Received: 17 May 2022 | Accepted: 26 April 2023

DOI: 10.1111/biom.13878

BIOMETRIC PRACTICE

Biometrics A JOURNAL OF THE INTERNATIONAL BIOMETRIC SOCIETY WILEY

A case study of glucose levels during sleep using multilevel fast function on scalar regression inference

Renat Sergazinov[1] | **Andrew Leroux**[2] | **Erjia Cui**[3] | **Ciprian Crainiceanu**[3] | **R. Nisha Aurora**[4] | **Naresh M. Punjabi**[5] | **Irina Gaynanova**[1]

[1]Department of Statistics, Texas A&M University, College Station, Texas, USA

[2]Department of Biostatistics & Informatics, University of Colorado Anschutz Medical Campus, Colorado, USA

[3]Department of Biostatistics, Johns Hopkins University, Baltimore, Maryland, USA

[4]New York University Grossman School of Medicine, New York, New York, USA

[5]Miller School of Medicine, University of Miami, Coral Gables, Florida, USA

Correspondence
Irina Gaynanova, Department of Statistics, Texas A&M University, College Station, TX, USA.
Email: irinag@stat.tamu.edu

Funding information
National Science Foundation, Grant/Award Number: NSF CAREER DMS-2044823

Abstract

Continuous glucose monitors (CGMs) are increasingly used to measure blood glucose levels and provide information about the treatment and management of diabetes. Our motivating study contains CGM data during sleep for 174 study participants with type II diabetes mellitus measured at a 5-min frequency for an average of 10 nights. We aim to quantify the effects of diabetes medications and sleep apnea severity on glucose levels. Statistically, this is an inference question about the association between scalar covariates and functional responses observed at multiple visits (sleep periods). However, many characteristics of the data make analyses difficult, including (1) nonstationary within-period patterns; (2) substantial between-period heterogeneity, non-Gaussianity, and outliers; and (3) large dimensionality due to the number of study participants, sleep periods, and time points. For our analyses, we evaluate and compare two methods: fast univariate inference (FUI) and functional additive mixed models (FAMMs). We extend FUI and introduce a new approach for testing the hypotheses of no effect and time invariance of the covariates. We also highlight areas for further methodological development for FAMM. Our study reveals that (1) biguanide medication and sleep apnea severity significantly affect glucose trajectories during sleep and (2) the estimated effects are time invariant.

KEYWORDS
actigraphy, CGM, diabetes, sleep apnea, wearables

1 | INTRODUCTION

Diabetes is a chronic disease characterized by elevated blood glucose levels. Type I diabetes results from the pancreas's inability to produce insulin, whereas type II is characterized by insulin resistance and insufficient amount of insulin. Diabetes is associated with considerable morbidity and mortality (Zimmet et al., 2001), and is linked to multiple complications (Kodl & Seaquist, 2008; Moxey et al., 2011; Resnick & Howard, 2002). In 2019, approximately 463 million people worldwide had diabetes , with type II diabetes constituting about 90% of the cases, and the rates are projected to rise (Saeedi et al., 2019). Therefore, it is important to understand the factors that affect

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.
© 2023 The Authors. Biometrics published by Wiley Periodicals LLC on behalf of International Biometric Society.

Downloaded from https://academic.oup.com/biometrics/article/79/4/3873/7587567 by guest on 13 August 2024

the development of diabetes and its progression with time. Obstructive sleep apnea (OSA) is a sleep-related breathing disorder that is prevalent among patients with type II diabetes (Foster et al., 2009). Multiple studies have reported associations between OSA, insulin resistance, and glucose intolerance (Lindberg et al., 2012; Punjabi et al., 2002). Despite this, the explicit effects of OSA severity on the glucose control of patients with type II diabetes are not well understood.

Traditionally, glucose control is quantified by Hemoglobin A1c (HbA_{1c}), which is a measure of the long-term average glucose levels. However, the glucose profiles are highly nonlinear and nonstationary, being sensitive to various environmental factors, including quantity and type of meals, medications, physical activity, and stress levels. As HbA_{1c} does not capture the within-night variability of glucose profiles, it cannot capture short-term and dynamic associations between exposures, such as OSA severity and blood glucose during sleep.

In this work, we aim to quantify the effects of OSA severity and medications on glucose control by analyzing measurements from continuous glucose monitors (CGMs) of patients with type II diabetes from the HYPNOS study (Rooney et al., 2021). Unlike HbA_{1c}, CGMs provide detailed quantification of blood glucose levels during the entire 24-h period (typical measurement interval is 5 min), thus playing an increasing role in clinical practice and disease management (Rodbard, 2016). A unique characteristic of the HYPNOS study is the availability of concurrent data from CGM and wearable actigraphy devices. For the purpose of this study, actigraphy was used to estimate sleep periods. Figure 1 provides examples of CGM profiles for three study participants during their estimated sleep periods. Time zero indicates the estimated sleep onset time, and time on the x-axis is the time from the sleep onset. Focusing on sleep periods is important as (1) OSA occurs during sleep and (2) sleep periods are less affected by confounding factors such as food intake or physical activity. However, when, what, and how much the person ate before going to sleep is likely to affect both the blood glucose level at the beginning and its dynamics during the sleep period. A subset of these data was previously analyzed in Gaynanova et al. (2022). However, this prior analysis focused on modeling variability using functional PCA and did not specifically investigate the potentially time-varying fixed effects of covariates, such as OSA severity or diabetes medications, on glucose profiles.

It is common practice to extract summary measures (e.g., mean, standard deviation, coefficient of variation) from CGM data and use these summaries in subsequent analyses (Broll et al., 2021; Rodbard, 2016). For example, a prior study of the HYPNOS data (Aurora et al., 2022) has used average glucose values over each sleep period and a linear mixed model to account for the multiple days of monitoring. Such approaches are easy to interpret as they use standard, well-known statistical models. The main disadvantage is that substantial information may be lost while compressing the CGM trajectory into a single number. Data compression may also make it impossible to test certain hypotheses of interest, such as whether the effects of covariates are time-varying.

