Using A Multi-Level B-Spline Model to Analyze and Compare Patient Glucose Profiles Based on Continuous Monitoring Data

Hui Zheng, Ph.D.,[1,2] David M. Nathan, M.D.,[2,3] and David A. Schoenfeld, Ph.D.[1,2]

Abstract

Objective: We show how continuous glucose monitoring (CGM) data can be analyzed using a three-level B-spline model, facilitating the estimation of inter-patient variability, within-patient inter-day variability, and measurement error. We propose methods for statistical comparison of glucose profiles among patient groups.

Methods: We applied a three-level random effects model using quadratic B-spline functions to analyze inter-patient and within-patient inter-day variations of the glucose trend. The estimated SD values of the glucose curves are time-dependent and were averaged over a 24-h period. We analyzed CGM data from 322 patients with type 1 diabetes, 223 patients with type 2 diabetes, and 86 subjects without diabetes using interstitial glucose levels measured every 5 min, for approximately 8 days per patient. We compared group-wide glucose profiles from the insulin pump–treated ($n = 124$) and multiple daily injection (MDI)–treated ($n = 144$) patients with type 1 diabetes.

Results: The average inter-patient SD values were 49 mg/dL, 43 mg/dL, and 15 mg/dL for type 1 diabetes patients, type 2 diabetes patients, and subjects without diabetes, respectively. The average within-patient, inter-day SD values were 67 mg/dL, 41 mg/dL, and 18 mg/dL, respectively. The residual SD values were 19 mg/dL, 14 mg/dL, and 8 mg/dL, respectively. We identified a statistically significant difference in glucose profiles during the morning between insulin pump–treated and MDI-treated type 1 diabetes patients.

Conclusions: B-spline models facilitate the analysis of CGM data and show that type 1 diabetes is associated with higher inter-day glucose variation than type 2 diabetes or being without diabetes. Pump therapy and MDI have different effects on glucose control during specific time periods.

Introduction

THE INTRODUCTION OF Food and Drug Administration–approved continuous glucose monitoring (CGM) devices that provide frequent (every 1–5 min) measurements of interstitial glucose levels has provided a powerful new tool for treating patients and studying blood glucose (BG) control strategies. Unlike techniques with infrequent sampling, which may provide unrepresentative and therefore misleading results, CGM provides more frequent measurements throughout the day, allowing a more comprehensive and detailed assessment of a patient's glucose trend. With this capability, CGM systems have been shown to be useful in improving BG control in selected diabetes patient populations.[1–3] However, the clinical role of these devices is still being determined, and how clinicians and patients can best utilize the mass of data, taking into account the frequent intra- and inter-day fluctuations and measurement error, remains unclear.

Data smoothing techniques including kernel smoothing and low-order polynomial regression[4] can reduce noise and provide a more stable indication of diurnal BG trends. Developments in spline-based statistical models[5,6] have led to techniques that balance the need to fit the data and to penalize the roughness of the estimated function. Rice and Wu[7] proposed using mixed effects models with spline basis functions to estimate a population mean function and the covariance. Mazze et al.[8] applied smoothing algorithms on percentiles of point-wise CGM measurements to obtain the median and confidence intervals of an individual patient's daily glucose pattern.

CGM data can be viewed as having multiple levels of variation. First, many days of CGM data from a patient often demonstrate a substantial degree of inter-day variability. The assessment of a 24-h curve representing the patient's typical daily glucose pattern would provide the basis for developing a glucose control strategy. This requires summarizing measurements in multiple days into a single representative trend.

[1]Biostatistics and [3]Diabetes Centers, Massachusetts General Hospital, Boston, Massachusetts.
[2]Harvard Medical School, Boston, Massachusetts.

It is also important to estimate the expected amount of inter-day variation in order to detect any significant deviation from the target range. Second, CGM data present varying patterns from patient to patient. For research purposes and in order to identify outliers in clinical practice, methodologies are needed to estimate a group-wide "average profile" and the amount of inter-patient variation. Finally, measurement errors must be factored into the assessment of self-monitoring clinical data.

