{z})p(\boldsymbol{\omega})p(\boldsymbol{\beta}|\mathbf{v})p(\mathbf{v})p(\vartheta)p(\mathbf{K}|\boldsymbol{\lambda})p(\boldsymbol{\lambda})p(\mathcal{A})p(\xi|\boldsymbol{\mu})p(\boldsymbol{\mu})p(\boldsymbol{\zeta})p(\boldsymbol{\Gamma}),$$

where $\boldsymbol{\varrho} = \{\boldsymbol{\beta}, \xi, \mathbf{K}, \boldsymbol{\Gamma}, \boldsymbol{\zeta}\}$. We use the Pólya-Gamma augmentation of Polson, Scott and Windle (2013) to efficiently sample the posterior distribution for the logistic regression model. Following Polson, Scott and Windle (2013), we express the likelihood contribution of $y_{i,j+1}$ as

$$f(y_{i,j+1}|\cdot) = \frac{(e^{\psi_{ij}})^{y_{i,j+1}}}{(1 + e^{\psi_{ij}})} \propto \exp(k_{i,j+1}\psi_{ij}) \int_0^\infty \exp(-\omega_{i,j+1}\psi_{ij}^2/2)p(\omega_{i,j+1}|n_{i,j+1}, 0)\partial\omega,$$

where $k_{i,j+1} = y_{i,j+1} - n_{i,j+1}/2$, $p(\omega_{i,j+1}|n_{i,j+1}, 0) \sim \text{PG}(n_{i,j+1}, 0)$ and PG is the Pólya-Gamma distribution. Using the notation presented in the previous sections, we set

$$\psi_{ij} = \sum_{p=1}^P (\beta_p^* \mathbf{u}_{ij}^* \xi_p + \beta_p^\circ u_{ij} + \beta_{0p})x_{ijp} + \mathbf{z}'_{ij} \mathbf{K}\boldsymbol{\Gamma}\zeta_i.$$

The MCMC sampler used to implement our model is outlined below in Algorithm 1. A more detailed description of the MCMC steps as well as a graphical representation of the model are provided in the Supplementary Material (Koslovsky et al. (2020)). After burn-in and thinning, the remaining samples obtained from running Algorithm 1 for \tilde{T} iterations

Algorithm 1 MCMC Sampler

```

1: Input data  $\mathbf{y}, \mathbf{x}, \mathbf{u}, \mathbf{z}$ 
2: Initialize parameters:  $\boldsymbol{\varrho}, \boldsymbol{\omega}, \mathbf{v}, \boldsymbol{\lambda}, \vartheta, \mathcal{A}, \boldsymbol{\mu}$ 
3: Set  $\text{DP}_{\bar{\beta}}$  and  $\text{DP}_K$  to True or False to indicate DP for slab on fixed or random effects,
   respectively.
4: for iteration  $\tilde{t} = 1, \dots, \tilde{T}$  do
5:   for  $i = 1, \dots, N$  do
6:     for  $j = 1, \dots, n_i - 1$  do
7:       Update  $\omega_{i,j+1} \sim \text{PG}(1, \psi_{ij})$ 
8:     end for
9:   end for
10:  if  $\text{DP}_{\bar{\beta}}$  then
11:    Update cluster assignment of  $\bar{\beta}$  following Neal (2000) algorithm 2.
12:  end if
13:  Jointly update  $\boldsymbol{\beta}$  and  $\mathbf{v}$  with Between and Within Step following Savitsky, Vannucci
   and Sha (2011).
14:  Update  $\boldsymbol{\xi}$  from FCD  $N(\mu_{\boldsymbol{\xi}}, V_{\boldsymbol{\xi}})$ .
15:  for  $p = 1, \dots, P$  do
16:    Rescale  $\boldsymbol{\xi}_p^*$  and  $\beta_p^*$  so  $\boldsymbol{\varphi}_p^*$  remains unchanged.
17:  end for
18:  for  $p = 1, \dots, P$  do
19:    for  $r = 1, \dots, r_p$  do
20:      Set  $\mu_{pr} = 1$  with probability  $1/(1 + \exp(-2\xi_{pr}))$ .
21:    end for
22:  end for
23:  Update  $\vartheta$  by the two-step Gibbs update of Escobar and West (1995).
24:  if  $\text{DP}_K$  then
25:    Update cluster assignment of  $\text{DP}_K$  following Neal (2000) algorithm 2.
26:  end if
27:  Jointly update  $\mathbf{K}$  and  $\boldsymbol{\lambda}$  with Between and Within Step following Savitsky, Vannucci
   and Sha (2011).
28:  Update  $\mathcal{A}$  following two-step Gibbs update of Escobar and West (1995).
29:  Update  $\boldsymbol{\Gamma}$  from FCD  $N(\hat{\boldsymbol{\gamma}}, \hat{V}_{\boldsymbol{\gamma}}) \cdot I(\boldsymbol{\gamma} \in \mathcal{Z})$ .
30:  for  $i = 1, \dots, N$  do
31:    Update  $\boldsymbol{\zeta}_i$  from FCD  $N(\hat{\boldsymbol{\zeta}}_i, \hat{V}_{\boldsymbol{\zeta}_i})$ .
32:  end for
33: end for

```

are used for inference. To determine a risk factor's inclusion in the model, its marginal posterior probability of inclusion (MPPI) is empirically estimated by calculating the average of its respective inclusion indicator's MCMC samples (George and McCulloch (1997)). Note that inclusion for both fixed and random effects is determined marginally for β_t and λ_d , respectively. Commonly, covariates are included in the model if their MPPI exceeds 0.50 (Barbieri and Berger (2004)) or a Bayesian false discovery rate threshold which controls for multiplicity (Newton et al. (2004)).

3. Case study. In this section we study the smoking behaviors in a group of adult smokers recruited from a smoking cessation research clinic. The overall research goal of this study was to identify and investigate the structural form of the relations between a set of risk factors

and smoking over a five-week period surrounding a scheduled quit attempt, using intensive longitudinal data collected with EMAs.

3.1. Data analysis. In the study design, momentary smoking, our outcome of interest, was defined as whether or not a subject reported smoking in the four hours prior to the current EMA. However, at each EMA a subject was prompted on their *current* psychological, social, environmental and behavioral status. Thus, to maintain temporality in this study, we assessed the relations between momentary smoking and measurements collected in the previous EMA. As such, regression coefficients are interpreted as the log odds of momentary smoking by the next assessment for a particular risk factor. In this study we investigated psychological and affective factors including *urge* to smoke, feelings of *restlessness*, *negative affect* (i.e., irritability, frustration/anger, sadness, worry, misery), *positive affect* (i.e., happiness and calmness), being *bored*, *anxiousness* and *motivation to quit smoking*. Additionally, we investigated numerous social and environmental factors, such as whether or not the subject was *interacting with a smoker*, if cigarettes were easily available (*cigarette availability*) and whether or not the subject was drinking alcohol (*alcohol consumption*). Also, we included a set of baseline, time-invariant measures (i.e., heaviness of smoking index (*HSI*), *age* (years), being *female* and treatment assignment) into the model. For each of these risk factors, we included a fixed main effect, linear interaction and nonlinear interaction term as well as a random main effect and linear interaction term. All interactions investigated in this analysis were between risk factors and assessment time (i.e., $u_{ij} = t_{ij}$), and t_{ij} were centered so that $t = 0$ represents the beginning of the scheduled quit attempt.

Only complete EMAs with corresponding timestamps were included in this analysis, resulting in 9634 total observations with the median number of assessments per individual 151 (IQR 101.5-162). All continuous covariates were standardized to mean zero and variance one before analysis to help reduce multicollinearity and place covariates on the same scale for interpretation. The spline functions were initially generated with 20 basis functions, but only the eigenvalues/eigenvectors that captured 99.9% of the variability were included in the model to reduce the parameter space and computation time, similar to Scheipl, Fahrmeir and Kneib (2012). This reduced the column space of the penalized covariates \mathcal{U}^* to 8 in our application. We applied our model with the traditional spike-and-slab prior as well as the spiked DP. When fitting each model, we chose a noninformative prior for the fixed and random effects' inclusion indicators, $a_{v_l} = b_{v_l} = a_{\lambda_d} = b_{\lambda_d} = 1$. This assumption reflects the exploratory nature of our study aimed at learning potential relations between risk factors and smoking behaviors with little or no information regarding their occurrence in the presence of other risk factors. We assumed a mildly informative prior on the fixed regression coefficients by setting $\tau^2 = 2$. This places a 95% prior probability of included regression coefficients between an odds ratio of 0.06 and 16. Additionally, we set $v_0 = v^* = 10$, $m_0 = m^* = 0$, and $\Gamma \sim N(\boldsymbol{\gamma}_0 = \mathbf{0}, \mathbf{V}_\gamma = \mathbf{I})$. Lastly, when using the SDP prior, the hyperparameters for the concentration parameters ϑ and \mathcal{A} were set to $a_\vartheta = b_\vartheta = a_{\mathcal{A}} = b_{\mathcal{A}} = 1$. For posterior inference we ran our MCMC algorithm with and without SDP priors for both fixed and random effects for 10,000 iterations, treating the first 5000 as burn-in and thinning to every 10th iteration. Trace plots of the parameters' posterior samples indicated good convergence and mixing. Additionally, we observed a relatively high correlation ($\sim 97\%$) between the posterior probabilities of inclusion obtained from two chains initiated with different parameter values and potential scale reduction factors, \hat{R} , for each of the selected $\boldsymbol{\beta}$ and \mathbf{K} below 1.1 (Gelman and Rubin (1992)), further demonstrating that the MCMC procedure was working properly and the chains converged. To assess model fit, a residual plot and a series of posterior predictive checks were performed in which we compared replicated data sets from the posterior predictive distribution of the model to the observed data (Gelman et al.