In this paper, we focus on modeling the complete glucose profile as a functional response. We are interested in how these profiles are associated with OSA severity and medication. The methodology used for analysis is in the general area of multilevel function-on-scalar regression (Bigelow & Dunson, 2009; Di et al., 2009; Goldsmith et al., 2015; Greven et al., 2010; Morris & Carroll, 2006). The term function-on-scalar regression was introduced in Reiss et al. (2010). While estimation in multilevel function-on-scalar regression has been well studied, statistical inference remains an active area of research. In particular, the HYPNOS data have (1) large dimensionality with 174 study participants, an average of 10 sleep periods per participant, and 84 time points per period (data observed in 5 min intervals for the first 7 h from sleep onset); (2) highly nonstationary within-period patterns; and (3) substantial between-period heterogeneity with non-Gaussianity and outliers. The main problems with existing methods for such data are (1) computational feasibility and scalability, (2) availability of software, and (3) validity of resulting uncertainty estimates in the presence of within-curve functional correlations, outliers, and non-Gaussian errors. Our goal is to conduct inference on the HYPNOS data using computationally feasible methods that account for its multilevel functional structure and heterogeneity. We have identified only one method, the recently published fast univariate inference (FUI) (Cui et al., 2022), which can be adapted to achieve these goals. We compare this approach with the FAMMs framework (Scheipl et al., 2015) implemented in the **R** package **refund** (Goldsmith et al., 2020), which is the current state-of-the-art for functional data analysis (FDA) (Morris, 2015).

One major contribution is to compare FUI and FAMM in a realistic scenario with a new data type (CGM) and provide practical guidance on their implementation. We conclude that FUI is much faster than FAMM when accounting for the within-sleep period correlation structure of the data (fitting the FAMM model with such a structure is computationally prohibitive for the HYPNOS data). Moreover, even when within-sleep period correlations are ignored, FAMM is still very slow using the default settings (over 12 h on a standard laptop). We show how to change the FAMM default settings to substantially reduce computation time to 7 min. These changes may be known to FAMM experts, but most users would be unaware of

Downloaded from https://academic.oup.com/biometrics/article/79/4/3873/7587567 by guest on 13 August 2024

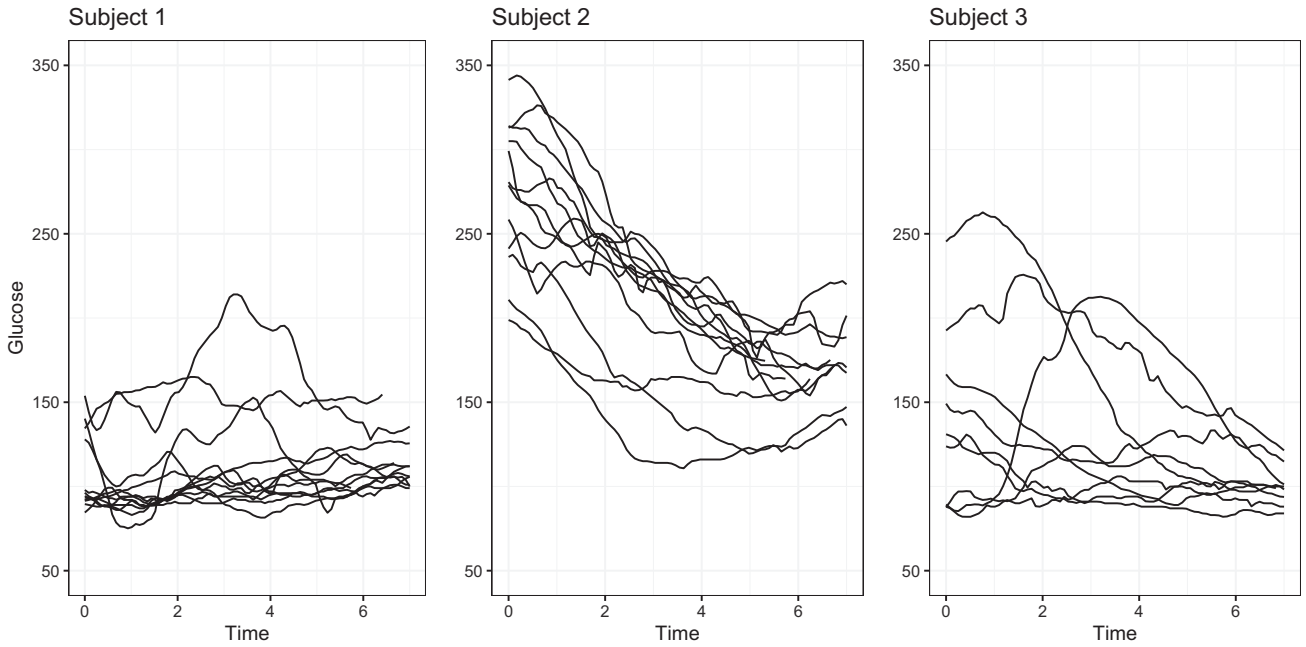

FIGURE 1 Glucose trajectories during sleep for three selected subjects. The *x*-axis is the time from estimated sleep onset, one observation every 5 min. The glucose values are measured in mg/dL.

where and how to adjust these settings. Our guidelines can substantially improve the user experience and, ultimately, the use of FAMM in practice. We also show that despite philosophical differences between the methods, FUI (accounting for within-period correlation) and FAMM (ignoring within-period correlation) lead to similar point estimators of fixed effects. The resulting FUI joint confidence bands are based on nonparametric case bootstrap, and thus, unlike FAMM, take into account within-curve functional correlations, outliers, and non-Gaussian distribution of the errors. While FAMM confidence bands are slightly narrower, they are pointwise rather than joint and could not account for within-sleep period correlations. This comparison is new, as the original FUI paper did not include simulations with correlated visit-specific data.

Moreover, we expand the utility of the FUI framework, proposing new methods for obtaining *p*-values for testing the global null hypotheses of (1) no fixed effects and (2) no time-invariant fixed effects. We implement simulation studies, confirming that the proposed approach has nominal rejection rates under the null and provide power curves for both tests under the alternative hypothesis. To our knowledge, this is the first time when joint confidence intervals are used to derive *p*-values for testing functional effects. For the no effect hypothesis on HYPNOS data, there are substantial differences between the newly proposed *p*-values and the approximate *p*-values using FAMM (Wood, 2013). This may be due to the misspecification of the FAMM model and correlations among the covariates. At this point, it is unclear how to conduct formal tests of time invariance based on FAMM, but they can be conducted using our novel proposed testing procedure.

