University of Cincinnati

Date: 6/11/2019

I, Resmi Gupta, hereby submit this original work as part of the requirements for the degree of Doctor of Philosophy in Biostatistics (Environmental Health).

It is entitled:

Flexible Multivariate Joint Model of Longitudinal Intensity and Binary Process for Medical Monitoring of Frequently Collected Data

Student's name: Resmi Gupta

This work and its defense approved by:

Committee chair: Roman Jandarov, Ph.D.

Committee member: Mekibib Altaye, Ph.D.

Committee member: Jane Khoury, Ph.D.

Committee member: Rhonda Szczesniak, Ph.D.

33129

Flexible Multivariate Joint Model of Longitudinal Intensity and Binary Process for Medical Monitoring of Frequently Collected Data

A dissertation submitted to the
Graduate School
of the University of Cincinnati
in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy

in the Division of Biostatistics and Bioinformatics
of the Department of Environmental Health
of the College of Medicine

by

Resmi Gupta

June 2019

MS (Biostatistics), State University of New York at Buffalo
MA (Applied Economics), University of Waterloo, Canada
MSc (Economics), University of Calcutta, India
B.Sc. (Hons. Economics), University of Calcutta, India

Committee:	Chair:	Roman Jandarov, PhD
	Research Mentor and Advisor:	Rhonda D Szczesniak, PhD
	Members:	Jane C Khoury, PhD
		Mekibib Altaye, PhD

Abstract

Objective: Medical monitoring plays a significant role in modern clinical practice and has become a powerful disease management tool that can inform treatment decisions for chronic diseases and disorders, such as diabetes, yielding improved quality of life and longevity in the affected patient. Even though advances in technology have the capacity to promote improved glycemic management strategies during pregnancy, these improvements have not been matched by statistical analysis and implementation science needed to aid the healthcare provider in interpreting the massive amounts of information available from glucose recordings. The overall objective of this dissertation is three-fold: (i) to develop novel statistical methodology, specifically a semiparametric mixed model, to assess the precise timing and degree of rapid fluctuations in the glycemic profiles of mothers with type 1 diabetes mellitus (T1DM), and to determine the extent to which these specific fluctuations are associated with adverse birth outcome (large for gestational age infant); (ii) to characterize the relationship between the maternal glycemic profile and risk of preterm birth (PB) (delivery prior to 37 completed weeks of gestation), using a joint model (JM) approach in a shared random effects framework; (iii) to develop a multivariate joint model (MVJM) within the Bayesian framework to predict adverse neonatal outcome, PB, and characterize its association with monitored glucose trajectories that are subject to irregular intensity of glucose sampling (IOGS) resulting in missing data.

Findings: (i) Using a novel semiparametric statistical approach, we have demonstrated that time-specific fluctuations in velocity and acceleration/deceleration (or change in velocity) of blood glucose (BG) levels differed across gestational age and between women delivering large for gestational age (LGA) and appropriate for gestational age (AGA) infants. These findings were not

detected in the conventional summary measures of glucose variability; (ii) we described a JM approach aimed at analyzing long sequences of longitudinal and event outcomes, and used it to simultaneously characterize the glycemic profile for women with type 1 diabetes and assess risk of PB. When comparing the JM approach to the commonly used time dependent cox model (TDCM), we found that the former was a better fit based on the model fit statistics and characterized association between risk of PB and glucose measurements over the entire pregnancy period to provide additional insights compared to TDCM ($DIC_{JM} = 19895$, $DIC_{TDCM} = 19932$); (iii) our proposed MVJM provided better model fit and higher predictive ability compared to a JM with multivariate linear mixed effects (LME) sub-model (ignoring (IOGS)) and a two stage (TS) model ($AUC_{MVJM} = 0.65$, $AUC_{JM} = 0.56$, $AUC_{TS} = 0.59$). Higher IOGS led to higher chances of PB ($OR_{MVJM(IOGS)} = 2.26$, 95% CI: (1.16, 3.12)). Additionally, presence of preeclampsia related to higher risk of PB; and, women with later diagnosis of T1DM had a lower chance of PB.