(2000)). Overall, we found strong evidence of good model fit; see the Supplementary Materials for details (Koslovsky et al. (2020)). Inclusion in the model was determined using the median model approach (Barbieri and Berger (2004)) (i.e., marginal posterior probability of inclusion (MPPI) ≥ 0.50). For the SDP model, clusters of regression coefficients were determined using sequentially-allocated latent structure optimization to minimize the lower bound of the variation of information loss (Wade and Ghahramani (2018), Dahl and Muller (2017)). To compare the predictive performance of both models, we performed a leave-one-out cross-validation approximation procedure, following the approach proposed by Vehtari, Gelman and Gabry (2017). This approach approximates leave-one-out (LOO) cross-validation with the expected log pointwise predictive density (epld). By using Pareto smoothed importance sampling (PSIS) for estimation, it provides a more stable estimate compared to the method of Gelfand (1996). We used the R package `loo` (Vehtari, Gelman and Gabry (2016)), which requires the pointwise log-likelihood for each subject $i = 1, \dots, N$ at each observation $j = 1, \dots, n_i$ calculated at each MCMC iteration $s = 1, \dots, S$ and produces an estimated $\widehat{\text{epld}}$ value with larger values implying a superior model.

3.2. *Results.* Overall, we found better predictive performance for the model with SDP priors vs. the traditional spike-and-slab priors, $\widehat{\text{epld}}_{\text{SDP}} = -2985.1$ and $\widehat{\text{epld}}_{\text{SS}} = -3062.7$, respectively. Plots of the marginal posterior probabilities of inclusion for the fixed and random effects selected using our proposed approach with SDP priors are found in Figure 1.

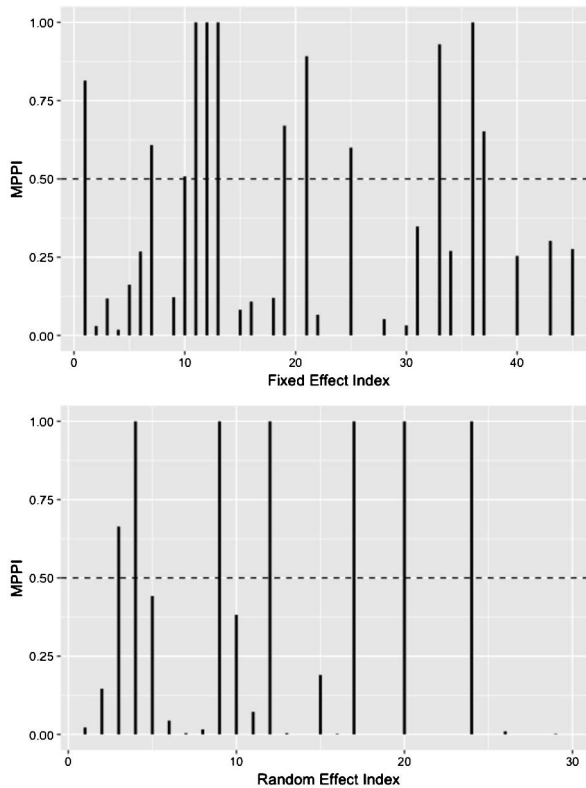


FIG. 1. *Smoking cessation study: Marginal posterior probabilities of inclusion (MPPI) for fixed (top) and random (bottom) effects. Selected fixed effects in ascending order: NCI (NL-INTX), urge to quit (NL-INTX), cigarette availability (all), interacting with a smoker (NL-INTX), negative affect (NL-INTX, main), being bored (NL-INTX), alcohol consumption (main), motivation to quit (main), HSI (NL-INTX). Selected random effects in ascending order: Urge (main), cigarette availability (main), being bored (main), motivation to quit (main), SmartT (L-INTX), interacting with a smoker (L-INTX), being bored (L-INTX). Dotted lines represent the inclusion threshold of 0.50. NL-INTX: Nonlinear interaction, L-INTX: Linear interaction.*

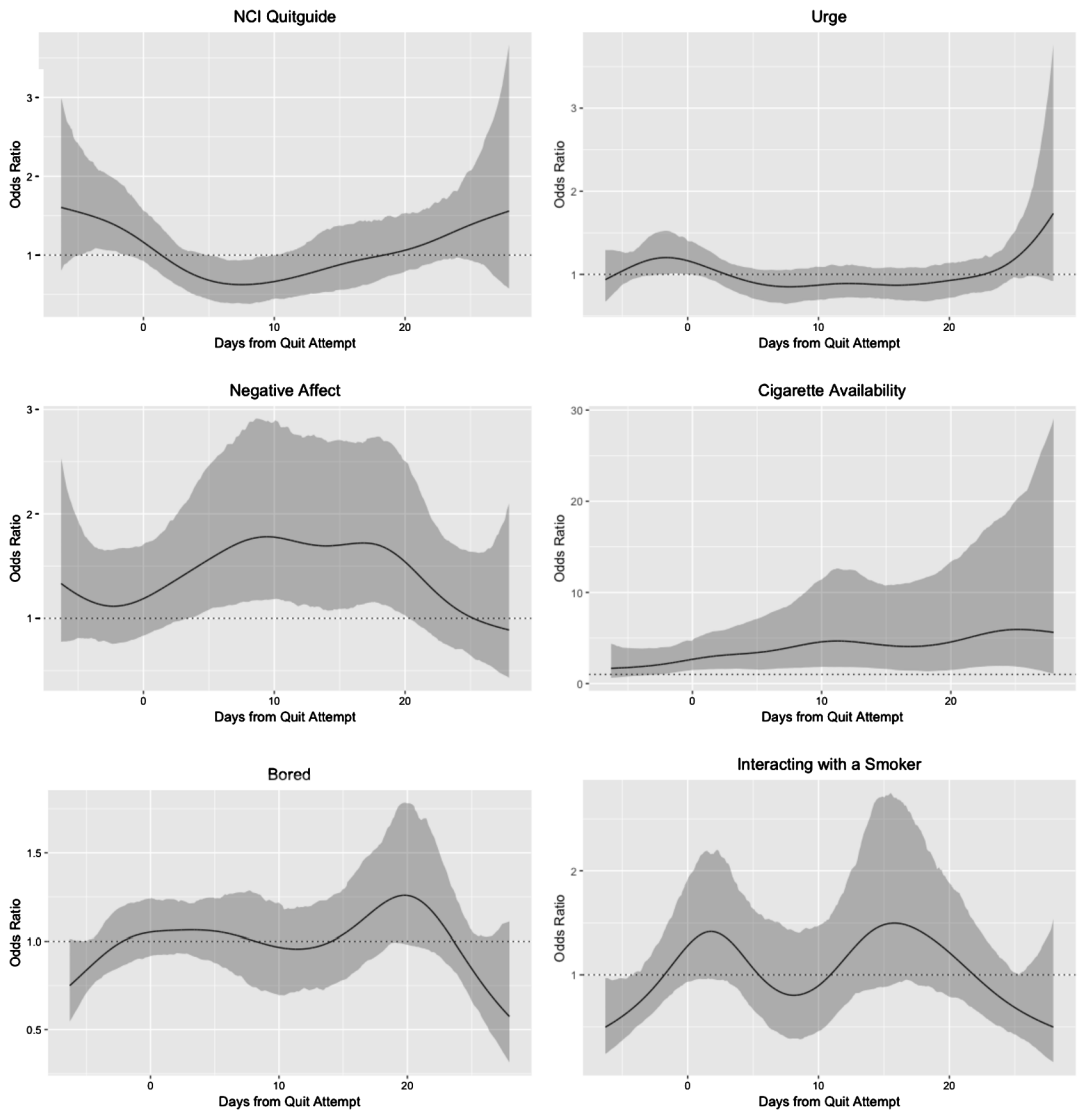


FIG. 2. *Smoking Cessation Study: Time-varying effects on momentary smoking by the next assessment of those covariates selected by our model with SDP priors. Shaded regions represent pointwise 95% CI. Dashed lines indicate an odds ratio of one.*

Figure 2 presents the time-varying effects selected using the same model. Compared to *usual care*, we found a higher odds of momentary smoking by the next assessment for those assigned to the *NCI QuitGuide* group prior to the quit attempt. However, immediately after the quit attempt we observed a lower odds of momentary smoking by the next assessment for those assigned to the *NCI QuitGuide* group which gradually increased to the initial level over the remainder of the study (top left panel). Similarly, we observed a positive relation between having the *urge* to smoke and momentary smoking by the next assessment prior to the quit attempt that diminished during the three weeks following the quit attempt before sharply increasing during the fourth week post-quit (top right panel). Throughout the assessment period we observed a positive relation between *negative affect* and momentary smoking by the next assessment that increased during the first week post-quit, leveling off at an odds ratio of 1.75 until the third week after the quit attempt. We additionally found a positive relation between *cigarette availability* and the odds of momentary smoking by the next assessment that strengthened over the assessment window. For a 1 SD increase in *cigarette availability*,

the odds of momentary smoking by the next assessment increased by 300% for the typical subject one week after the quit attempt, holding all else constant. In the two lower panels of Figure 2, we observe a relatively weak, oscillating effect of being *bored* and *interacting with a smoker* on momentary smoking by the next assessment, respectively. In addition to these effects, the model identified a constant effect for *alcohol consumption* in the last hour and *motivation to quit smoking* over the assessment period. A similar set of fixed effect relations were identified by our model without the SDP prior, with the exception of not selecting being *bored*.