In summary, our case study (1) characterizes the association between covariates and glucose trajectories as a function of time from sleep onset; (2) provides confidence intervals and measures of statistical significance of fixed effects, including *p*-values for testing time invariance; (3) evaluates and compares two analytic methods (FUI and FAMM); and (4) provides guidelines for the use of these methods in future studies. Such case studies are crucial to the ultimate success of FDA approaches. Indeed, despite their success in the statistical literature, FDA is rarely used in practice. Successful case studies that go beyond didactic examples could help improve the acceptability of FDA methods in realistic scenarios.

2 | DATA DESCRIPTION

2.1 | Data collection

The data analyzed in this paper were collected as part of the hyperglycemic profiles in obstructive sleep apnea (HYPNOS) randomized clinical trial. The study population consisted of adults between 21 and 75 years old with type II diabetes and mild-to-severe OSA recruited from the community. The primary objective of the trial was to determine whether treatment with positive airway pres-

Downloaded from https://academic.oup.com/biometrics/article/79/4/3873/7587567 by guest on 13 August 2024

sure (PAP) therapy is associated with improvements in glycemic measures. To investigate the effects of OSA severity and medication on the glucose control of patients with type II diabetes, we consider the data at the baseline visit (prior to the randomization of study participants into control and PAP therapy groups). The research protocol was approved by the Institutional Review Board on human research (Number: NA_00093188). A detailed description of the trial protocol and implementation can be found in Rooney et al. (2021). Below we provide a short summary.

Study participants were screened based on the point-of-care HbA_{1c} measured with a DC Vantage Analyzer (Siemens, Malvern, PA) and a home sleep apnea test using Apnealink (Resmed, San Diego, CA). The oxygen desaturation index (ODI) was determined using the number of times the oxyhemoglobin decreased by at least 4% per hour of sleep. Participants with $HbA_{1c} \geq 6.5\%$ and ODI $\geq$ 5 events pre hour were invited to enroll in the study. For each study participant, OSA severity was characterized as mild ($5 \leq$ ODI < 15) or moderate-to-severe (ODI ≥ 15). Exclusion criteria included pregnancy, any prior therapy for OSA, insulin use, change in glycemic medications in the previous 6 weeks, current oral steroid use, other sleep disorders, habitual sleep duration of < 6 h/night, and any unstable medical condition. Study participants completed an actigraphy study using the Actiwatch (Philips Respironics, Murraysville, PA) and continuous glucose monitoring (CGM) using the Dexcom G4 Platinum sensor, which produces one measurement every 5 min. The actigraph was worn on the nondominant wrist, and the CGM was placed 6 cm lateral to the umbilicus. Participants were instructed to wear both monitors for at least 7 days and provide calibration glucose data for the Dexcom sensor twice a day according to the manufacturer's instructions.

2.2 | Extraction of glucose curves corresponding to sleep periods

Sleep periods were estimated using the actigraphy data and the proprietary algorithm of the Phillips Actiware software. Actigraphy-estimated sleep periods that were shorter than 5 h were excluded from the analysis. Only participants who had at least five sleep periods with concurrent CGM measurements were included in our analyses. This led to 1812 actigraphy-estimated sleep periods for 174 study participants ranging from 5–20 sleep periods per study participant, with a median of 11 sleep periods. Rather than rescaling the time of each sleep period to account for different lengths (which may lead to distortion of circadian and sleep-related biological rhythms), we work with absolute time from the actigraphy-estimated sleep onset, in accordance with established practices in physiological studies of glucose regulation during sleep (Van Cauter et al., 1991). The median sleep period duration is 7.5 h with lower and upper quartiles of 7 and 8 h, respectively, and we focus on the first 7 h from the estimated sleep onset (time zero). The measurement times are synchronized across participants by linearly interpolating glucose trajectories at 5-min intervals from time zero. The interpolation interval matches the frequency of the CGM device, and the interpolated trajectories are visually indistinguishable from the original ones. Additional details on data processing and filtering steps are in Web Appendix A.

Figure 1 displays the glucose trajectories during the first 7 h of actigraphy-estimated sleep for three study participants (selected to illustrate the variety of BG profiles observed in the data). Each solid black line corresponds to one period of actigraphy-estimated sleep, and time zero corresponds to the actigraphy-estimated sleep onset. While typical glucose values range between 70 and 120 mg/dL for people without diabetes, the glucose values for patients with diabetes are much more variable, even during sleep. All three study participants in Figure 1 exhibit high glucose values, with Subject 2 having measurements in the [120, 350] mg/dL range. Furthermore, while glucose values are expected to decrease during sleep due to the absence of food intake, this decreasing trend is not consistently observed across all study participants and sleep periods. Subject 1 has trajectories that tend to increase throughout the night, with several trajectories having a peak in the middle of sleep. In contrast, all trajectories for Subject 2 are decreasing. For Subject 3, trajectories are highly variable across nights, with most, but not all, trajectories decreasing. For example, one trajectory starts at in-range values of 90 mg/dL, reaches hyperglycemic value of 210 mg/dL at 3 h from sleep onset, and goes back to around 100 mg/dL at 7 h.

2.3 | Research questions and statistical challenges

In addition to the glucose trajectories, multiple covariates are available: age, sex BMI (coded as 0 for BMI $<$ 35, and as 1 for BMI $\geq$ 35, corresponding to severe obesity), use of hypoglycemic medications (biguanides and sulfonylureas), point-of-care HbA_{1c}, and OSA severity status (mild OSA and moderate-to-severe OSA). Since point-of-care HbA_{1c} measurements were obtained prior to the device placements, we use HbA_{1c} as a baseline marker of diabetes severity. These covariates are completely observed for all 174 study participants, and Table 1 provides summary statistics.

Our primary scientific goal is to investigate the effects of OSA severity and hypoglycemic medications on glu-

Downloaded from https://academic.oup.com/biometrics/article/79/4/3873/7587567 by guest on 13 August 2024

TABLE 1 Descriptive statistics of the study participants. Values are the medians (min-max) for continuous covariates or size (%) for binary covariates.

Covariates	Full sample $n = 174$	
Age	61	(35–75)
Male Sex	95	(55%)
Biguanide use	145	(83%)
Sulfonylurea use	66	(38%)
Severe OSA	87	(50%)
BMI >= 35 (severe obesity)	60	(34%)
HbA_{1c}	7.2	(6.5–11.2)

cose trajectories during sleep after accounting for baseline HbA_{1c}, age, sex and BMI. Statistically, this is an inference question about the association between fixed scalar covariates (e.g., OSA severity) and functional response (CGM trajectory during sleep). However, this is not a standard function-on-scalar regression (terminology introduced by Reiss et al. (2010)) because the CGM exhibits: (1) highly nonstationary within-day patterns; (2) high between-day heterogeneity and substantial departures from the normality of the marginal distributions; and (3) multilevel structure with a large number of study participants (174) and time points (84). Our methodological goal is to conduct inference using computationally feasible methods that account for the known structure of the data.