In this report we use a modeling strategy similar to the one of Rice and Wu[7] by extending to a three-level random effects model. We focus on estimating patient-level and group-level profiles and their variability. We propose statistical methodology for comparison of profiles between groups. This model can accomplish the following goals: (1) smoothing CGM measurements into spline curves; (2) estimating an individual patient's glucose profiles; (3) estimating group-wide glucose profiles; (4) estimating inter-day variation within a patient; (5) estimating inter-patient variation; and (6) drawing statistical comparisons among groups of patients who differ in characteristics such as disease type, treatment method, or genetic trait.

The proposed analysis of the frequent glucose levels measured with CGM should help clinicians make sense of these data, provide insights into the patterns of glycemia and the treatments necessary to control glucose, and serve as a useful analytic tool in research studies.

Subjects and Methods

CGM data were obtained from the A1c-Derived Average Glucose (ADAG) study, an international observational study designed to establish a reliable regression equation to translate hemoglobin A1c levels into estimated average glucose levels. ADAG recruited patients with type 1 and type 2 diabetes and subjects without diabetes from centers in Asia, Europe, and North America. CGM was performed for at least 2 days at baseline and at the end of months 1, 2, and 3.[9] The MiniMed Medtronic (Northridge, CA) device (MiniMed CGMS® System Gold™ model MMT-7102W) measures interstitial glucose levels every 5 min, resulting in 288 values recorded each day of CGM. The CGM data were downloaded electronically and transmitted to the data coordinating center in Boston, MA for analysis.

The three-level B-spline model

We modeled a glucose curve as the linear combination of 14 quadratic B-spline basis functions. Quadratic B-splines are piece-wise quadratic functions that are continuous to the first derivative and have discontinuities in the second derivative at some points called knots. These basis functions can be derived recursively using the De Boor algorithm[10–12] or using the explicit form given in Appendix A. Figure 1 gives a visual illustration of three basis functions, $B_3(t)$, $B_4(t)$, and $B_5(t)$.

We used a three-level quadratic B-spline model to model the group-wide, patient-level, and intra-day BG variation.

First, we assume Patient j's BG level at time $t \in (0–24)$ h on Day k is $BG_{jk}(t) = f_{jk}(t) + \varepsilon_{jk}(t)$, where $f_{jk}(t)$ is a quadratic spline function:

$$f_{jk}(t) = \sum_{l=1}^{14} a_{ljk} B_l(t) \tag{1}$$

with basis functions $B_l(t), l = 1, \ldots, 14$ and parameters $(a_{1jk}, \ldots, a_{14jk})$. The 14 basis functions were based on 17 evenly
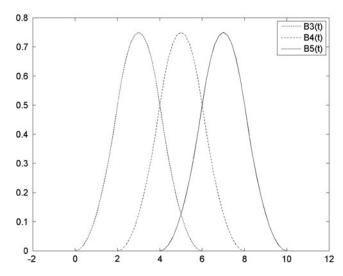

FIG. 1. Three quadratic B-spline basis functions.

spaced knots at $-4, -2, 0, 2, \ldots, 24, 26, 28$ h. The knots at -4 and -2 and at 26 and 28 h were added to aid the computation of basis functions in the (0–4) and (20–24) h intervals, but did not bear physical interpretation. We assumed the errors $\varepsilon_{jk}(t)$ were independent and identically distributed (i.i.d.) normally distributed with mean zero and variance σ^2.

Next, we assumed that the daily spline coefficients of Patient j were normally distributed around the patient mean with a common covariance matrix, i.e.,

$$(a_{1jk}, \ldots, a_{14jk}) \sim \text{Normal}((a_{1j}, \ldots, a_{14j}), \Phi) \tag{2}$$

Using this structure, we modeled that Patient j's daily glucose trends vary around a mean trend function (i.e., a patient-level profile) of $f_j(t) = \sum_{l=1}^{14} a_{lj} B_l(t)$. We assumed that all within-patient, inter-day covariance matrices were the same (Φ). If necessary and computationally feasible, this assumption can be relaxed to allow the variance matrix to vary with patient.