Compared to standard TVEMs, our approach deconstructs the structure of the relations between risk factors and smoking behaviors over time, aiding the interpretation of the underlying trends. This information may help the development and evaluation of tailored intervention strategies targeting smoking cessation using mHealth data. For example, *negative affect* has an obvious positive association with momentary smoking by the next assessment that wavers around an odds ratio of 1.2 to 1.5 for a majority of the study. However, based on Figure 2, it is unclear whether or not the effect linearly diminishes over time. By performing selection on the main effect, linear interaction, and nonlinear interaction terms separately, we are able to obtain an actual point estimate for the constant effect of *negative affect* (OR 1.40), as opposed to subjectively assuming a range of values from the plot. Additionally, since the linear interaction term was not selected, we can claim that the effect was not linearly decreasing over time and that it was simply wavering around the constant effect throughout the study.

Tables 1 and 2 present the estimated variances and corresponding 95% credible intervals (CI) for the random effects selected using SDP priors and traditional spike-and-slab priors, respectively. Using SDP priors, our method identified a random main effect for *urge to smoke*, *cigarette availability*, being *bored* and *motivation to quit smoking* as well as a random linear interaction between being assigned to the *SmartT* treatment group, *interacting with smokers* and being *bored* with time. Thus, even though we did not discover an overall difference in the odds of momentary smoking by the next assessment for those assigned to the *SmartT* treatment vs. *usual care*, we observed evidence that the subjects responded differently to the *SmartT* treatment across the assessment window. With the traditional spike-and-slab priors, we found similar results overall. However, the model only selected a random main effect for *interacting with smokers* and, additionally, suggested a random effect for *anxiousness*.

By using SDP priors, our approach is capable of clustering covariates that share similar linear trends with momentary smoking by the next assessment over time. In practice, this information can be used to help construct decision rules when designing future intervention strategies. In our analysis, only five main effect and linear interaction terms were selected,

TABLE 1
Smoking cessation study: Estimated variances with corresponding 95% credible intervals (CI) for selected random effects with SDP priors based on $MPPI \geq 0.50$

Random effect	$\hat{\sigma}^2$	95% CI
Intercept	0.923	(0.539, 1.528)
Urge	0.152	(0.031, 0.278)
Cigarette Availability	0.865	(0.394, 1.467)
Bored	0.183	(0.076, 0.398)
Motivation to Quit Smoking	0.156	(0.045, 0.311)
SmartT \times Time	0.077	(0.010, 0.210)
Interacting with a Smoker \times Time	0.016	(0.002, 0.050)
Bored \times Time	0.002	(0.000, 0.005)

TABLE 2
Smoking cessation study: Estimated variances with corresponding 95% credible intervals (CI) for selected random effects with traditional spike-and-slab priors based on $\text{MPPI} \geq 0.50$

Random effect	$\hat{\sigma}^2$	95% CI
Intercept	1.317	(0.676, 2.487)
Urge	0.099	(0.011, 0.248)
Cigarette Availability	0.905	(0.503, 1.607)
Interacting with a Smoker	0.848	(0.286, 1.924)
Bored	0.244	(0.065, 0.517)
Anxiousness	0.140	(0.001, 0.361)
Motivation to Quit Smoking	0.212	(0.076, 0.448)
SmartT \times Time	0.062	(0.016, 0.155)
Bored \times Time	0.002	(0.000, 0.004)

and each of them were allocated to their own cluster. With this knowledge, researchers can prioritize targeting risk factors based on their relative strength of association at a given moment. Had some of these risk factors' effects been clustered together, researchers may rely more heavily on other pieces of information, such as the cost or success rates for a particular intervention strategy, when assessing which risk factors to target during a high-risk moment.

Similar to previous studies investigating the temporal relation between risk factors and smoking behaviors around a quit attempt, our results show a convex relation between *urge* to smoke and momentary smoking after the quit attempt, a positive association with *cigarette availability* throughout the quit attempt and a positive, increasing relation between *negative affect* and momentary smoking during the first week after the quit attempt (Koslovsky et al. (2018), Vasilenko et al. (2014)). Existing TVEMs approaches, however, typically model the repeated measures structure of the data by simply including a random intercept term in the model, neglecting to investigate random main effects or interaction terms. They also do not incorporate variable selection. Our approach, on the other hand, delivers insights on how relations vary over time as well as how they vary across individuals.

3.3. Sensitivity analysis. To investigate our model's sensitivity to prior specification, we set each of the hyperparameters to default values and then evaluated the effect of manipulating each term on the results obtained in Section 3. For the default parameterization we set the hyperparameters for the prior inclusion indicators \mathbf{v} and λ to $a_{v_i} = b_{v_i} = a_{\lambda_d} = b_{\lambda_d} = 1$. For interpretation, $a_{v_i} = b_{v_i} = 1$ implies that the prior probability of inclusion for a fixed effect is $a_{v_i} / (a_{v_i} + b_{v_i}) = 0.50$. The default values for the variance of the normal distribution for the slab of β_0 and β° as well as the base distribution for β^* were each fixed at 5. Additionally, the mean and variance for the random effect terms' proposal and prior distributions were set to 0 and 5, respectively. The hyperparameters for the concentration parameters ϑ and \mathcal{A} were set to $a_\vartheta = b_\vartheta = a_{\mathcal{A}} = b_{\mathcal{A}} = 1$. Lastly, we assumed $\Gamma \sim N(\boldsymbol{\gamma}_0 = \mathbf{0}, V_\gamma = \mathbf{I})$. We ran our MCMC algorithm for 10,000 iterations, treating the first 5000 iterations as burn-in and thinning to every 10th iteration for the SDP model, similar to our case study. For each of the fixed and random effects, inclusion in the model was determined using the median model approach (Barbieri and Berger (2004)).

Since the true model is never known in practice, we evaluated each model parameterization in terms of sparsity levels and overlap with the results reported in the case study section. Specifically, we present the total number terms selected for both fixed and random effects (# Fixed and # Random). We also provide the proportion of active risk factors in our case study that were also included by each model and the proportion of inactive risk factors that

TABLE 3

Case study data: Sensitivity results for the proposed model with SDP across various prior specifications. Total number of terms selected for both fixed and random effects are indicated as # Fixed and # Random, respectively. The proportion of active (inactive) risk factors presented in the case study that were also included (excluded) by each model is reported as f-IN and r-IN (f-EX and r-EX), for fixed and random effects, respectively. Finally, the overall proportion of active (inactive) risk factors presented in the case study that were also included (excluded) by each model is represented as IN (EX)

	$a_{v_i} = a_{\lambda_d} = 1, b_{v_i} = b_{\lambda_d} = 9$	$\tau^2 = v_0 = 2$	$a_{\vartheta} = b_{\vartheta} = a_{\mathcal{A}} = b_{\mathcal{A}} = 0.1$
# Fixed	4	8	6
# Random	5	5	7
IN	0.60	0.70	0.80
f-IN	0.44	0.67	0.56
r-IN	0.50	0.50	0.83
EX	0.80	0.60	1.00
f-EX	1.00	1.00	1.00
r-EX	0.78	0.78	0.89
	$a_{v_i} = a_{\lambda_d} = 9, b_{v_i} = b_{\lambda_d} = 1$	$\tau^2 = v_0 = 10$	$a_{\vartheta} = b_{\vartheta} = a_{\mathcal{A}} = b_{\mathcal{A}} = 10$
# Fixed	10	7	6
# Random	8	5	4
IN	1.00	0.60	0.80
f-IN	0.78	0.67	0.56
r-IN	0.83	0.50	0.33
EX	0.60	1.00	1.00
f-EX	0.83	0.80	1.00
r-EX	0.67	0.78	0.78