3 | METHODOLOGY

We start by introducing the notation and model structure. Denote by $y_{ij}(t_k)$ the glucose measurement for subject $i = 1, \dots, I$, sleep period $j = 1, \dots, J_i$ (number of periods varies across study participants), at time t_k, $k = 1, \dots, K$, from sleep onset. In this application, we use an equally spaced time index for the 7 h interval from sleep onset. The data structure is the multilevel functional (Bigelow & Dunson, 2009; Crainiceanu et al., 2013; Di et al., 2009; Greven et al., 2010; Morris & Carroll, 2006; Meyer et al., 2015) because multiple functions are observed for each study participant. Let $\mathbf{x}_i = (x_{i1}, \dots, x_{iR})^\top$ be the $R \times 1$-dimensional vector of covariates for participant i and $X = (\mathbf{x}_1, \dots, \mathbf{x}_I)^\top$ be the $I \times R$-dimensional matrix, where each row contains the covariates for one study participant. We are interested in how these covariates affect the blood glucose trajectories from sleep onset.

We consider the following multilevel function-on-scalar regression model:

$$y_{ij}(t) = \beta_0(t) + \sum_{r=1}^{R} \beta_r(t) x_{ir} + b_i(t) + \epsilon_{ij}(t). \qquad (1)$$

Here, $\beta_0(t)$ is the global intercept and could be interpreted as the average blood glucose level at time t over study participants, i, and sleep periods, j, when covariates are equal to zero. The component $\beta_0(t) + \sum_{r=1}^{R} \beta_r(t) x_{ir}$ is the average fixed effect at time t across visits and subjects with the covariates $\mathbf{x}_i$. The functions, $\beta_r(\cdot)$ for $r = 0, \dots, R$, are assumed to be continuous and smooth. The random intercept $b_i(t)$ is the participant-specific deviation from the population mean at time t. It plays a similar role to that of random intercepts in standard linear mixed models but allows for a subject-specific shift at each time t, which is assumed to be smooth. More precisely, assume that $b_i(\cdot) \sim N(\mathbf{0}_K, \mathbf{\Sigma}_{b,K})$, mutually independent, where $\mathbf{0}_K$ is the $K \times 1$-dimensional vector of zeros, and $\mathbf{\Sigma}_{b,K}$ is the $K \times K$-dimensional between-participant covariance matrix after accounting for fixed effects. We also assume that $\epsilon_{ij}(\cdot) \sim N(\mathbf{0}_K, \mathbf{\Sigma}_{\epsilon,K})$, where $\mathbf{\Sigma}_{\epsilon,K}$ is the $K \times K$-dimensional within-participant covariance matrix. Note that it would be too restrictive to assume that $\epsilon_{ij}(t)$ are uncorrelated across t because model (1) would not be generative for the type of data observed in the HYPNOS study. In the actual data, the period-specific deviations from the participant-specific means are highly structured and smooth.

We consider two competing methods for estimation and inference of model (1): FAMM (Scheipl et al., 2015) and FUI (Cui et al., 2022). We review FAMM in Section 3.1 and FUI in Section 3.2. In Section 3.3, we build upon the FUI framework to develop p-values for fixed effects to test the null hypotheses of (1) no effect and (2) time-invariant effect.

3.1 | Estimation and inference with FAMM

The FAMM is based on the extension of the generalized additive models. For the fixed effects coefficient functions $\beta_r(\cdot), r = 0, \dots, R$, FAMM uses spline basis expansion. For the subject-specific random intercept $b_i(\cdot)$ and period-specific time-dependent $\epsilon_{ij}(\cdot)$, FAMM uses the tensor product expansion (Scheipl et al., 2015; Wood, 2006). The tensor product uses a Kronecker product to cross the bases for the random and time-dependent components. This creates a significant computational bottleneck as the total number of basis functions is a product between the number of participants, the number of sleep periods, and the number of basis functions for the time component. With 10 basis functions for the time component and 174 participants, the basis for the subject-specific random intercept $b_i(\cdot)$ consists of 1740 functions. With 1812 sleep periods, the basis for the period-specific time-dependent $\epsilon_{ij}(\cdot)$ consists of 18,120 functions. Given the size of the basis, we found it impossible to fit the full model (1) using FAMM

Downloaded from https://academic.oup.com/biometrics/article/79/4/3873/7587567 by guest on 13 August 2024

on our data. To achieve computational feasibility, we use FAMM to fit a simpler model that assumes that the residuals, $\epsilon_{ij}(\cdot)$, are independent across time. However, this model is misspecified as the CGM data have a strong within-sleep period structure (Figure 1). Web Appendix B provides guidelines for fitting both models with FAMM using refund package (Goldsmith et al., 2020).

FAMM produces pointwise confidence intervals for fixed effects based on Bayesian credible intervals. To obtain joint confidence intervals, Wood (2017) suggests using posterior simulations or the bootstrap procedure. However, both approaches require significant computational resources on our data. Even without access to joint confidence intervals, approximate p-values for testing the hypothesis of no effect can be calculated as in Wood (2013). The approximation, however, neglects the uncertainty in the smoothing parameters (Wood, 2013), which, combined with model misspecification, could lead to inflated rejection rates under the null (see simulation study in Web Appendix D).

3.2 | Estimation and inference with FUI

FUI fits univariate linear mixed models at each time point. To be specific, for each t from the discrete set $\{t_1, \dots, t_K\}$, FUI fits a univariate mixed effects model using, for example, **lme4** in the **R** (Bates et al., 2015). From the pointwise univariate models, we obtain estimators of the coefficient vectors $\hat{\boldsymbol{\beta}}_r = (\hat{\beta}_r(t_1), \dots, \hat{\beta}_r(t_K))$. These coefficients can then be smoothed; see, for example, Cui et al. (2022).