Finally, we assumed that the patient-specific parameters were normally distributed around a group-wide mean with covariance matrix Σ,

$$(a_{1j}, \ldots, a_{14j}) \sim \text{Normal}((a_1, \ldots, a_{14}), \Sigma) \tag{3}$$

This structure assumes that patient-specific trend functions vary around a group-wide trend function (i.e., a group-wide profile):

$$f(t) = \sum_{l=1}^{14} a_l B_l(t)$$

At a typical time point, three consecutive quadratic B-spline basis functions take non-zero values, suggesting that the adjacent parameters are correlated with $(a_{l_1jk}, a_{l_2jk}) \neq 0$ for any $|l_1 - l_2| \leq 2$. An examination of the unrestricted autocorrelation functions revealed a banded covariance structure for the intra-day glucose curve (Fig. 2A) and an unstructured covariance matrix for patient-level curve (Fig. 2B). Numerical comparisons showed that unstructured and banded Toeplitz within-patient inter-day Φ matrices[12–14] lead to very similar estimations for individual and group-wide profiles. We used
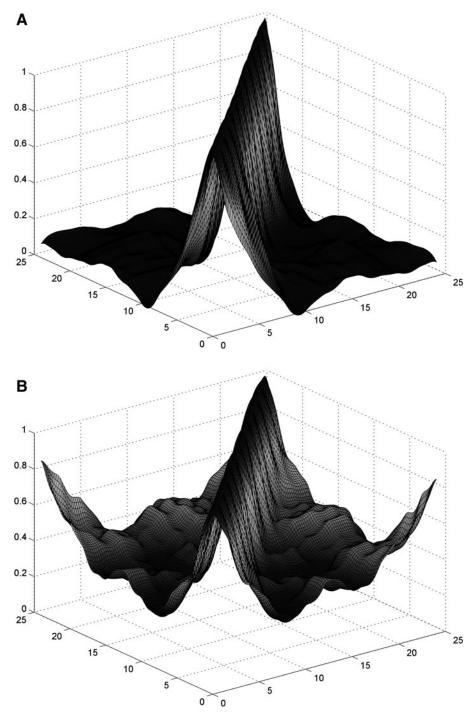
FIG. 2. Autocorrelation of (**A**) the daily glucose curve and (**B**) the patient average glucose profile for type 1 diabetes. The z-axis represents the correlation coefficient between glucose values at a time point on the x-axis and a time point on the y-axis. The time unit is h.

a Toeplitz (3) structure as the within-patient inter-day co-variance Φ for its efficiency in computer memory and CPU time usage without sacrificing numerical accuracy.

This model can be parameterized as a linear mixed effects model[15–17] and fitted using conventional statistical software such as SAS version 9.2 (SAS Institute, Inc., Cary, NC). In SAS PROC MIXED, this model requires two "random" statements to specify the patient-level and within-patient, day-to-day random effects. Some sample SAS code is given in Appendix B.

One can view the proposed model as a hierarchical Bayesian model with normal priors for the B-spline coefficients and use Bayesian algorithms to draw inferences.

Model-based variance estimation of the glucose profiles

Using the proposed hierarchical model, we were able to estimate the inter-patient and inter-day covariance matrices of

model coefficients and calculate the point-wise variance of the mean glucose functions.

The inter-patient variance of BG function at time point t was given by

$$Var(f_j(t)) = Var\left(\sum_{l=1}^{14} a_{lj}B_l(t)\right) = B(t)^T \Sigma B(t) \qquad (4)$$

where $B(t)$ is the 14×1 matrix containing the values $B_1(t), \ldots, B_{14}(t)$.

Similarly, for Patient j, the inter-day variation of BG function at time point t was given by

$$Var(f_{jk}(t)|j) = Var\left(\sum_{l=1}^{14} a_{ljk}B_l(t)\right) = B(t)^T \Phi B(t) \qquad (5)$$

The above two quantities were estimated by replacing Σ and Φ by their maximum likelihood (ML) or restricted maximum likelihood (REML) estimates, $\hat{\Sigma}$ and $\hat{\Phi}$, respectively.