were also excluded by each model, for fixed (f-IN and f-EX) and random effects (r-IN and r-EX) as well as overall (IN and EX). Results of the sensitivity analysis are reported in Table 3. Compared to the results presented in the case study, we found relatively consistent overlap in the risk factors included and excluded by each model overall. We observed moderate sensitivity to hyperparameter values in terms of percent overlap for fixed and random effects of risk factors included in the model, an artifact of the relatively weak associations identified for some of the risk factors. Notably, risk factors showing stronger associations with momentary smoking at the next assessment (e.g., *negative affect*, *cigarette availability* and *motivation to quit smoking*) were selected by the model regardless of prior specification. Likewise, weaker relations between momentary smoking at the next assessment and risk factors, such as *being bored* and *interacting with a smoker*, were more sensitive to hyperparameters. We also observed that the number of selected fixed and random effects increased (decreased) as the prior probability of inclusion increased (decreased), as expected. In practice, there are a variety of factors researchers should consider when setting the prior probability of inclusion, including the aim of the research study, the desired sparsity of the model and prior knowledge of covariates inclusion as well as results from simulation and sensitivity analyses to name a few. From a clinical perspective, $\tau^2 = 10$ reflects a relatively diffuse prior for a given risk factor (i.e., odds ratio between 0.002 and roughly 500). To further investigate the model's sensitivity to regression coefficients' variances, we set $\tau^2 = v_0 = 1000$, and found somewhat similar results to the model with $\tau^2 = v_0 = 10$ overall (i.e., IN = 0.8, EX = 0.8). Here, we unexpectedly found nonmonotonic behavior in the proportion of included and excluded terms as a function of the coefficients' variance, which might also reflect our model's sensitivity to relatively weak associations, as previously noted. In theory, the selection of random effects may

be sensitive to the order in which the columns of \mathbf{Z} are ordered, since the Cholesky decomposition is, itself, order dependent (Müller, Scealy and Welsh (2013)). In our case study we did not observe any differences regarding which random effects were selected with a random permutation of the \mathbf{Z} columns. In Section 5 we further demonstrate our model's robustness to the ordering of \mathbf{Z} on simulated data.

4. Simulation study. In this section we evaluate our model in terms of variable selection and clustering performance on simulated data similar in structure to our case study data. We compared our method with and without SDP priors on varying-coefficient and random effects to two other Bayesian methods which are designed to handle this class of models. The first is the method of Scheipl, Fahrmeir and Kneib (2012) which has previously shown promising results performing function selection in structural additive regression models using continuous spike-and-slab priors. Their approach differs from ours in that they assume parameter-expanded normal-mixture-of-inverse-gamma (peNMIG) distribution priors for selection, inspired by Ishwaran and Rao (2005), and design a Metropolis–Hastings with penalized iteratively weighted least-squares algorithm for updating regression coefficients within the logistic framework. A popular alternative to spike-and-slab priors to induce sparsity in high-dimensional regression settings is to assume global-local shrinkage priors on the regression coefficients (see van Erp, Oberski and Mulder (2019), Bhadra et al. (2019) for detailed reviews). At the request of a reviewer, we additionally compared our proposed model to a reparameterized version with shrinkage priors (Carvalho, Polson and Scott (2009)). To achieve this, we replaced the spike-and-slab priors on $\boldsymbol{\beta}$ with horseshoe priors which belong to the class of global-local scale mixtures of normal priors (Polson and Scott (2010)). For random effects, \mathbf{K} , we assumed a similar global-local structure for the scale parameters of the folded-normal distribution, v_0 . To our knowledge, the theoretical properties and selection performance of global-local scale mixtures of nonnormal priors have yet to be explored. However, we conjectured that the global-local framework should effectively shrink inactive random effects toward zero and allow active terms to be freely estimated. Details of the resulting model and accompanying MCMC algorithm are found in the Supplementary Material (Koslovsky et al. (2020)).

We simulated $N = 100$ subjects with 20–40 observations randomly spaced across an assessment window with $t_{ij} \in [0, 1]$, without loss of generality. For each observation, we generated a set of 15 covariates, \mathbf{x}_i , comprised of an intercept term and 14 continuous covariates simulated from a $N_{14}(\mathbf{0}, \Sigma)$, where $\Sigma_{st} = w^{|s-t|}$ and $w = 0.3$. To simulate time-varying covariate trajectories, we randomly jittered half of the elements within \mathbf{x}_i by $N(0, 1)$. Additionally, we set $\mathbf{z}_{ij} = \mathbf{x}_{ij}$. Thus, each full model contained 15 main effects, linear interactions, nonlinear interactions and random main effects, corresponding to 60 potential terms (or groups of terms for the nonlinear interaction components) to select. The first five functional terms in the true model were defined as:

- $f_1(t_{ij}) = \pi \sin(3\pi t_{ij}) + 1.4t_{ij} - 1.6$
- $f_2(t_{ij}) = \pi \cos(2\pi t_{ij}) + 1.6$
- $f_3(t_{ij}) = -\pi t \sin(5\pi t_{ij}) + 1.7t_{ij} - 1.5$
- $f_4(t_{ij}) = -1.5t_{ij} + 1.6$
- $f_5(t_{ij}) = -1.6,$

and the random effects $\mathbf{a}_i \sim N(\mathbf{0}, \Sigma_\alpha)$ with $\sigma_{kk} = 0.75$ and $\sigma_{jk} = 0.4$ for $j, k = 1, \dots, 5$. Thus, in the true model, $\psi_{ij} = \sum_{p=1}^5 f_p(t_{ij})x_{ijp} + \mathbf{z}'_{ij}\mathbf{a}_i$. Note that to impose an inherent clustering for the main effects and linear interaction terms, their values were specified to center around ± 1.5 .

We ran each of the MCMC algorithms on 50 replicated data sets, using 7500 iterations, treating the first 3750 iterations as burn-in and thinning to every 10th iteration for each model.

The spline functions were generated similar to our application. We set the hyperparameters for the inclusion indicators, $a_{v_i} = b_{v_i} = a_{v_i} = b_{v_i} = 1$, imposing a noninformative prior for selection of fixed and random effect terms. Additionally, we fixed the regression coefficient hyperparameters to $\tau^2 = 2$ and $m_0 = 0$ with $v_0 = 10$. For the concentration parameters ϑ and \mathcal{A} , we assumed $a_\vartheta = b_\vartheta = a_{\mathcal{A}} = b_{\mathcal{A}} = 1$. Before analysis, the covariates were standardized to mean 0 and variance 1.

For each of the models with spike-and-slab priors, inclusion in the model for both fixed and random effects was determined using the median model approach (Barbieri and Berger (2004)). For the horseshoe model, fixed effects were considered active if their corresponding 95% credible interval did not contain zero, similar to Bhadra et al. (2019). The 95% credible interval for random effects will almost surely not contain zero. As a naive alternative we assumed a random effect was active in the model if its posterior mean exceeded a given threshold. For the sake of demonstration, we evaluated the performance of the model over a grid of potential threshold values and presented the results for the best performing model overall. Notably, this solution is only feasible when the true answer is known, which is never the case in practice. Variable selection performance was evaluated via sensitivity (SENS), specificity (SPEC) and Matthew's correlation coefficient (MCC) for fixed and random effects separately. These metrics are defined as

$$\begin{aligned} \text{SENS} &= \frac{\text{TP}}{\text{FN} + \text{TP}}, \\ \text{SPEC} &= \frac{\text{TN}}{\text{FP} + \text{TN}}, \\ \text{MCC} &= \frac{\text{TP} \times \text{TN} - \text{FP} \times \text{FN}}{\sqrt{(\text{TP} + \text{FP})(\text{TP} + \text{FN})(\text{TN} + \text{FP})(\text{TN} + \text{FN})}}, \end{aligned}$$

where TN, TP, FN and FP represent the true negatives, true positives, false negatives and false positives, respectively. For the SDP models, clusters of regression coefficients were determined using sequentially-allocated latent structure optimization to minimize the lower bound of the variation of information loss (Wade and Ghahramani (2018), Dahl and Muller (2017)). Once clusters were determined, clustering performance was evaluated using the variation of information, a measure of distance between two clusterings ranging from 0 to $\log R$, where R is the number of items to cluster and lower values imply better clustering (Meilă (2003)).

Figure 3 presents the estimated smooth functions obtained using our proposed method with SDP priors on a randomly selected replicated data set from the simulation study. Here, $f_1(t_{ij})$ represents the global intercept comprised of a main effect, linear interaction and non-linear interaction term that were forced into the model. Of interest is the ability of the model to properly select the influential components in $f_2(t_{ij})$ and $f_3(t_{ij})$ and, additionally, capture their structure. Using the method proposed in Dahl and Muller (2017) to identify latent clusters of fixed main effect and linear interaction terms, our method successfully clustered the linear interaction in $f_1(t_{ij})$ and the main effects in $f_2(t_{ij})$ and $f_4(t_{ij})$ while incorrectly assigning the linear interaction term in $f_3(t_{ij})$ to its own cluster. Additionally, the main effects in $f_1(t_{ij})$, $f_3(t_{ij})$ and $f_5(t_{ij})$ were appropriately clustered together, while the linear interaction term in $f_4(t_{ij})$ was incorrectly assigning to its own cluster. The remaining, uninfluential terms were all allocated to the trivial group. Despite $f_1(t_{ij})$ and $f_3(t_{ij})$ having similar main effect and linear interaction terms, they are dramatically different in terms of their nonlinear interaction terms. However, by clustering their underlying linear trajectories our model with SDP priors was able to uncover similarities in their relations with the outcome over time that traditional approaches would fail to discover.