To conduct fixed-effects inference, FUI uses a nonparametric case bootstrap of study participants. For a bootstrap $b = 1, \dots, B$, study participants $\{1, \dots, I\}$ are sampled with replacement and estimators of the effects of interest $\{\hat{\boldsymbol{\beta}}_r^{(b)}\}$ are obtained. For each effect of interest $r = 1, \dots, R$, we obtain the mean, $\hat{\boldsymbol{\beta}}_r$, and covariance, $\text{var}(\hat{\boldsymbol{\beta}}_r)$, from the B bootstrap estimators. The approach proceeds by sampling $\hat{\boldsymbol{\beta}}_r^{(n)} \sim \mathcal{N}\{\hat{\boldsymbol{\beta}}_r, \text{var}(\hat{\boldsymbol{\beta}}_r)\}$ N times and calculating the distribution of the standardized maximum deviations $u_n = \max |\hat{\boldsymbol{\beta}}_r^{(n)} - \hat{\boldsymbol{\beta}}_r| / \sqrt{\text{diag}(\text{var}\{\hat{\boldsymbol{\beta}}_r\})}$. The empirical $1-\alpha$ quantile $q_{1-\alpha}$ of the distribution $\{u_1, \dots, u_N\}$ is used to compute the joint confidence bands for the effects as $\hat{\boldsymbol{\beta}}_r \pm q_{1-\alpha}\sqrt{\text{diag}(\text{var}(\hat{\boldsymbol{\beta}}_r))}$.

FUI implicitly allows for time-dependent error curves, $\epsilon_{ij}(\cdot)$, because the fitting is done separately across time points. The correlation of the errors is accounted for when calculating the confidence intervals based on the bootstrap of study participants. This is a simple but highly effective work-around for the direct modeling of highly complex covariance structures. It does not impact the pointwise confidence intervals but provides a way of calculating joint confidence intervals that account for complex dependence structures. Moreover, this approach allows for more flexible modeling of the covariance that could include, for example, latent subgroups with different dependence structures. In turn, joint confidence intervals can be used for global testing of parameter coefficients. While FUI is a powerful approach to inference, the original paper (Cui et al., 2022) does not provide (1) an assessment of the method in the presence of within-period correlations (which is the case for our HYPNOS data); (2) a one-number summary (e.g., p-value) for testing the global null hypotheses about the functional effects, $\beta_r(\cdot)$; and (3) a formal way to assess time invariance of fixed effects.

3.3 | Obtaining p-values from FUI for testing the no effect and time invariance hypotheses

FUI provides a way of constructing joint confidence intervals for $\beta_r(\cdot)$ at any level α. However, it does not provide a one-number summary (e.g., p-value) for testing the null and time invariance hypotheses about the functional effects. Specifically, given evaluation points $[t_1, \dots, t_K]$, and a fixed matrix $W \in \mathbb{R}^{m \times K}$, we are interested in testing the hypothesis:

$$H_0 : W\boldsymbol{\beta}_r = \mathbf{0},$$

where $\boldsymbol{\beta}_r = (\beta_r(t_1), \dots, \beta_r(t_K))^T$. The test of no effect corresponds to $W = I_K$. The test of time invariance corresponds to $W \in \mathbb{R}^{(K-1)\times K}$ being a first-order difference matrix:

$$W = \begin{bmatrix} 1 & -1 & & \\ & \ddots & & \\ & & 1 & -1 \end{bmatrix}.$$

To obtain p-values and perform the test, we propose to use the duality between the confidence intervals and the hypothesis tests. The original FUI paper (Cui et al., 2022) uses bootstrap to construct joint confidence interval for $W\boldsymbol{\beta}_r$ only when $W = I_K$. Our first contribution is to provide an extension of this approach to any W using the original bootstrap samples $\{\hat{\boldsymbol{\beta}}_r^{(b)}\}$. Second, we observe that a level α test rejects the null hypothesis if there exists at least one element in $W\boldsymbol{\beta}_r$ for which the corresponding $1-\alpha$ joint confidence interval does not contain zero. Thus, we propose to calculate the p-value for these tests as the smallest level α for which $\mathbf{0} \in \mathbb{R}^m$ is not inside the joint confidence interval for $W\boldsymbol{\beta}_r$. Web Appendix C provides details of the corresponding algorithm.

An advantage of having the p-values is that they provide an interpretable, quick summary of the evidence

Downloaded from https://academic.oup.com/biometrics/article/79/4/3873/7587567 by guest on 13 August 2024

against the null hypotheses, which complements the visual inspections of joint confidence intervals. Throughout this paper, we are using these newly proposed p-values. In Web Appendix D, we conduct simulation studies to empirically validate the proposed p-values for both no effect and time invariance tests. The p-values have approximately uniform distribution with close to nominal rejection rates under the null hypothesis and have good power under the alternative. We also examine whether the no effect test can be overly sensitive due to the $\beta_r(t) = 0$ for all t requirement and conclude that the test remains robust to small deviations from the null. To the best of our knowledge, this is the first time p-values are implemented based on the bootstrap joint confidence intervals.

4 | RESULTS

Our first question of interest is investigating the effects of OSA severity and hypoglycemic medications on glucose trajectories during sleep after accounting for baseline HbA_{1c}, age, sex and BMI. BMI is coded as 0 for BMI $<$ 35, and as 1 for BMI $\geq$ 35 (corresponding to severe obesity). OSA severity is coded as 0 for mild OSA and as 1 for moderate-to-severe OSA. For hypoglycemic medications, we consider the biguanide family (coded as 0 for no use and 1 for active use) and the sulfonylurea family (coded as 0 for no use and 1 for active use). Both biguanides and sulfonylureas are families of oral drugs commonly prescribed for type II diabetes to reduce glucose levels. Metformin is the most recognized drug in the biguanides group and is commonly used as a first-line treatment for type II diabetes. While biguanides work by suppressing the production of glucose in the liver, sulfonylureas promote the body's production of insulin. The two families of medications are not mutually exclusive. Out of 174 patients, 18 take neither type of drug, 90 take biguanide only, 11 take sulfonylurea only, and 55 take both drugs. Sex is modeled as a binary variable with 1 corresponding to males and 0 to females. Age and HbA_{1c} are modeled as continuous variables. We apply both FAMM and FUI methods to conduct inference on fixed-effect coefficients $\beta_r(t)$ in model (1).