Acknowledgements

I would like to thank Dr. Rhonda Szczesniak for guiding me throughout the PhD process and serving as a research and career mentor, and, advisor. She has shown me by example what a successful Biostatistician and researcher is, and I am very thankful for her encouragement, motivation, and time. I want to thank Dr. Roman Jandarov for his insightful comments, support, and, serving as my dissertation chair. I am also very thankful to Dr. Jane Khoury for her guidance, support, and mentorship, and inspiring me to develop myself as a researcher in the best possible way. Thanks also to Dr. Mekibib Altaye for his guidance, support, and serving on my dissertation committee. I am grateful to my colleagues in the Division of Biostatistics and Epidemiology at Cincinnati Children's Hospital for their help and support.

The research included in this dissertation was supported by:

U.S.P.H.S. M01 RR 08084; General Clinical Research Centers Program, National Center for Research Resources, NIH, and HD 11725 Diabetes in Pregnancy Program Project Grant, NIH/NICHD, NIH NCATS Grant 8UL1TR000077-04 for Clinical and Translational Science.

R01 DK109956; Level and timing of diabetic hyperglycemia in utero: the transgenerational effect on adult morbidity (TEAM Study).

Preface

This dissertation includes the work of several multi-author papers for which input came from several people. I am the first author in these papers. In the following, I will detail my contributions for each of these papers. Chapter 2 of this dissertation is partially based on a paper describing a semiparametric statistical approach to identify sensitive time points during gestation for glycemic excursions in type 1 diabetes in pregnancy. My contributions in this work include designing the details and development of the statistical modeling. Chapter 3 is partially based on a joint modelling approach in characterizing maternal glycemic profiles and risk of preterm birth in type 1 diabetes. My contributions in this work include the design and development of the statistical methodology. Chapter 4 is partially based on developing a Bayesian multivariate joint model for medical monitoring data. My contributions here include the design and development of details, and, implementation of our inferential approach. Finally, Chapter 5 summarizes the three papers.

Dedication

This work is dedicated to my grandparents, my mother, and my husband. I will always be grateful for your endless support, and the sacrifices you have made to help me succeed.

Abbreviation table

AGA: Appropriate for gestational age

AUC: Area under the receiver operating system curve

BG: Blood glucose

BIC: Bayesian information criteria

BMI: Body mass index

CI: Confidence interval

COV: Coefficient of variation

DIC: Deviance information criteria

GEE: Generalized estimating equation

IOGS: Intensity of glucose sampling

IQR: Interquartile range

IRB: Institutional review board

JM: Joint model

LGA: Large for gestational age

LME: Linear mixed effects

LMP: Last menstrual period

MAGE: mean amplitude of glycemic excursion

MCMC: Markov chain Monte Carlo

MVJM: Multivariate joint model

OR: Odds ratio

PB: Preterm birth

SD: Standard deviation

T1DM: Type 1 diabetes mellitus

TDCM: Time dependent Cox model

TS: Two-stage

Contents

Abstract.....	i-ii
Acknowledgements	iv
Preface.....	v
Dedication	vi
Abbreviation Table	vii-viii
1. Introduction.....	1
1.1 Background	2-3
1.2 Overall Objective	3
1.3 Specific Aims and Hypotheses	3-5
1.4 Significance.....	5
1.5 IRB Approval and Research Ethics	6
2. A Semiparametric Statistical Approach to Identify Time-specific Detection in Glycemic Profiles of Type 1 Diabetes in Pregnancy	7
Abstract	8
2.1 Introduction.....	9-11
2.2 Methods.....	11
2.2.1 Participant Characteristics and Study Variables	11-12
2.2.2 Model Set Up and Notation	13-15

2.3 Results.....	15-17
2.4 Discussion.....	18-20
2.5 Conclusions.....	20
Figure 2.1	21
Figure 2.2	22
Figure 2.3	23
Figure 2.4	24
Table 2.1	25
Appendix A. SAS Implementation	26
Table 2.A1.....	26
SAS Code.....	26
3. A Joint Modeling Approach in Characterizing Maternal Glycemic Profiles and Risk of Preterm Birth in Type 1 Diabetes.....	27
Abstract	28
3.1 Introduction.....	29-31
3.2 Methods.....	31
3.2.1 Study Design and Participant Characteristic.....	31-32