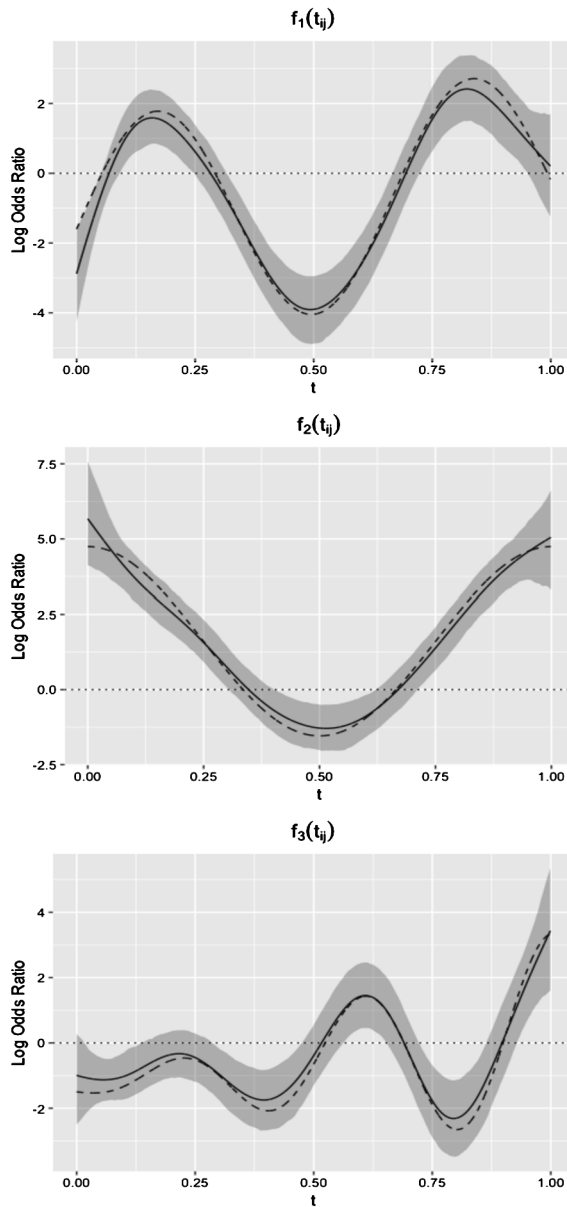


FIG. 3. *Simulated Data: Estimated smooth function $f_1(t_{ij})$, $f_2(t_{ij})$, $f_3(t_{ij})$ for a randomly selected replicate data set generated in the simulation study. The estimated smooth function is represented by a solid black line with pointwise 95% credible regions in gray. Dashed lines represent the true log odds ratios as a function of time.*

Table 4 reports results for our proposed method with SDP priors (PGBVSDP), our proposed method without SDP priors (PGBVS), peNMIG and our model with horseshoe priors (PGHS) in terms of average sensitivity, specificity and MCC for fixed (fSENS, fSPEC, fMCC) and random (rSENS, rSPEC, rMCC) effects across the replicate data sets with standard errors in parentheses. Additionally, for the PGBVSDP model we provide clustering performance results for fixed (fCLUST) and random effects (rCLUST). Since each of the random effects were simulated similarly, clusterings were compared to a single cluster for the nonzero terms. Overall, the methods had relatively similar results for fixed effects, with PGBVS and PGHS performing the best in terms of sensitivity (1.00 and 1.00) and MCC (0.96 and 0.99), respectively. Our method with SDP priors, PGBVSDP, obtained the highest specificity for fixed effects overall. Given that the maximum possible values fCLUST and

TABLE 4

Simulated Data: Results for the proposed model with and without the SDP on regression coefficients compared to peNMIG (Scheipl, Fahrmeir and Kneib (2012)) and our model with horseshoe priors (Carvalho, Polson and Scott (2009)). Results are averaged over 50 replicate data sets with standard deviations in parentheses

	PGBVSDP	PGBVS	peNMIG	PGHS
fSENS	0.96 (0.09)	1.00 (0.02)	0.93 (0.11)	1.00 (0.00)
fSPEC	0.99 (0.02)	0.98 (0.02)	0.94 (0.04)	0.96 (0.01)
fMCC	0.94 (0.08)	0.96 (0.05)	0.83 (0.10)	0.99 (0.02)
fCLUST	0.39 (0.21)	–	–	–
rSENS	0.76 (0.21)	0.62 (0.25)	0.46 (0.23)	0.86 (0.24)
rSPEC	0.88 (0.10)	0.96 (0.05)	0.64 (0.16)	0.90 (0.11)
rMCC	0.63 (0.26)	0.64 (0.23)	0.11 (0.33)	0.76 (0.21)
rCLUST	0.92 (0.50)	–	–	–
Time (s)	4658 (271)	2235 (46)	10,720 (1116)	3076 (74)

rCLUST could take on were 3.4 and 2.7, respectively, we found fairly strong clustering performance for both fixed (0.39) and random (0.92) effects with PGBVSDP. We observed more variability in the selection of random effects across models. Random effect selection sensitivity was significantly lower compared to the fixed effects for all of the models. In terms of specificity (1-false positive rate) for random effects, our methods, regardless of prior formulation, dramatically outperformed peNMIG, with PGBVS obtaining the highest specificity overall (0.96). However, PGBVSDP and PGBVS had lower sensitivity with respect to random effects, compared to PGHS. While PGHS performed well separating active from inactive random effects, recall that the truth was used to select the optimal selection threshold. The improved performance of PGBVS, PGBVSDP and PGHS in terms of variable selection was achieved in considerably less computation time compared to peNMIG. Our core method was able to run 7500 iterations in a fifth of the time compared to peNMIG, accessed via Scheipl (2011). Using the SDP priors, which requires additional updates for clustering the regression coefficients, we observed a two-fold increase in computation time for PGBVSDP compared to PGBVS. However, on average, the PGBVSDP approach still achieved about a 50% reduction in computation time compared to peNMIG. It is important to note that, for comparison, all algorithms were run in series, even though the R package spikeSlabGAM (Scheipl (2011)) provides functionality to run multiple chains in parallel.

5. Sensitivity analysis. To assess the model's sensitivity to hyperparameter settings, we set each of the hyperparameters to default values and then evaluated the effect of manipulating each term on selection and clustering performance. For the default parameterization, we set the hyperparameters for the prior inclusion indicators ν and λ to $a_{\nu_i} = b_{\nu_i} = a_{\lambda_d} = b_{\lambda_d} = 1$. The default values for the variance of the normal distribution for the slab of β_0 and β° as well as the base distribution for β^* were each fixed at 5. Additionally, the mean and variance for the random effect terms' proposal and prior distributions were set to 0 and 5, respectively. The hyperparameters for the concentration parameters, ϑ and \mathcal{A} $a_\vartheta = b_\vartheta = a_{\mathcal{A}} = b_{\mathcal{A}} = 1$. Lastly, we assumed $\Gamma \sim N(\gamma_0 = \mathbf{0}, V_\gamma = \mathbf{I})$. We ran our MCMC algorithm on the 50 replicated data sets generated in the simulation study, using 7500 iterations, treating the first 3750 iterations as burn-in and thinning to every 10th iteration for the SDP model.

Results of the sensitivity analysis are reported in Table 5. As expected, we found that the sensitivity (specificity) increased (decreased) as the prior probability of inclusion for the fixed and random effects increased. The model did not seem sensitive to the variance assumed for the normal and folded normal priors assigned to the fixed and random effect slab distributions, respectively. Similarly, we found comparable results in terms of sensitivity and specificity for

TABLE 5

Simulated data: Sensitivity results for the proposed model with SDP on regression coefficients. Results are averaged over 50 replicated data sets with standard errors in parentheses

	$a_{v_t} = a_{\lambda_d} = 1, b_{v_t} = b_{\lambda_d} = 9$	$\tau^2 = v_0 = 2$	$a_{\vartheta} = b_{\vartheta} = a_{\mathcal{A}} = b_{\mathcal{A}} = 0.1$
fSENS	0.92 (0.13)	0.97 (0.08)	0.94 (0.12)
fSPEC	0.99 (0.02)	0.99 (0.02)	0.99 (0.02)
fMCC	0.93 (0.10)	0.96 (0.07)	0.94 (0.09)
fCLUST	0.45 (0.30)	0.35 (0.20)	0.45 (0.25)
rSENS	0.50 (0.20)	0.79 (0.20)	0.54 (0.28)
rSPEC	0.87 (0.09)	0.88 (0.08)	0.85 (0.10)
rMCC	0.40 (0.26)	0.66 (0.23)	0.41 (0.29)
rCLUST	1.30 (0.36)	0.91 (0.44)	1.25 (0.50)
	$a_{v_t} = a_{\lambda_d} = 9, b_{v_t} = b_{\lambda_d} = 1$	$\tau^2 = v_0 = 10$	$a_{\vartheta} = b_{\vartheta} = a_{\mathcal{A}} = b_{\mathcal{A}} = 10$
fSENS	0.99 (0.03)	0.96 (0.07)	0.94 (0.11)
fSPEC	0.96 (0.03)	0.99 (0.02)	0.99 (0.02)
fMCC	0.91 (0.06)	0.95 (0.07)	0.93 (0.11)
fCLUST	0.40 (0.20)	0.39 (0.23)	0.41 (0.24)
rSENS	0.85 (0.20)	0.78 (0.20)	0.74 (0.23)
rSPEC	0.84 (0.10)	0.89 (0.10)	0.86 (0.10)
rMCC	0.66 (0.19)	0.67 (0.25)	0.60 (0.27)
rCLUST	0.84 (0.49)	0.89 (0.47)	0.97 (0.49)

different values of the concentration parameters' hyperparameters. In terms of clustering, we saw marginally better variation of information measures with larger concentration parameter hyperparameters. However, across simulations runs we observed relatively high standard errors in terms of the variation of information measures. To assess potential sensitivity to the order of random effects in our simulations, we reran the simulation study with a random permutation of the columns of \mathbf{Z} . Similar to the case study, we found no evidence of sensitivity to random effect ordering with our model as the results were almost identical to those presented in Table 4 with PGBVSDP (rSENS = 0.76 (0.20), rSPEC = 0.87 (0.09), rMCC = 0.62 (0.20), rCLUST = 0.94 (0.42)).