Figure 2 displays the estimated coefficient curves, $\hat{\beta}_r(t)$, for each covariate, together with the corresponding 95% confidence bands. It is reassuring that the point estimators for FUI and FAMM are relatively close, though some differences can be observed. In particular, FUI estimates are, in general, more variable across time. For FAMM, we only display the pointwise confidence bands obtained under the misspecified model that does not account for the sleep-period-specific correlation. For FUI, we display both the pointwise and joint confidence bands (obtained as described in Section 3.2). FUI pointwise bands are, on average, 3% wider than FAMM across all covariates with a minimum of 1% (for biguanide and sulfonylurea) and a maximum of 4% (for all remaining covariates). FUI joint bands are, on average, 22% wider than FAMM, with a minimum of 11% (for sex) and a maximum of 57% (for HbA_{1c}). The FUI joint band for BMI is shorter than both FUI pointwise band and FAMM joint band. We suspect that this discrepancy is due to differences in the calculation of standard errors. FUI pointwise bands use standard errors from pointwise mixed models (corresponding to unsmoothed coefficients), whereas the joint bands use standard errors from bootstrap, where the coefficients are smoothed at each replication. If the true underlying coefficient is very smooth, bootstrap standard errors can be smaller.

TABLE 2 P-values for each fixed effect testing (1) $H_0^{(1)} : \beta(t) \equiv 0$ using FUI and FAMM, (2) $H_0^{(2)} : \beta(t) \equiv \text{const}$ using FUI. For FUI, we compute the values using the method in Section 3.3. For FAMM, we use the default method as in Wood (2013).

Covariates	$H_0^{(1)}$ (FUI)	$H_0^{(1)}$ (FAMM)	$H_0^{(2)}$ (FUI)
Age	0.008	0.085	0.809
Sex	0.099	0.126	0.968
BMI	0.832	0.728	0.901
OSA Severity	0.026	0.038	0.877
Biguanide	0.033	< 0.001	0.868
Sulfonylurea	0.673	0.024	0.736
HbA_{1c}	< 0.001	< 0.001	0.788

Next, we test the hypothesis of no effect based on FUI and FAMM p-values, as well as time invariance based on FUI. All p-values are reported in Table 2. For the no effect hypothesis, based on FUI, BMI and sex were not significant at the level $\alpha = 0.05$. As expected, HbA_{1c} is very strongly associated with the CGM curves (p-value$<$0.001). This makes sense, as HbA_{1c}, is thought to measure an average of glucose values over the past 3 months (Nathan et al., 2007). The FUI analysis seems to indicate that this association is stronger in the first part of the sleep period, though the association is statistically significant across the entire sleep period. Results suggest that patients with higher baseline HbA_{1c} have overall higher glucose levels. While the effect of sulfonylurea medication on glucose profiles during sleep is not significant (p-value=0.673), the effect of biguanide is (p-value=0.033). Specifically, biguanide is associated with lower glucose values during sleep after accounting for HbA_{1c}. Panels labeled biguanide in Figure 2 indicate that the effect is the strongest approximately 3 h after the sleep onset. A possible explanation may be that biguanide is typically taken with dinner in the evening, and the drug effect may be delayed. Age is also significant (p-value $= 0.008$), with higher age being associated with lower glucose values. One possible explanation is

Downloaded from https://academic.oup.com/biometrics/article/79/4/3873/7587567 by guest on 13 August 2024

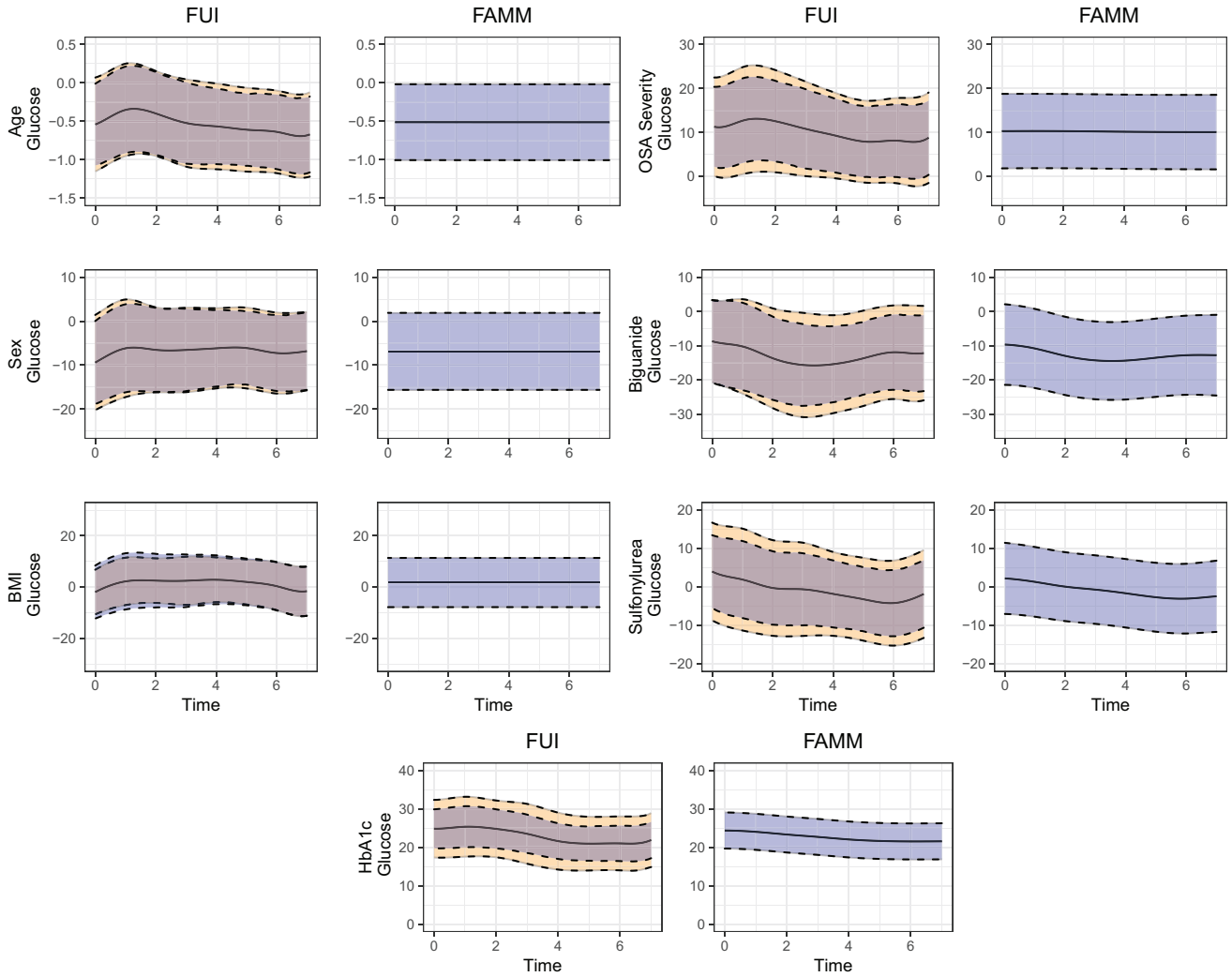

FIGURE 2 Estimated coefficient functions together with the 95% confidence intervals from fast univariate inference method (FUI) and functional additive mixed model (FAMM) described in Sections 3.2 and 3.1. For FAMM, the confidence intervals are pointwise. For FUI, both pointwise and joint confidence intervals are displayed, with joint intervals obtained as in Section 3.2. This figure appears in color in the electronic version of this article, and any mention of color refers to that version.