3.2.2 Statistical Methods.....	33-35
3.2.3 Individual Dynamic Prediction.....	35
3.3 Results.....	36-37
3.4 Discussion and Conclusion	37-40
Tables and Figures	41
Table 3.1	41
Table 3.2	42
Table 3.3	42
Figure 3.1	43
Figure 3.2	43
4. A Bayesian Multivariate Joint Model for Medical Monitoring Data.....	44
Abstract	45
4.1 Introduction.....	46-49
4.2 MVJM Formulation for Mixed Outcomes.....	49-51
4.3 Estimation	51-52
4.4 Application and Results	52
4.4.1 Cohort Data and Study Variables	52-53
4.4.2 MVJM Application	53-57

4.5 Comparison with Other Models.....	57-58
4.6 Simulation.....	58
4.6.1 Design and Analysis	58-60
4.6.2 Simulation Results	60
4.7 Discussion and Conclusion.....	60-64
Tables and Figures	65
Table 4.1	65
Table 4.2	66-67
Table 4.3	68
Figure 4.1	69
Figure 4.2	70
Figure 4.3	71
5. Conclusion.....	72
5.1 Summary.....	73
5.2 Discussion.....	73-75
5.3 Future Research	75
6. References	76-86

Chapter 1

Introduction

1.1 Background

Medical monitoring plays a significant role in modern clinical practice and has become a powerful disease management tool that can inform treatment decisions for chronic diseases and disorders, such as diabetes, yielding improved quality of life and longevity in affected patients^{1,2}. This monitoring may include collection of static variables on individuals, such as certain demographics, and a longitudinal sequence of measurements to act as a surrogate for disease state. This sequence often yields a complex, and usually unbalanced data structure. Advancements in technology and clinical studies have improved diabetes self-management practices, allowing monitoring of BG for example for those with diabetes, on a continuous basis³⁻⁶. In clinical care, patient BG levels may be monitored through self-report data, where the patient records glucose at pre-specified times of the day as directed by the provider or by electronic data capture of this data, the latter being most common in more recent years. Data obtained from patient-based monitoring provides insights into patterns and can be used to assess trends in BG levels over time. These data have been used as a tool to predict impending glucose excursions (hypo- and hyperglycemia), and to assess glycemic variability^{7,8}. This type of self-monitoring data typically involves unequally-spaced glucose measurements, since patients tend to monitor glucose at different times throughout a 24-hour period. The number of glucose measurements varies within and between subjects for various reasons. These types of unbalanced and irregular data may be related to the underlying missing data mechanism, which may be informative. In such instances, irregular or intermittent missingness in BG measurement patterns may be embedded within the subject's glycemic monitoring profile, which during pregnancy for instance, may be correlated with adverse neonatal outcomes. Furthermore, the following critical areas remain largely unexplored:

- (i) No previous studies have examined the complex form of time and subject-specific

rates of change within the glucose profiles and its association with adverse birth outcome, such as delivery of an LGA baby, in women with T1DM.

(ii) No studies have focused on gaining better understanding by simultaneously modeling longitudinal measurements, such as maternal glycemic profiles, and the event outcomes to gain a better understanding of the risk of an event such as PB.

(iii) No studies have been conducted with an aim to improve the predictive ability of a modeling technique by simultaneously modeling the longitudinal (BG measurements), count (IOGS), and, binary outcome (PB) processes, where the longitudinal processes involve unbalanced and irregular data structures.

This dissertation is a compilation of research work that I have led as a Biostatistics PhD student to address the above gaps. While I have published one manuscript (Aim 1: Chapter 2) as first author, the other two are in draft manuscript forms. Chapters 2-4 of this dissertation present the three research activities mentioned above. These manuscripts involve the work that was conducted in partial fulfillment of the requirements for a PhD in Biostatistics in the Department of Environmental Health at the University of Cincinnati College Of Medicine.

1.2 Overall Objectives

The overarching aim of this dissertation is to improve predictive accuracy from simultaneously modeling longitudinal and event processes, where the former involves unbalanced and irregular repeated measurements.

1.3 Specific Aims and Hypotheses

Aim 1. (Manuscript 1):

A 1.1: To develop a semiparametric mixed model to assess the precise timing and degree

of rapid fluctuations of BG levels over gestation in the glycemic profiles of mothers with T1DM.