6. Conclusions. Using a semiparametric Bayesian time-varying effect modeling framework, in this paper we have investigated intensive longitudinal data, collected in a novel, smartphone-based smoking cessation study to better understand the relation between potential risk factors and smoking behaviors in the critical moments surrounding a quit attempt. Unlike standard TVEMs, our approach deconstructs the structure of the relations between risk factors and smoking behaviors over time which aids in formulating hypotheses regarding dynamic relations between risk factors and smoking in the critical moments around a quit attempt. By performing variable selection on random effects, the approach delivers additional insights on how relations vary over time as well as how they vary across individuals. Furthermore, the use of non- and semiparametric prior constructions allows simultaneous variable selection for fixed and random effects while learning latent clusters of regression coefficients. As such, our model is designed to discover various forms of latent structures within the data without requiring strict model assumptions or burdensome tuning procedures. Results from our analysis have confirmed previously identified temporal relations between smoking behaviors and *urge* to smoke, *cigarette availability* and *negative affect*. They have also identified subject-specific heterogeneity in the effects of *urge* to smoke, *cigarette availability* and *motivation to quit*. Additionally, we have found that subjects differed in how they

responded to the *SmartT* treatment (compared to usual care), *interacting with a smoker* and being *bored* over time. This has practical relevance as researchers can use this information to design adaptive interventions that prioritize targeting risk factors based on their relative strength of association at a given moment. They also reinforce the importance of designing dynamic intervention strategies that are adaptive to subjects' current risk profiles.

Throughout this work we have demonstrated how our method is well suited to aide the development and evaluation of future JITAI strategies targeting smoking cessation using mHealth data. The existing *SmartT* algorithm delivers treatment based on the presence of six lapse triggers which are weighted based on their relative importance in predicting risk of lapse (Businelle et al. (2016)). The results of this study allow for a more dynamic algorithm that takes into account not only the time-varying relationships between psychosocial and environmental variables and smoking lapse but also the different ways in which individuals experience a quit attempt. For example, the results suggest that providing momentary support to cope with *urge* to smoke and *negative affect* may be more useful if delivered in the early stages of a quit attempt but become less important by week 4 post-quit. However, messages that address *cigarette availability*, *alcohol consumption* and *motivation to quit smoking* may be a more important focus for the entire quit attempt. Although the findings for this small sample may not be generalizable to larger, more diverse populations, these methods are the next step in developing a personalized smoking risk algorithm that can inform highly specific, individualized treatment to each smoker.

It is important to note that selection of a risk factor by our proposed method (or any variable selection technique), does not imply clinical significance. Notably, the pointwise credible intervals often contained odds ratios of one, and most risk factors were only influential for brief moments throughout the study period. While these results highlight the importance of understanding risk factors' dynamic relations with smoking to design tailored intervention strategies, we recommend using our method for hypothesis generation in practice and conducting confirmatory studies before generalizing results.

Compliance rates for EMA studies typically range between 70% and 90%, with a recommended threshold of 80% (Jones, Xu and Grunwald (2006)). In our case study the compliance rate was 84%. Additionally, 97.3% of all assessments were completed once initiated, and subjects were unable to skip questions within an assessment. Since subjects were assessed multiple times per day, nonresponse was attributed more to situational context (e.g., driving) than smoking status. Thus, for this study we found the missing completely at random assumption for missing observations justified. However, future studies may consider the development of advanced analytical methods for EMA data sets that can handle different types of missingness assumptions and other potential biases, such as social desirability bias.

In this analysis we focus on time-varying effects due to their recent popularity in smoking behavior research—Tan et al. (2012), Vasilenko et al. (2014), Koslovsky et al. (2018), Lanza et al. (2014), Shiyko et al. (2012, 2014). A promising alternative for investigating the complexity of smoking behaviors around a quit attempt is the varying index coefficient model which allows a covariate's effect to vary as a function of multiple other variables (Ma and Song (2015)). By incorporating variable selection priors, researchers could identify which variables are responsible for modifying a covariate's effect. Oftentimes, behavioral researchers are interested in exploring other forms of latent structure, such as clusters of individuals who respond similarly to treatments or have similar risk profiles over time. Taking advantage of the flexibility and efficiency of our approach, future work could extend our core model to address these research questions by recasting it into a mixture modeling framework. In addition, while we have developed our method for binary outcomes due to their prevalence in smoking behavior research studies, our approach is easily adaptable to other data structures found within and outside of smoking behavior research, such as time to event data

(Sha, Tadesse and Vannucci (2006)) and continuous outcomes. While our method borrows information across regression coefficients, we avoided imposing structure among covariates via heredity constraints which restrict the model space for higher order terms depending on the inclusion status of the lower order terms that comprise them. Researchers interested in extending our approach to accommodate these and other forms of hierarchical constraints may adjust the prior probabilities of inclusion (Chipman (1996)). Lastly, while we were hesitant to present variable selection results for PGHS, due to the limited understanding of global-local priors for non-Gaussian distributions, this showed good results in simulations. Furthermore, when applied to the case study data, we obtained promising predictive performance (i.e., $\widehat{\text{epId}}_{\text{HS}} = -2955.5$) that warrant future investigation of its theoretical properties.

Acknowledgments. Matthew Koslovsky is supported by NSF Research Training Group award DMS-1547433.

SUPPLEMENTARY MATERIAL

Supplemental code (DOI: [10.1214/20-AOAS1402SUPPA](https://doi.org/10.1214/20-AOAS1402SUPPA); .zip). R-package PGBVS contains code to perform the methods described in the article. The package also contains functionality for reproducing the data used in the sensitivity and simulation studies and for posterior inference. The R package is located at <https://github.com/mkoslovsky/PGBVS>.

Supplemental simulations and results (DOI: [10.1214/20-AOAS1402SUPPB](https://doi.org/10.1214/20-AOAS1402SUPPB); .pdf). This file contains a description of the full joint distribution of our model with a graphical representation, a detailed description of our proposed MCMC algorithm with and without SDP priors, and derivations for the prior marginal likelihood used to sample latent cluster assignments. Additionally, we include details of the goodness-of-fit analysis for the case study.

REFERENCES

- BARBIERI, M. M. and BERGER, J. O. (2004). Optimal predictive model selection. *Ann. Statist.* **32** 870–897. MR2065192 <https://doi.org/10.1214/009053604000000238>
- BERARDI, V., CARRETERO-GONZÁLEZ, R., BELLETTIERE, J., ADAMS, M. A., HUGHES, S. and HOVELL, M. (2018). A Markov approach for increasing precision in the assessment of data-intensive behavioral interventions. *J. Biomed. Inform.* **85** 93–105.
- BHADRA, A., DATTA, J., POLSON, N. G. and WILLARD, B. (2019). Lasso meets horseshoe: A survey. *Statist. Sci.* **34** 405–427. MR4017521 <https://doi.org/10.1214/19-STS700>
- BOLMAN, C., VERBOON, P., THEWISSEN, V., BOONEN, V., SOONS, K. and JACOBS, N. (2018). Predicting smoking lapses in the first week of quitting: An ecological momentary assessment study. *J. Addict. Med.* **12** 65–71. <https://doi.org/10.1097/ADM.0000000000000365>
- BROOK, D. W., BROOK, J. S., ZHANG, C., WHITEMAN, M., COHEN, P. and FINCH, S. J. (2008). Developmental trajectories of cigarette smoking from adolescence to the early thirties: Personality and behavioral risk factors. *Nicotine Tob. Res.* **10** 1283–1291.
- BROWN, P. J., VANNUCCI, M. and FEARN, T. (1998). Multivariate Bayesian variable selection and prediction. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **60** 627–641. MR1626005 <https://doi.org/10.1111/1467-9868.00144>
- BUSINELLE, M. S., MA, P., KENDZOR, D. E., REITZEL, L. R., CHEN, M., LAM, C. Y., BERNSTEIN, I. and WETTER, D. W. (2014). Predicting quit attempts among homeless smokers seeking cessation treatment: An ecological momentary assessment study. *Nicotine Tob. Res.* **16** 1371–1378.
- BUSINELLE, M. S., MA, P., KENDZOR, D. E., FRANK, S. G., VIDRINE, D. J. and WETTER, D. W. (2016). An ecological momentary intervention for smoking cessation: Evaluation of feasibility and effectiveness. *J. Med. Internet Res.* **18** e321.
- CAI, B. and BANDYOPADHYAY, D. (2017). Bayesian semiparametric variable selection with applications to periodontal data. *Stat. Med.* **36** 2251–2264. MR3660129 <https://doi.org/10.1002/sim.725>
- CANALE, A., LIJOI, A., NIPOTI, B. and PRÜNSTER, I. (2017). On the Pitman–Yor process with spike and slab base measure. *Biometrika* **104** 681–697. MR3694590 <https://doi.org/10.1093/biomet/asx041>
- CARVALHO, C. M., POLSON, N. G. and SCOTT, J. G. (2009). Handling sparsity via the horseshoe. In *Artificial Intelligence and Statistics* 73–80.