The SE of the group-wide function and patient-specific mean function can be estimated by replacing the variance–covariance matrices in the above equations by the estimated variance–covariance matrices of $(\hat{a}_{1j}, \ldots, \hat{a}_{14j})$ and $(\hat{a}_{1jk}, \ldots, \hat{a}_{14jk})$, respectively.

The matrices $\hat{\Sigma}$ and $\hat{\Phi}$, the covariance matrices of $(\hat{a}_{1j}, \ldots, \hat{a}_{14j})$ and $(\hat{a}_{1jk}, \ldots, \hat{a}_{14jk})$, and the error variance σ^2 can all be estimated under the mixed effects model theory.

Validation of the model-based SE estimation

The aforementioned model-based variance estimation requires fairly accurate estimation of the variance–covariance matrices. A robust, but computationally intensive, alternative is the replication-based method, which does not involve the estimation of covariance matrices. We used a replication-based method to verify the model-based estimation.

We applied the following bootstrap method to estimate the SE of group-wide profiles. We created 10 replication samples by drawing with replacement from the sample the same number of patients. We applied the three-level model to each replication sample to obtain a replicate estimate of the mean glucose trend function. From the replicate estimates, we calculated the bootstrap estimate of the point-wise SE of the mean trend function. To check the validity of the model-based SE, we compared the bootstrap variance estimation with the model-based estimation on the type 1 diabetes patients' data.

Comparison of glucose profiles between patient groups

In order to demonstrate the utility of the proposed model in comparing BG profiles, we compared two groups of type 1 diabetes patients: group 1 was insulin pump–treated, and group 2 was treated with multiple daily injection (MDI) therapy. The goal was to assess how much, on average, the use of an insulin pump influenced the glucose profile in type 1 diabetes. In order to perform this exemplary comparison, we defined a group indicator I_j which was 0 if Patient j was in group 1 and 1 if the patient was in group 2.

We introduced 14 additional parameters $(\Delta_1, \ldots, \Delta_{14})$ that represented the differences in spline coefficients between group 2 and group 1. The group model (3) was then modified

to be $(a_{1j}, \ldots, a_{14j}) \sim$ Normal $((a_1 + \Delta_1 I_j, \ldots, a_{14} + \Delta_{14} I_j), \Sigma)$ so that group 1 and group 2 were modeled to have different profiles. In the mixed effects model, parameters $(\Delta_1, \ldots, \Delta_{14})$ were treated as fixed effects. One can use the Hotelling T-square test[18] for the null hypothesis that $\Delta_1 = \Delta_2 = \ldots = \Delta_{14} = 0$. One can use the Bayesian model to allow either identical or different covariance matrices in the comparison groups. For pump versus MDI comparison, we assume the two groups have the same covariance matrices.

The difference of the two spline functions is still a spline function with parameters $(\Delta_1, \ldots, \Delta_{14})$. The estimates $(\hat{\Delta}_1, \ldots \hat{\Delta}_{14})$ and their covariance matrix are useful for making various comparisons. Point-wise estimates and credible intervals (CI) (Bayesian language for confidence interval) of the difference between the two profiles can be obtained at any time point. The difference can be plotted versus time in the form of a spline function (D-curve) and its CI. From the 95% CI of the D-curve one can identify time periods in which two profiles are significantly different.

The area under the curve (AUC) values of two group-wide trend functions can also be compared. This requires the calculation of AUC for each basis function, which equals $\frac{1}{3}$ for $B_1(t)$ and $B_{14}(t)$, $\frac{5}{3}$ for $B_2(t)$ and $B_{13}(t)$, and 2 for $B_3(t)$ to $B_{12}(t)$.

Comparisons of nonlinear quantities such as maximum and minimum differences are possible using simulations. One can draw from the joint posterior distribution of $(\Delta_1, \ldots, \Delta_{14})$ and obtain distributions of such quantities for statistical inference.