H_{1.1}: Velocity and acceleration in glucose levels will vary across gestation in women with T1DM.

A_{1.2}: To determine the extent to which these specific fluctuations are associated with delivery of an LGA or AGA baby.

H_{1.2}: Time-specific fluctuations in velocity and acceleration/deceleration of BG levels will differ between women delivering LGA and AGA infants.

A_{1.3}: To contrast this novel approach with conventional glucose variability calculations.

H_{1.3}: The first and second order derivatives of semiparametric mixed regression will yield additional information about glucose fluctuations compared to conventional approaches.

Aim 2. (Manuscript 2):

A_{2.1}: To characterize the association between maternal glycemic profiles and PB using a JM approach.

H_{2.1}: The proposed JM approach will provide additional insights compared to using a conventional approach.

H_{2.2}: Model fit will be improved using a JM technique in comparison to using a conventional approach.

Aim 3. (Manuscript 3):

A_{3.1}: To develop a MVJM in a mixed data type outcomes framework to minimize irregular sampling bias in medical monitoring studies.

H_{3.1}: Our proposed MVJM model will provide higher prediction compared to common approaches employed in biomedical studies.

A3.2: To apply the MVJM for BG monitoring related to adverse perinatal outcome.

H3.2: Mothers who deliver premature babies will exhibit different sampling intensity profiles than mothers who had full-term delivery.

1.4 Significance

Our research methods will accommodate the analysis of longitudinal data when the observation times are possibly outcome dependent. To date, JM prediction involving nonlinear and unbalanced longitudinal trends, such as those exhibited by BG profiles, remains poorly understood and has been recently identified as an area in need of research^{9,10}. Our contribution will be the techniques to develop an enhanced prediction model tailored to the dynamic, nonlinear progression of BG throughout pregnancy. This contribution is significant because it is the next step in a continuum of research to improve prenatal care management and monitoring in T1DM pregnancies. We expect what we learn will have direct impact on BG monitoring during pregnancy and implications for other women's health conditions that require medical monitoring and intensive care management, providing a new paradigm to link perinatal risk in real-time to disease-specific markers. Our proposed model may be used for modeling association between intermittent missing patterns in sampling glucose and other perinatal disease specific risks, such as development of microvascular disease, or for women with diabetes who suffer miscarriage. Finally, the above-mentioned research studies could be directly used to model monitoring data from other diseases and disorders which requires active surveillance, such as blood pressure monitoring and risk of developing coronary artery disease.

1.5 IRB Approval and Research Ethics

The Diabetes in Pregnancy program was first approved by the IRB at the University of Cincinnati. Informed consent was obtained for each pregnancy. We performed secondary analysis of the data available from this T1DM diabetes women cohort, which has been approved by the Cincinnati Children's Hospital Medical Center IRB. The database is de-identified with respect to patient name, identification number, zip code, and date of birth. The data were stored in a shared drive which were restricted solely to the purposes of this research.

Chapter 2

A Semiparametric Statistical Approach to Identify Time-specific Detection in Glycemic Profiles of Type 1 Diabetes in Pregnancy

Glycemic Excursions in Type 1 Diabetes in Pregnancy: A Semiparametric Statistical Approach to Identify Sensitive Time Points During Gestation

Resmi Gupta, Jane Khoury, Mekibib Altaye, Lawrence Dolan, Rhonda D. Szczesniak

Journal of Diabetes Research, Volume 2017, Article ID 2852913,
<https://doi.org/10.1155/2017/2852913>

Abstract

Aim: To examine the gestational glycemic profile and identify specific times during pregnancy that changes in glucose levels, measured by velocity and acceleration of BG fluctuations, are associated with delivery of LGA baby, in women with T1DM.

Methods: Retrospective analysis of capillary BG levels measured multiple times daily throughout gestation in women with T1DM was performed. Semiparametric mixed modeling was used to evaluate the time-specific rates of change throughout gestation while appropriately characterizing progression and variation in glucose levels.

Results: Velocity and acceleration in glucose levels varied across gestation regardless of delivery outcome. Compared to women delivering LGA babies, those delivering AGA babies exhibited significantly lower rates of change and less variation in glucose levels between 180 days gestation and birth.