- CHEN, Z. and DUNSON, D. B. (2003). Random effects selection in linear mixed models. *Biometrics* **59** 762–769. MR2025100 <https://doi.org/10.1111/j.0006-341X.2003.00089.x>
- CHENG, J., EDWARDS, L. J., MALDONADO-MOLINA, M. M., KOMRO, K. A. and MULLER, K. E. (2010). Real longitudinal data analysis for real people: Building a good enough mixed model. *Stat. Med.* **29** 504–520. MR2751785 <https://doi.org/10.1002/sim.3775>
- CHIPMAN, H. (1996). Bayesian variable selection with related predictors. *Canad. J. Statist.* **24** 17–36. MR1394738 <https://doi.org/10.2307/3315687>
- CURSIO, J. F., MERMELSTEIN, R. J. and HEDEKER, D. (2019). Latent trait shared-parameter mixed models for missing ecological momentary assessment data. *Stat. Med.* **38** 660–673. MR3902605 <https://doi.org/10.1002/sim.7989>
- DAHL, D. B. and MULLER, P. (2017). sdols: Summarizing distributions of latent structures. R Package Version 1.4. 591.
- DE HAAN-RIETDIJK, S., KUPPENS, P., BERGEMAN, C. S., SHEEBER, L. B., ALLEN, N. B. and HAMAKER, E. L. (2017). On the use of mixed Markov models for intensive longitudinal data. *Multivar. Behav. Res.* **52** 747–767. <https://doi.org/10.1080/00273171.2017.1370364>
- DUNSON, D. B., HERRING, A. H. and ENGEL, S. M. (2008). Bayesian selection and clustering of polymorphisms in functionally related genes. *J. Amer. Statist. Assoc.* **103** 534–546. MR2523991 <https://doi.org/10.1198/016214507000000554>
- 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. *Psychol. Methods* **20** 444.
- ESCOBAR, M. D. and WEST, M. (1995). Bayesian density estimation and inference using mixtures. *J. Amer. Statist. Assoc.* **90** 577–588. MR1340510
- FITZMAURICE, G. M., LAIRD, N. M. and WARE, J. H. (2012). *Applied Longitudinal Analysis*, 2nd ed. Wiley Series in Probability and Statistics. Wiley, Hoboken, NJ. MR2830137
- GEISER, C., BISHOP, J., LOCKHART, G., SHIFFMAN, S. and GRENARD, J. L. (2013). Analyzing latent state-trait and multiple-indicator latent growth curve models as multilevel structural equation models. *Front. Psychol.* **4** 975. <https://doi.org/10.3389/fpsyg.2013.00975>
- GELFAND, A. E. (1996). Model determination using sampling-based methods. In *Markov Chain Monte Carlo in Practice. Interdiscip. Statist.* 145–161. CRC Press, London. MR1397969
- GELMAN, A. and RUBIN, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statist. Sci.* **7** 457–472.
- GELMAN, A., GOEGEBEUR, Y., TUERLINCKX, F. and VAN MECHELEN, I. (2000). Diagnostic checks for discrete data regression models using posterior predictive simulations. *J. R. Stat. Soc. Ser. C. Appl. Stat.* **49** 247–268. MR1821324 <https://doi.org/10.1111/1467-9876.00190>
- GEORGE, E. I. and MCCULLOCH, R. E. (1993). Variable selection via Gibbs sampling. *J. Amer. Statist. Assoc.* **88** 881–889.
- GEORGE, E. I. and MCCULLOCH, R. E. (1997). Approaches for Bayesian variable selection. *Statist. Sinica* 339–373.
- HASTIE, T. and TIBSHIRANI, R. (1993). Varying-coefficient models. *J. Roy. Statist. Soc. Ser. B* **55** 757–796. MR1229881
- HERON, K. E. and SMYTH, J. M. (2010). Ecological momentary interventions: Incorporating mobile technology into psychosocial and health behaviour treatments. *Br. J. Health Psychol.* **15** 1–39. <https://doi.org/10.1348/135910709X466063>
- HUI, F. K. C., MÜLLER, S. and WELSH, A. H. (2017). Hierarchical selection of fixed and random effects in generalized linear mixed models. *Statist. Sinica* **27** 501–518. MR3674683
- ISHWARAN, H. and RAO, J. S. (2005). Spike and slab variable selection: Frequentist and Bayesian strategies. *Ann. Statist.* **33** 730–773. MR2163158 <https://doi.org/10.1214/009053604000001147>
- JONES, R. H., XU, S. and GRUNWALD, G. K. (2006). Continuous time Markov models for binary longitudinal data. *Biom. J.* **48** 411–419. MR2240089 <https://doi.org/10.1002/bimj.200510224>
- KIM, S., DAHL, D. B. and VANNUCCI, M. (2009). Spiked Dirichlet process prior for Bayesian multiple hypothesis testing in random effects models. *Bayesian Anal.* **4** 707–732. MR2570085 <https://doi.org/10.1214/09-BA426>
- KINNEY, S. K. and DUNSON, D. B. (2007). Fixed and random effects selection in linear and logistic models. *Biometrics* **63** 690–698. MR2395705 <https://doi.org/10.1111/j.1541-0420.2007.00771.x>
- KLASNJA, P., HEKLER, E. B., SHIFFMAN, S., BORUVKA, A., ALMIRALL, D., TEWARI, A. and MURPHY, S. A. (2015). Microrandomized trials: An experimental design for developing just-in-time adaptive interventions. *Health Psychology* **34** 1220.
- KOSLOVSKY, M. D., HÉBERT, E. T., SWARTZ, M. D., CHAN, W., LEON-NOVELO, L., WILKINSON, A. V., KENDZOR, D. E. and BUSINELLE, M. S. (2018). The time-varying relations between risk factors and smoking before and after a quit attempt. *Nicotine Tob. Res.* **20** 1231–1236.