Results

We analyzed CGM data from three groups of patients separately using the proposed model: 322 patients with type 1 diabetes, 223 patients with type 2 diabetes, and 86 subjects without diabetes. Among the type 1 diabetes patients for whom treatment was consistent during the study, 124 were treated with an insulin pump, and 144 were treated with MDI. The median age of the patients was 45 years (range, 16–70 years), and 53% were female. The characteristics of the study group are shown in Table 1.

Using the three-level spline model, we were able to obtain patient-specific and group-wide glucose functions for type 1 and type 2 diabetes and for the subjects without diabetes. We were also able to estimate the 95% CI for each diabetes type.

Table 2 gives the 24-h average inter-day and inter-patient variations in the three groups: type 1 diabetes, type 2 diabetes, and subjects without diabetes. Type 1 and type 2 diabetes patients had similar inter-patient variability with an average SD of approximately 49 and 43 mg/dL, respectively, compared with 15 mg/dL for the subjects without diabetes. Type 1 diabetes patients showed higher inter-day variability than type 2 patients (SD = 67 vs. 41 mg/dL, $P < 0.001$ according to the estimated asymptotic distribution of the covariance parameters). Type 1 diabetes patients' inter-day variability was higher than their inter-patient variability (SD = 67 vs. 49 mg/dL). Type 1 diabetes had the highest residual SD of 19 mg/dL versus 14 mg/dL for type 2 diabetes and 8 mg/dL for subjects without diabetes.

The model-based SE estimate was close to the bootstrap estimates for type 1 diabetes patients. The 24-h average model-based SE estimate was 3.1 mg/dL, which is close to the bootstrap estimate of 2.9 mg/dL.

| | Type 1 diabetes | | | | |
	Overall*	Pump-treated	MDI-treated	Type 2 diabetes	Without diabetes
Number	322	124	144	223	86
Age (mean ± SD) (years)	43 ± 13	43 ± 13	44 ± 13	55 ± 9	40 ± 14
Number (%) of females	166 (52%)	70 (57%)	70 (49%)	111 (50%)	59 (69%)
HbA1c (mean ± SD) (%)	7.3 ± 1.2	7.2 ± 1.1	7.4 ± 1.2	6.8 ± 1.1	5.2 ± 0.3

*Includes patients whose treatment information was not available or not determined.
HbA1c, hemoglobin A1c.

The quadratic B-spline model leads to smooth curves tightly tracking the individual patient's CGM measurements. Figure 3 demonstrates the smoothed intra-day BG curve, and its 95% CI overlapped with the measured BG of a type 1 diabetes patient over a 24-h period.

Figure 4A shows the group-wide 24-h glucose profile for type 1 diabetes patients treated with an insulin pump. The group-wide glucose curve of such patients trended within the 145–180 mg/dL range and had peaks at three time points: at approximately 9 a.m., 3 p.m., and midnight, corresponding to the effects of three meals. By comparison, the group-wide 24-h glucose curves for type 1 diabetes patients treated with MDI trended within the 145–190 mg/dL range and peaked at approximately 9–10 a.m. and midnight (Fig. 4B).

Figure 5 illustrates the estimated difference between MDI-treated and pump-treated type 1 diabetes patients' glucose profiles and the 95% CI of the differences. Figure 5 shows that glucose levels of MDI-treated patients are significantly higher than those of pump-treated patients from around 6 a.m. to around 10 a.m. The Hotelling[18] T-square test for the equality of the two group level profiles leads to a P value of 0.14, suggesting no significant difference between the two group-wide profiles as a whole.

Discussion

The three-level B-spline model provides a framework for the smoothing and inference of the dense glucose data provided by CGM. The model allows us to estimate the patient-level and group-level glucose profiles and their variability, as well as to make statistical comparisons of profiles among different groups.

The model yields smooth curves tightly tracking the CGM measurements. Not surprisingly, the population mean glucose profiles reflect the well-recognized effects of eating on glycemic excursions. The model provides insights into the different patterns of glycemia achieved with therapy in type 1 diabetes and, potentially, the timing and types of interventions necessary to address the glucose excursions.