Conclusions: Use of innovative statistical methods enabled detection of gestational intervals in which BG fluctuation parameters might influence the likelihood of delivering a LGA baby in mothers with T1DM. Understanding the dynamics of gestational changes in BG is a potentially useful tool to assist care providers in determining the optimal timing to initiate continuous glucose monitoring.

2.1 Introduction

Pregnancy in women with T1DM is associated with an increased risk of various adverse outcomes—both for mothers and their offspring. These morbidities include preeclampsia, pre-term delivery, cesarean section delivery, and LGA infant^{11,12}. Long-term adverse health outcomes in the offspring have also been reported such as obesity, insulin resistance, beta cell dysfunction, type 2 diabetes mellitus and cardiovascular dysfunction¹³⁻¹⁵. Management of pregnant women with diabetes focuses on the importance of good glycemic control before and during pregnancy, to decrease the frequency of adverse outcomes for both infants and mothers^{16,17}. Tracking maternal glycemic control over the entire pregnancy has provided insights into the effects of maternal glucose control on various neonatal outcomes, including LGA babies¹⁸. LGA babies born to mothers with diabetes are likely to be significantly overweight both in childhood and as adults, putting them at risk of developing type 2 diabetes¹⁹.

In recent years, advancement in technology has improved diabetes self-management practices allowing monitoring of BG on both a programmed and continuous basis. The sheer volume of data produced and the variability that these data contain has proved challenging to analyze and interpret. Regulation of glucose is a dynamic process that varies in response to meals as well as gestational period. The maternal glucose profile may be considered as measurements of a continuous-time monitoring process whose outputs are samples of functions or curves. Each curve alludes to BG oscillations that occur throughout the day, and perhaps at the same time on different days throughout pregnancy. In recent studies, glucose variabilities have been attributed to both hyperglycemic spikes and hypoglycemic troughs^{20,21}. Both have been attributed to a wider range of chronic and severe complications associated with diabetes mellitus, and hence can provide

crucial information about risk factors for the development of complications in diabetes. Different methods, such as, calculating the SD, COV, and MAGE are popular approaches to measure and monitor fluctuations or glycemic excursions in BG levels²². These approaches, although computationally straightforward, have drawbacks in investigating highly nonlinear profiles^{23,24}. For example, SD is sensitive to outliers and may overestimate the magnitude of a glycemic excursion. Using COV, which is computed based on SD, or using MAGE, which only takes peaks and nadirs at greater than one SD, can also lead to misleading estimation of glycemic excursions.

Vandyke et al.²⁵ used semiparametric mixed models to assess time specific differences in the glycemic profiles of mothers with T1DM who delivered LGA and AGA babies. LGA and AGA profiles had greatest distinctions in the first and third trimesters of pregnancy. The study showed that there is a difference in slopes between glycemic profiles arising from LGA and AGA births, suggesting that rate of change (i.e., velocity) and acceleration (or deceleration) of BG levels vary throughout pregnancy. Even though the above study estimated the mean curvature in the entire glycemic profile, which avoided the typical loss of information with using summary measures, it did not examine the complex form of time and subject-specific rates of change within the glucose profiles.

Analysis of the first- and second-order rates of change is a way to evaluate the dynamics of BG fluctuations on the gestational time scale. First-order rates of change, which are obtained by taking the derivative of estimated glucose trends over pregnancy, correspond to the velocity of glycemic fluctuations. Large rates of change imply rapid BG fluctuations and hence a less stable system. Second-order rates of change indicate how rapidly these fluctuations occur, estimating them in terms of acceleration or deceleration. The acceleration of glucose, which is computed as the derivative of the rate of glucose, can be described as a change in glucose velocity (rate of change)

over time. Acceleration can be termed as a measure of variability, which indicates the change in curvature of maternal glycemic profiles across gestation. Identifying first- and second-order glucose rates of change with respect to gestational time will be helpful in allowing treatment to be targeted to precise points during the pregnancy.

We hypothesize that estimating glycemic variability in terms of derivatives will provide better insight into hypoglycemia and hyperglycemia, which have been shown to be associated with various adverse outcomes and chronic diabetes complications²⁶. Measures of oxidative stress, such as reactive oxygen species, which have been associated with hyperglycemia, are closely related to glucose peaks and fluctuating high and low glucose concentrations²⁷.