- KOSLOVSKY, M. D., SWARTZ, M. D., CHAN, W., LEON-NOVELO, L., WILKINSON, A. V., KENDZOR, D. E. and BUSINELLE, M. S. (2018). Bayesian variable selection for multistate Markov models with interval-censored data in an ecological momentary assessment study of smoking cessation. *Biometrics* **74** 636–644. MR3825350 <https://doi.org/10.1111/biom.12792>
- KOSLOVSKY, M. D., HÉBERT, E. T., BUSINELLE, M. S. and VANNUCCI, M. (2020). Supplement to “A Bayesian time-varying effect model for behavioral mHealth data.” <https://doi.org/10.1214/20-AOAS1402SUPPA>, <https://doi.org/10.1214/20-AOAS1402SUPPB>
- KÜRÜM, E., LI, R., SHIFFMAN, S. and YAO, W. (2016). Time-varying coefficient models for joint modeling binary and continuous outcomes in longitudinal data. *Statist. Sinica* **26** 979–1000. MR3559939
- LANG, S. and BREZGER, A. (2004). Bayesian P-splines. *J. Comput. Graph. Statist.* **13** 183–212. MR2044877 <https://doi.org/10.1198/1061860043010>
- LANZA, S. T., VASILENKO, S. A., LIU, X., LI, R. and PIPER, M. E. (2014). Advancing the understanding of craving during smoking cessation attempts: A demonstration of the time-varying effect model. *Nicotine Tob. Res.* **16** S127–S134.
- LI, Y., MÜLLER, P. and LIN, X. (2011). Center-adjusted inference for a nonparametric Bayesian random effect distribution. *Statist. Sinica* **21** 1201–1223. MR2827521 <https://doi.org/10.5705/ss.2009.180>
- LUCKETT, D. J., LABER, E. B., KAHKOSKA, A. R., MAAHS, D. M., MAYER-DAVIS, E. and KOSOROK, M. R. (2020). Estimating dynamic treatment regimes in mobile health using V-learning. *J. Amer. Statist. Assoc.* **115** 692–706. MR4107673 <https://doi.org/10.1080/01621459.2018.1537919>
- MA, S. and SONG, P. X.-K. (2015). Varying index coefficient models. *J. Amer. Statist. Assoc.* **110** 341–356. MR3338507 <https://doi.org/10.1080/01621459.2014.903185>
- MASON, M., MENNIS, J., WAY, T., LANZA, S., RUSSELL, M. and ZAHARAKIS, N. (2015). Time-varying effects of a text-based smoking cessation intervention for urban adolescents. *Drug Alcohol Depend.* **157** 99–105. <https://doi.org/10.1016/j.drugalcdep.2015.10.016>
- MCCARTHY, D. E., EBSSA, L., WITKIEWITZ, K. and SHIFFMAN, S. (2016). Repeated measures latent class analysis of daily smoking in three smoking cessation studies. *Drug Alcohol Depend.* **165** 132–142.
- MCCLURE, J. B., ANDERSON, M. L., BRADLEY, K., AN, L. C. and CATZ, S. L. (2016). Evaluating an adaptive and interactive mHealth smoking cessation and medication adherence program: A randomized pilot feasibility study. *JMIR mHealth uHealth* **4** e94.
- MEILÄ, M. (2003). Comparing clusterings by the variation of information. In *Learning Theory and Kernel Machines* 173–187. Springer, Berlin.
- MINAMI, H., YEH, V. M., BOLD, K. W., CHAPMAN, G. B. and MCCARTHY, D. E. (2014). Relations among affect, abstinence motivation and confidence, and daily smoking lapse risk. *Psychol. Addict. Behav.* **28** 376–388. <https://doi.org/10.1037/a0034445>
- MÜLLER, P., QUINTANA, F. A. and ROSNER, G. L. (2007). Semiparametric Bayesian inference for multilevel repeated measurement data. *Biometrics* **63** 280–289, 316. MR2345598 <https://doi.org/10.1111/j.1541-0420.2006.00668.x>
- MÜLLER, S., SCEALY, J. L. and WELSH, A. H. (2013). Model selection in linear mixed models. *Statist. Sci.* **28** 135–167. MR3112403 <https://doi.org/10.1214/12-sts410>
- NAHUM-SHANI, I., SMITH, S. N., SPRING, B. J., COLLINS, L. M., WITKIEWITZ, K., TEWARI, A. and MURPHY, S. A. (2017). Just-in-time adaptive interventions (JITAs) in mobile health: Key components and design principles for ongoing health behavior support. *Ann. Behav. Med.* **52** 446–462.
- NAUGHTON, F., HOPEWELL, S., LATHIA, N., SCHALBROECK, R., BROWN, C., MASCOLO, C., MCEWEN, A. and SUTTON, S. (2016). A context-sensing mobile phone app (Q sense) for smoking cessation: A mixed-methods study. *JMIR mHealth uHealth* **4** e106.
- NEAL, R. M. (2000). Markov chain sampling methods for Dirichlet process mixture models. *J. Comput. Graph. Statist.* **9** 249–265. MR1823804 <https://doi.org/10.2307/1390653>
- NEWTON, M. A., NOUEIRY, A., SARKAR, D. and AHLQUIST, P. (2004). Detecting differential gene expression with a semiparametric hierarchical mixture method. *Biostatistics* **5** 155–176.
- PIASECKI, T. M., FIORE, M. C., MCCARTHY, D. E. and BAKER, T. B. (2002). Have we lost our way? The need for dynamic formulations of smoking relapse proneness. *Addiction* **97** 1093–1108.
- PIASECKI, T. M., JORENBY, D. E., SMITH, S. S., FIORE, M. C. and BAKER, T. B. (2003). Smoking withdrawal dynamics: II. Improved tests of withdrawal-relapse relations. *J. Abnorm. Psychol.* **112** 14–27.
- PIASECKI, T. M., TRELA, C. J., HEDEKER, D. and MERMELSTEIN, R. J. (2013). Smoking antecedents: Separating between- and within-person effects of tobacco dependence in a multiwave ecological momentary assessment investigation of adolescent smoking. *Nicotine Tob. Res.* **16** S119–S126.
- POLSON, N. G. and SCOTT, J. G. (2010). Shrink globally, act locally: Sparse Bayesian regularization and prediction. *Bayesian Stat.* **9** 105.

- POLSON, N. G., SCOTT, J. G. and WINDLE, J. (2013). Bayesian inference for logistic models using Pólya-Gamma latent variables. *J. Amer. Statist. Assoc.* **108** 1339–1349. MR3174712 <https://doi.org/10.1080/01621459.2013.829001>
- RILEY, W. T., RIVERA, D. E., ATIENZA, A. A., NILSEN, W., ALLISON, S. M. and MERMELSTEIN, R. (2011). Health behavior models in the age of mobile interventions: Are our theories up to the task? *Transl. Behav. Med.* **1** 53–71. <https://doi.org/10.1007/s13142-011-0021-7>
- RIVERA, D. E., PEW, M. D. and COLLINS, L. M. (2007). Using engineering control principles to inform the design of adaptive interventions: A conceptual introduction. *Drug Alcohol Depend.* **88** S31–S40.
- SAVITSKY, T. and VANNUCCI, M. (2010). Spiked Dirichlet process priors for Gaussian process models. *J. Probab. Stat.* **2010** Art. ID 201489, 14. MR2745498 <https://doi.org/10.1155/2010/201489>
- SAVITSKY, T., VANNUCCI, M. and SHA, N. (2011). Variable selection for nonparametric Gaussian process priors: Models and computational strategies. *Statist. Sci.* **26** 130–149. MR2849913 <https://doi.org/10.1214/11-STS354>
- SCHIEPL, F. (2011). spikeSlabGAM: Bayesian variable selection, model choice and regularization for generalized additive mixed models in R. Preprint. Available at [arXiv:1105.5253](https://arxiv.org/abs/1105.5253).
- SCHIEPL, F., FAHRMEIR, L. and KNEIB, T. (2012). Spike-and-slab priors for function selection in structured additive regression models. *J. Amer. Statist. Assoc.* **107** 1518–1532. MR3036413 <https://doi.org/10.1080/01621459.2012.737742>
- SELYA, A. S., UPDEGROVE, N., ROSE, J. S., DIERKER, L., TAN, X., HEDEKER, D., LI, R. and MERMELSTEIN, R. J. (2015). Nicotine-dependence-varying effects of smoking events on momentary mood changes among adolescents. *Addict. Behav.* **41** 65–71.
- SHA, N., TADESSE, M. G. and VANNUCCI, M. (2006). Bayesian variable selection for the analysis of microarray data with censored outcomes. *Bioinformatics* **22** 2262–2268.
- SHIFFMAN, S. (2013). Conceptualizing analyses of ecological momentary assessment data. *Nicotine Tob. Res.* **16** S76–S87.
- 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. *J. Consult. Clin. Psychol.* **64** 366–379. <https://doi.org/10.1037/0022-006x.64.2.366>
- SHIFFMAN, S., BALABANIS, M. H., PATY, J. A., ENGBERG, J., GWALTNEY, C. J., LIU, K. S., GNYS, M., HICKCOX, M. and PATON, S. M. (2000). Dynamic effects of self-efficacy on smoking lapse and relapse. *Health Psychology* **19** 315.
- 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. *J. Abnorm. Psychol.* **111** 531.
- SHIYKO, M. P., LANZA, S. T., TAN, X., LI, R. and SHIFFMAN, S. (2012). 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. *Prev. Sci.* **13** 288–299.
- SHIYKO, M., NAAB, P., SHIFFMAN, S. and LI, R. (2014). Modeling complexity of EMA data: Time-varying lagged effects of negative affect on smoking urges for subgroups of nicotine addiction. *Nicotine Tob. Res.* **16** S144–S150.
- TAN, X., SHIYKO, M. P., LI, R., LI, Y. and DIERKER, L. (2012). A time-varying effect model for intensive longitudinal data. *Psychol. Methods* **17** 61.
- TIMMS, K. P., RIVERA, D. E., COLLINS, L. M. and PIPER, M. E. (2013). A dynamical systems approach to understanding self-regulation in smoking cessation behavior change. *Nicotine Tob. Res.* **16** S159–S168.
- 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. *Psychol. Methods* **19** 175.
- VAN ERP, S., OBERSKI, D. L. and MULDER, J. (2019). Shrinkage priors for Bayesian penalized regression. *J. Math. Psych.* **89** 31–50. MR3903921 <https://doi.org/10.1016/j.jmp.2018.12.004>
- VASILENKO, S. A., PIPER, M. E., LANZA, S. T., LIU, X., YANG, J. and LI, R. (2014). Time-varying processes involved in smoking lapse in a randomized trial of smoking cessation therapies. *Nicotine Tob. Res.* **16** S135–S143.
- VEHTARI, A., GELMAN, A. and GABRY, J. (2016). loo: Efficient leave-one-out cross-validation and WAIC for Bayesian models. R Package Version 0.1 **6**.
- VEHTARI, A., GELMAN, A. and GABRY, J. (2017). Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Stat. Comput.* **27** 1413–1432. MR3647105 <https://doi.org/10.1007/s11222-016-9696-4>
- WADE, S. and GHAHRAMANI, Z. (2018). Bayesian cluster analysis: Point estimation and credible balls (with discussion). *Bayesian Anal.* **13** 559–626. MR3807860 <https://doi.org/10.1214/17-BA1073>

- WALLS, T. A. and SCHAFER, J. L., eds. (2006). *Models for Intensive Longitudinal Data* Oxford Univ. Press, Oxford. MR2334216 <https://doi.org/10.1093/acprof:oso/9780195173444.001.0001>
- YANG, M. (2012). Bayesian variable selection for logistic mixed model with nonparametric random effects. *Comput. Statist. Data Anal.* **56** 2663–2674. MR2915153 <https://doi.org/10.1016/j.csda.2011.12.014>