TABLE 2. AVERAGE INTER-PATIENT AND WITHIN-PATIENT, INTER-DAY SD OF BLOOD GLUCOSE PROFILE BY GROUP

| | Average SD (mg/dL) | | |
	Inter-patient	Inter-day	Residual
Type 1 diabetes	49	67	19
Type 2 diabetes	43	41	14
Without diabetes	15	18	8

The evenly spaced knots are a natural fit for the CGM measurements by including the same number of measurements in each time interval between consecutive knots. Statistical criteria for selection of number and placement of knots have been widely discussed. Such methods include Akaike information criterion, Bayesian information criterion, cross-validation, and Bayesian model selection methods such as the reversible jump Markov chain Monte Carlo method.[19] In practice, the choice of knots should be based on a combination of statistical and practical considerations. In a situation where more detailed examination of the glucose curve is necessary, a larger number of knots can be used.

The choice of covariance structures has a limited impact on the estimated shapes of trend functions. Mis-specified covariance structures still lead to unbiased, albeit not the most efficient, estimates of the glucose curves. However, the covariance structures play an important role in variance estimation and hypothesis testing. This model assumes i.i.d. Gaussian random errors with mean zero and does not account for correlations among measurement errors.

Although the replication-based method provides robust variance estimation, it is computationally resource-demanding, making it infeasible for large datasets. In contrast, the model-based method is computationally efficient for large datasets without sacrificing the quality of variance estimation.

The proposed model allows statistical comparisons of point-wise and overall differences between two (or more) profiles, thus providing a useful statistical tool for future clinical trials using glucose profiles as an endpoint. Power calculations for such statistical tests are possible using estimated variance components and the expected effect size.

The B-spline basis functions demonstrated a numerical advantage in variance estimation over an alternative, the truncated polynomial functions. B-splines lead to more sparse and well-conditioned covariance matrices, whereas truncated polynomials are more prone to yield unsatisfactory variance estimation.

For large samples, the three-level random effects model is resource-demanding in terms of computer memory and CPU time. It would have been impossible to fit such a model on the ADAG data using a regular PC a few years ago. However, with the computational power of today's personal computers, such models can be fitted on sizeable samples in a reasonable amount of time using a common PC.

This model could be used in software to help patients better control their glucose. The mathematical methods for this are beyond the scope of this article, but they would involve using variance covariance matrix as part of a prior distribution for the individual parameters that define a patient's daily profile.
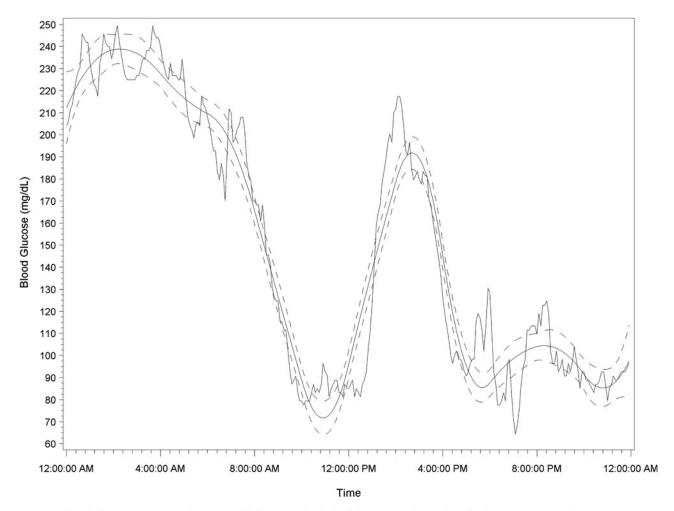
FIG. 3. Blood glucose curve and 95% credible interval (dashed lines) overlapped with the continuous glucose monitoring measurement of a type 1 diabetes patient over a 24-h period.