Being able to estimate rate of change and acceleration in BG levels across specific gestational periods can provide insight into glycemic fluctuation and risk for hyperglycemia and hypoglycemia for specific gestational time periods on a continuous basis. The purpose of this study is three-fold: 1) to develop a semiparametric mixed model²⁸ to assess the precise timing and degree of rapid fluctuations in the glycemic profiles of mothers with T1DM; 2) to determine the extent to which these specific fluctuations are associated with delivery of an LGA or AGA baby; 3) to contrast this novel approach with conventional glucose variability calculations.

2.2 Methods

2.2.1 Participant Characteristics and Study Variables

The study methods and cohort characteristics have been described in detail elsewhere¹⁶. Briefly, pregnant women or women who were planning a pregnancy and had T1DM were recruited and enrolled in the study. The women were prospectively followed as part of a 17-year interdisciplinary program of diabetes in pregnancy between 1978 and 1995 at the University of Cincinnati Medical

Center. The study subjects were managed with intensive insulin therapy, involving a split mixed-dose regimen of three to four injections per day using short- and intermediate-acting insulin combined with dietary regulation. After 1981, women were instructed to check BG concentrations 6–8 times a day: fasting, pre-prandial (before each meal), 90 min postprandial (after each meal), at bedtime, and occasionally at 3:00 AM. This analysis included women who used a reflectance meter through pregnancy and delivered a singleton live fetus beyond 32 weeks of gestation. The glucose measurements recorded between gestational days 50 and 250 were included in the study. Data from profiles corresponding to neonatal death within 28 days of delivery or presence of a major congenital malformation were excluded from the study. Birthweight was measured within the first hour of delivery using an electronic scale (Toledo Scale, Worthington, Ohio). LGA was defined as birth weight greater than the 90th percentile for gestational age, based on race and gender-specific growth curves. We estimated the rate of change and rate of acceleration (or deceleration) in mother's glycemic profiles with respect to gestational time for LGA and AGA outcomes separately by taking respective first and second-order derivatives of the model equations. A semiparametric mixed effects model using penalized regression splines with a cubic truncated power basis was employed to provide smooth estimates of the longitudinal glycemic profiles^{29,20}. To reflect uncertainty in the estimated rates of change in glucose over pregnancy, 95% pointwise CIs of first order (velocity) and second order (acceleration / deceleration) were constructed using bootstrap sampling. Analyses were implemented in SAS 9.4 (SAS Institute, Cary, NC). Implementation is provided in the online supplement. We considered LGA and AGA profiles to have different trends across gestation and characterized them as described below.

2.2.2 Model Setup and Notation

The observed glucose recordings from the i^{th} pregnancy and j^{th} measurement observed at gestation t_{ij} can be expressed in the semiparametric mixed effects model, separately for LGA and AGA profiles, as:

$$Y_{ij}^{LGA} = \beta_0 + \beta_1 t_{ij} + \beta_2 t_{ij}^2 + \beta_3 t_{ij}^3 + \beta_{02} LGA_i + \beta_{12} LGA_i t_{ij} + \beta_{22} LGA_i t_{ij}^2 + \beta_{32} LGA_i t_{ij}^3 + \sum_{l=1}^L b_l (t_{ij} - K_l)_+^3 + \gamma_{0i} + \gamma_{1i} t_{ij} + \varepsilon_{ij}$$

$$Y_{ij}^{AGA} = \beta_0 + \beta_1 t_{ij} + \beta_2 t_{ij}^2 + \beta_3 t_{ij}^3 + \sum_{l=1}^L b_l (t_{ij} - K_l)_+^3 + \gamma_{0i} + \gamma_{1i} t_{ij} + \varepsilon_{ij}$$

The term LGA_i indicates whether the subject delivered LGA or AGA baby (1=LGA, 0=AGA).

The polynomial terms coefficients, $\beta_0, \beta_1, \beta_2, \beta_3$ are for the respective, intercept, slope, quadratic and cubic terms to model global curvature of the glycemic profile. The polynomial of degree three provided good fit for the data. The summed expression is comprised of the knots at distinct locations K_l ($l=1, \dots, L$) evaluated at gestational time t_{ij} with corresponding coefficients $b_1, \dots, b_L$.