that older patients may have more experience in managing their diabetes. OSA severity has a significant negative effect on blood glucose even after accounting for HbA_{1c}, with larger glucose values during sleep for patients with moderate-to-severe OSA compared to patients with mild OSA. The *p*-values from FAMM lead to the same conclusions at 5% level for sex, BMI, OSA severity, biguanide, and HbA_{1c}. However, the methods differ in remaining covariates: FUI concludes that age is significant while sulfonylurea is not, whereas FAMM indicates vice versa. In general, FAMM gives considerably lower *p*-value estimates, up to 30 times lower, for both drug treatments: biguanide and sulfonylurea. This may be due to model misspecification and/or not unaccounted-for uncertainty in the smoothing parameters (see Web Appendix D for similar results on simulated data). In Web Appendix E, we provide additional FAMM fit diagnostics and a comparison with FUI, which support this conclusion. For the time-invariance hypothesis, based on FUI *p*-values, we cannot reject the null, corroborating the visual inspection of the plots in Figure 2 that suggest time invariance.

Overall, our empirical results highlight that FAMM and FUI methods result in similar widths of the confidence bands and comparable estimates of the coefficients functions, with FUI providing slightly wider joint confidence intervals. Further, FUI estimates are, in general, more variables across time. Both FUI and FAMM are powerful estimation approaches that are designed to work with complex dependencies in the functional data. In practice, however, fitting FAMM may be computationally challenging and may need case-by-case adjustment of the model specifications. Based on 10 runs, the average fitting time on a standard laptop for FUI was *less than 2 s*, and for FAMM with the options selected

Downloaded from https://academic.oup.com/biometrics/article/79/4/3873/7587567 by guest on 13 August 2024

in Web Appendix B was *7 min*. Without computation speedup, FAMM takes more than 12 h (the computations were interrupted at 12 h). Combining FUI with case bootstrap to perform inference leads to an overall computation time of 2 min with 100 bootstrap replications performed sequentially.

5 | DISCUSSION

In this paper, we consider the problem of inference in a multilevel function-on-scalar regression framework using two distinct methodologies, FAMM and FUI, in the context of glucose trajectories measured by CGM during sleep. Our results indicate that the glucose levels during sleep are significantly higher in patients with moderate-to-severe OSA compared to patients with mild OSA. Since heightened glucose levels in type II diabetes are associated with major adverse health effects, our findings suggest that OSA treatment options may need to be considered as part of an overall diabetes management plan in addition to traditional diet and exercise interventions. We have also found significant effects of participants' age, baseline diabetes severity as measured by HbA_{1c}, and biguanide medication. While heterogeneity in the times of sleep onset among participants may present potential confounding, we found the estimated effect sizes and conclusions to be similar when restricting the analysis to the sleep curves with the same time of sleep onset (Web Appendix E).

In terms of methods comparison, we found that FUI outperforms FAMM in terms of the fitting time and allows changes in model complexity without large losses in computational time. Indeed, fitting the full model (1) with FAMM is computationally prohibitive on our data without further methodological developments. When fitting the simplified FAMM model, FUI and FAMM provide similar estimates of the fixed effects. However, the FUI confidence bands are slightly wider as they are joint and account for within-period correlations, whereas the FAMM confidence bands are pointwise and are based on a misspecified model. We also provide an important improvement for FUI: computation of p-values for testing the null of no effect and time invariance for fixed effects based on joint confidence intervals. As the proposed p-values are intuitive, we expect that they will become popular in quantifying the significance of functional fixed effects in other applications.

There are multiple opportunities for further methodological and scientific research. First, the joint inference across the time domain is a unique feature of FUI, which FAMM currently does not support for large data sets due to computational limitations associated with model estimation. To perform joint inference, FUI relies on the bootstrap of the study participants, which is computationally efficient given its low model fitting cost. In contrast, fitting FAMM, even for a misspecified model, requires considerable computational resources, which reduces the appeal of the bootstrap. An alternative approach could be to specify and fit FAMM as a fully Bayesian model and conduct inference via posterior sampling. Second, while we focused on inference for fixed effects, a subject-specific inference could be of substantial interest in the context of the personalized assessment of glucose control during sleep. The main difficulties for subject-specific inference are (1) substantial between-subject variability and (2) the small number of sleep periods per subject. Third, it may be important to dynamically predict individual glucose trajectories and identify early unusual patterns that could be predictive of adverse health outcomes (e.g., hyper- or hypoglycemia).

ACKNOWLEDGMENTS

The authors thank the Editor, the AE, and two referees for valuable suggestions that help improve this manuscript. This work was partially supported by NSF DMS–2044823.

CONFLICT OF INTEREST

Dr. Crainiceanu is consulting with Bayer, Johnson and Johnson, and Cytel on methods development for wearable devices in clinical trials. The details of the contracts are disclosed through the Johns Hopkins University eDisclose system and have no direct or apparent relationship with the current paper.

DATA AVAILABILITY STATEMENT

Data used in this paper to support our findings were collected as part of the clinical trial (ClinicalTrials.gov Identifier NCT02454153), and are confidential. Supporting Information contains synthetically generated CGM data that mimics the characteristics of data analyzed in the paper.

ORCID

Irina Gaynanova https://orcid.org/0000-0002-4116-0268

REFERENCES

Aurora, R.N., Gaynanova, I., Patel, P. & Punjabi, N.M. (2022) Glucose profiles in obstructive sleep apnea and type 2 diabetes mellitus. *Sleep Medicine*, 95, 105–111.

Bates, D., Mächler, M., Bolker, B. & Walker, S. (2015) Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1–48.

Bigelow, J.L. & Dunson, D.B. (2009) Bayesian semiparametric joint models for functional predictors. *Journal of the American Statistical Association*, 104(485), 26–36.