Then the patient's daily profile can be estimated from several weeks of data. This profile, along with a record of insulin injections, food intake, and exercise, can be used to determine behavior changes that might improve the patient's glucose control. We contemplate clinical trials comparing such a strategy to other treatment strategies. The methods we propose could also help improve estimation of instantaneous insulin values by using the model parameters in a prior distribution for each reading. This would involve building software into the device.
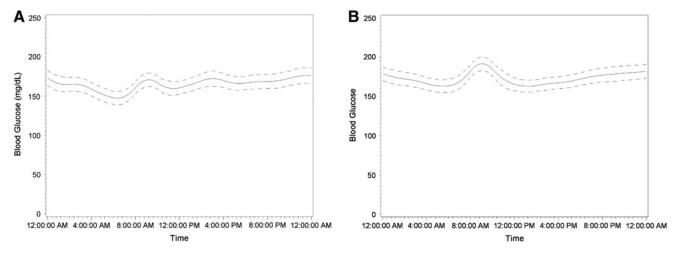
FIG. 4. Group-wide blood glucose curve and 95% credible interval (dashed lines) of type 1 diabetes patients treated with (**A**) an insulin pump and (**B**) multiple daily injections.
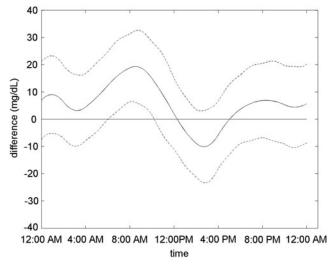
FIG. 5. Estimated mean and 95% credible interval (dashed lines) of the difference in glucose profiles between multiple daily injection–treated and insulin pump–treated type 1 diabetes patients. Higher values represent higher values in multiple daily injection–treated compared with insulin pump–treated patients.

The B-splines-based Bayesian modeling techniques provide a promising tool for clinical decision-making and ultimately for the development of "closed-loop" systems.

Author Disclosure Statement

None of the authors has a conflict of interest regarding the submitted manuscript.

References

1. Deiss D, Bolinder J, Riveline JP, Battelino T, Bosi E, Tubiana-Rufi N, Kerr D, Phillip M: Improved glycemic control in poorly controlled patients with type 1 diabetes using real-time continuous glucose monitoring. Diabetes Care 2006;29:2730–2732.
2. Tamborlane WV, Beck RW, Bode BW, Buckingham B, Chase HP, Clemons R, Fiallo-Scharer R, Fox LA, Gilliam LK, Hirsch IB, Huang ES, Kollman C, Kowalski AJ, Laffel L, Lawrence JM, Lee J, Mauras N, O'Grady M, Ruedy KJ, Tansey M, Tsalikian E, Weinzimer S, Wilson DM, Wolpert H, Wysocki T, Xing D: Continuous glucose monitoring and intensive treatment of type 1 diabetes. N Engl J Med 2008;359:1464–1476.
3. Tamborlane WV, Beck RW: Continuous glucose monitoring in type 1 diabetes mellitus. Lancet 2009;373:1744–1746.
4. Boyne MS, Silver DM, Kaplan J, Saudek CD: Timing of changes in interstitial and venous blood glucose measured with a continuous subcutaneous glucose sensor. Diabetes 2003;52:2790–2794.
5. Eilers P, Marx B: Flexible Smoothing with B-splines and penalties. Stat Sci 1996;11:89–121.
6. Ruppert D, Carroll R: Spatially adaptive penalties for spline fitting. Aust N Z J Stat 2000;42:205–223.
7. Rice JA, Wu CO: Nonparametric mixed effects models for unequally sampled noisy curves. Biometrics 2001;57:253–259.
8. Mazze RS, Strock E, Borgman S, Wesley D, Stout P, Racchini J: Evaluating the accuracy, reliability, and clinical applicability of continuous glucose monitoring (CGM): is CGM ready for real time? Diabetes Technol Ther 2009;11:11–18.
9. Nathan DM, Kuenen J, Borg R, Zheng H, Schoenfeld D, Heine RJ: Translating the A1C assay into estimated average glucose values. Diabetes Care 2008;31:1473–1478.
10. De Boor C: Package for calculating with B-splines. SIAM J Numerical Anal 1977;14:441–472.
11. De Boor C: A Practical Guide to Splines. Berlin: Springer, 1978.
12. Grenander U, Szegö G: Toeplitz Forms and Their Applications. Berkeley and Los Angeles: University of California Press, 1958.
13. Widom H: Toeplitz matrices. In: Hirschmann II Jr, ed. MAA Studies in Mathematics: Studies in Real and Complex Analysis. Englewood Cliffs, NJ: Prentice-Hall, 1965:179–209.
14. Böttcher A, Silbermann B: Introduction to Large Truncated Toeplitz Matrices. New York: Springer, 1999.
15. Laird NM, Ware JH: Random-effects models for longitudinal data. Biometrics 1982;38:963–974.
16. Raudenbush SW, Bryk AS: Hierarchical Linear Models: Applications and Data Analysis Methods. Thousand Oaks, CA: Sage, 2002.
17. Singer JD: Using SAS PROC MIXED to fit multilevel models, hierarchical models, and individual growth models. J Educ Behav Stat 1998;24:323–355.
18. Hotelling H: The generalization of Student's ratio. Ann Math Stat 1931;2:360–378.
19. Green PJ: Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. Biometrika 1995;82:711–732.

Address correspondence to:
Hui Zheng, Ph.D.
Biostatistics Center
Massachusetts General Hospital
5th Floor, Suite 560
50 Staniford Street
Boston, MA 02446

E-mail: hzheng1@partners.org

Appendix A

De Boor algorithm and the explicit form for B-spline basis functions

The 0th degree B-spline basis functions are $B_{i,0}(t) = 1$ if $\kappa_i \leq t \leq \kappa_{i+1}$, and $B_{i,0}(t) = 0$ if $t < \kappa_i$ or $t > \kappa_{i+1}$. The d-th degree basis functions are obtained by the recursive relation

$$B_{i,d}(t) = \frac{(t - \kappa_i)}{(\kappa_{i+d} - \kappa_i)} B_{i,d-1}(t) + \frac{(\kappa_{i+d+1} - t)}{(\kappa_{i+d+1} - \kappa_{i+1})} B_{i+1,d-1}(t)$$

Based on 17 knots positioned at $\kappa_i = -4, -2, 0, \ldots, 24, 26, 28$, the 14 quadratic B-spline basis functions take the following form:

$$B_{i,2}(t) = \begin{cases} 0 & \text{if } t > \kappa_{i+3} \text{ or } t < \kappa_i \\ \frac{(t - \kappa_i)^2}{8} & \text{if } \kappa_i \leq t < \kappa_{i+1} \\ \frac{(\kappa_{i+2} - t)(t - \kappa_i) + (t - \kappa_{i+1})(\kappa_{i+3} - t)}{8} & \text{if } \kappa_{i+1} \leq t < \kappa_{i+2} \\ \frac{(\kappa_{i+3} - t)^2}{8} & \text{if } \kappa_{i+2} \leq t \leq \kappa_{i+3} \end{cases}$$

Appendix B

Sample SAS code for fitting the three-level spline model

```
/*
SAS PROGRAM FOR A THREE-LEVEL B-SPLINE MODEL
HUI ZHENG, PHD
MAY 2010


DATA1 CONTAINS VARIABLES ID, PLASMABG, DAY, AND B-SPLINE FUNCTIONS B1--B14.
SMOOTHED GLUCOSE CURVES AND 95% CI ARE OUTPUT INTO ''OUTP1''.
GROUP-WIDE CURVE AND 95% CI ARE OUTPUT INTO ''OUTPM1''.

*/
PROC MIXED DATA=DATA1 METHOD=MIVQUE0;
CLASS ID DAY;
MODEL PLASMABG=B1-B14/NOINT SOLUTION DDFM=BETWITHIN OUTP=OUTP1 OUTPM=OUTPM1;
RANDOM B1-B14/ SUBJECT=ID G TYPE=UN;
RANDOM B1-B14/ SUBJECT=DAY(ID) G TYPE=TOEP(3);
RUN;
```