Downloaded from https://academic.oup.com/biometrics/article/79/4/3873/7587567 by guest on 13 August 2024

Broll, S., Urbanek, J., Buchanan, D., Chun, E., Muschelli, J., Punjabi, N.M. et al. (2021) Interpreting blood glucose data with r package iglu. *PLoS One*, 16(4), e0248560.

Crainiceanu, C., Caffo, B. & Morris, J. (2013) Multilevel functional data analysis. In: Scott, M.A., Simonoff, J.S. & Marx, B.D. (Eds.) *The SAGE handbook of multilevel modeling*. London: SAGE Publications Ltd, pp. 223–248.

Cui, E., Leroux, A., Smirnova, E. & Crainiceanu, C.M. (2022) Fast univariate inference for longitudinal functional models. *Journal of Computational and Graphical Statistics*, 31(1), 219–230.

Di, C.-Z., Crainiceanu, C., Caffo, B. & Punjabi, N. (2009) Multilevel functional principal component analysis. *Annals of Applied Statistics*, 3(1), 458–488.

Foster, G.D., Sanders, M.H., Millman, R., Zammit, G., Borradaile, K.E., Newman, A.B. et al. (2009) Obstructive sleep apnea among obese patients with type 2 diabetes. *Diabetes Care*, 32, 1017–1019.

Gaynanova, I., Punjabi, N. & Crainiceanu, C. (2022) Modeling continuous glucose monitoring (CGM) data during sleep. *Biostatistics*, 23(1), 223–239.

Goldsmith, A., Zipunnikov, V. & Schrack, J. (2015) Generalized multilevel function-on-scalar regression and principal component analysis. *Biometrics*, 71(2), 344–353.

Goldsmith, J., Scheipl, F., Huang, L., Wrobel, J., Di, C., Gellar, J. et al. (2020) *refund: regression with functional data*. R package version 0.1-23.

Greven, S., Crainiceanu, C.M., Caffo, B. & Reich, D. (2010) Longitudinal functional principal component analysis. *Electronic Journal of Statistics*, 4, 1022–1054.

Kodl, C.T. & Seaquist, E.R. (2008) Cognitive dysfunction and diabetes mellitus. *Endocrine Reviews*, 29(4), 494–511.

Lindberg, E., Theorell-Haglöw, J., Svensson, M., Gislason, T., Berne, C. & Janson, C. (2012) Sleep apnea and glucose metabolism a long-term follow-up in a community-based sample. *Chest*, 142(4), 935–942.

Meyer, M.J., Coull, B.A., Versace, F., Cinciripini, P. & Morris, J.S. (2015) Bayesian function-on-function regression for multilevel functional data. *Biometrics*, 71(3), 563–574.

Morris, J.S. (2015) Functional regression. *Annual Review of Statistics and Its Application*, 2, 321–359.

Morris, J.S. & Carroll, R.J. (2006) Wavelet-based functional mixed models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 68(2), 179–199.

Moxey, P., Gogalniceanu, P., Hinchliffe, R., Loftus, I., Jones, K., Thompson, M. et al. (2011) Lower extremity amputations–a review of global variability in incidence. *Diabetic Medicine*, 28(10), 1144–1153.

Nathan, D., Turgeon, H. & Regan, S. (2007) Relationship between glycated haemoglobin levels and mean glucose levels over time. *Diabetologia*, 50(11), 2239–2244.

Punjabi, N.M., Sorkin, J.D., Katzel, L.I., Goldberg, A.P., Schwartz, A.R., & Smith, P.L. (2002) Sleep-disordered breathing and insulin resistance in middle-aged and overweight men. *American Journal of Respiratory and Critical Care Medicine*, 165(5), 677–682.

Reiss, P.T., Huang, L. & Mennes, M. (2010) Fast function-on-scalar regression with penalized basis expansions. *The International Journal of Biostatistics*, 6(1), Article 28.

Resnick, H.E. & Howard, B.V. (2002) Diabetes and cardiovascular disease. *Annual Review of Medicine*, 53(1), 245–267.

Rodbard, D. (2016) Continuous glucose monitoring: a review of successes, challenges, and opportunities. *Diabetes Technology & Therapeutics*, 18(S2), S2–3.

Rooney, M.R., Aurora, R.N., Wang, D., Selvin, E. & Punjabi, N.M. (2021) Rationale and design of the hyperglycemic profiles in obstructive sleep apnea (hypnos) trial. *Contemporary Clinical Trials*, 101, 106248.

Saeedi, P., Petersohn, I., Salpea, P., Malanda, B., Karuranga, S., Unwin, N. et al. (2019) Global and regional diabetes prevalence estimates for 2019 and projections for 2030 and 2045: results from the international diabetes federation diabetes atlas. *Diabetes Research and Clinical Practice*, 157, 107843.

Scheipl, F., Staicu, A.-M. & Greven, S. (2015) Functional additive mixed models. *Journal of Computational and Graphical Statistics*, 24(2), 477–501.

Van Cauter, E., Blackman, J.D., Roland, D., Spire, J.P., Refetoff, S., Polonsky, K.S. et al. (1991) Modulation of glucose regulation and insulin secretion by circadian rhythmicity and sleep. *The Journal of Clinical Investigation*, 88(3), 934–942.

Wood, S.N. (2006) Low-rank scale-invariant tensor product smooths for generalized additive mixed models. *Biometrics*, 62(4), 1025–1036.

Wood, S.N. (2013) On p-values for smooth components of an extended generalized additive model. *Biometrika*, 100(1), 221–228.

Wood, S.N. (2017) *Generalized additive models: an introduction with R*. Boca Raton, FL: CRC Press.

Zimmet, P., Alberti, K. & Shaw, J. (2001) Global and societal implications of the diabetes epidemic. *Nature*, 414(6865), 782–787.

SUPPORTING INFORMATION

Web Appendices referenced in Sections 2.2–5 and **R** codes with corresponding analyses are available with this paper at the Biometrics website on Wiley Online Library.

How to cite this article: Sergazinov, R., Leroux, A., Cui, E., Crainiceanu, C., Aurora, R.N., Punjabi, N.M. et al. (2023) A case study of glucose levels during sleep using multilevel fast function on scalar regression inference. *Biometrics*, 79, 3873–3882. https://doi.org/10.1111/biom.13878

Downloaded from https://academic.oup.com/biometrics/article/79/4/3873/7587567 by guest on 13 August 2024