The expression $(t_{ij} - K_l)_+$ is the positive part of the function (ie., $\max\{0, t_{ij} - K_l\}$). To account for variation for both between- and within-subject terms, random intercept (γ_{0i}) and slope (γ_{1i}) terms

were included in the model. In accordance with semiparametric model fitting strategy, the smoothing parameters are incorporated as random effects. We assumed that the random effects are

mutually independent and distributed as $b_l \sim N(0, \sigma_b^2)$, $\gamma_{0i} \sim N(0, \sigma_{\gamma_0}^2)$, $\gamma_{1i} \sim N(0, \sigma_{\gamma_1}^2)$, and measurement error $\varepsilon_{ij} \sim N(0, \sigma_\varepsilon^2)$. For this analysis we considered the same spline coefficients

and same degree of smoothing for both LGA and AGA profiles (i.e., $b_l^{LGA} = b_l^{AGA}$ and

$$Var(b_l^{LGA}) = Var(b_l^{AGA})).$$

Next, we computed the rate of change in glucose levels for women with LGA and AGA births corresponding to the first-order derivative with respect to gestation. More specifically, rate of change, or velocity can be thought of as change in glucose levels over the whole gestational time period, indicating the speed at which glucose levels are changing throughout pregnancy.

The rate of change in glucose levels across gestation for LGA_i can be written as the first derivative of the model:

$$\frac{d(Y_{ij}^{LGA})}{dt_{ij}} = \beta_1 + 2\beta_2 t_{ij} + 3\beta_3 t_{ij}^2 + \beta_{12} LGA_i + 2\beta_{22} LGA_i t_{ij} + 3\beta_{32} LGA_i t_{ij}^2 + 3 \sum_{l=1}^L b_l (t_{ij} - K_l)_+^2$$

Similarly, the rate of change in glucose levels across specific gestational periods for mothers with AGA babies can be expressed as:

$$\frac{d(Y_{ij}^{AGA})}{dt_{ij}} = \beta_1 + 2\beta_2 t_{ij} + 3\beta_3 t_{ij}^2 + 3 \sum_{l=1}^L b_l (t_{ij} - K_l)_+^2$$

The acceleration or the second order derivative with respect to gestation was computed to measure how fast (or slow) the velocity of glucose levels is changing across gestation.

And the acceleration in glucose levels across gestational time for LGA is:

$$\frac{d^2(Y_{ij}^{LGA})}{dt_{ij}^2} = 2\beta_2 + 6\beta_3 t_{ij} + 2\beta_{22} LGA_i + 6\beta_{32} LGA_i t_{ij} + 6 \sum_{l=1}^L b_l (t_{ij} - K_l)_+;$$

the acceleration in glucose levels for mothers with AGA births is:

$$\frac{d^2(Y_{ij}^{AGA})}{dt_{ij}^2} = 2\beta_2 + 6\beta_3 t_{ij} + 6 \sum_{l=1}^L b_l (t_{ij} - K_l)_+$$

Commonly used summary glycemic variability measures, SD, COV, and MAGE¹² [12] were computed for each participant and trimester, then compared with the acceleration (and deceleration) in glucose fluctuations estimated from the semiparametric mixed model. First, SD around a mean glucose level for each of the three trimesters was computed. Next, COV was calculated by dividing the SD by the mean of the corresponding glucose readings. MAGE was computed to measure the average heights of glucose excursions between peaks (highs) and nadirs (lows) that exceed the SD for an individual within a given day. It generates a value for the variation around a mean glucose value by summing the absolute rises or falls encountered. These measures were compared between LGA and AGA groups for each trimester using repeated measure analysis of variance, in order to contrast these longitudinal findings, which are typically used to make conclusions about glycemic fluctuations, with results from the semiparametric mixed modeling.

2.3 Results

The analysis cohort was comprised of a total of 199 women with T1DM with 246 pregnancies and 139,991 glucose readings. Glucose concentration recording times and number sampled varied across women, with each woman having a mean (range) number of recordings of 634 (10-2728). A random sample of five women showed large within-subject and between-subject variations in glucose levels over the gestational period (Figure 2.1). Duration of follow up among women was 155 (62-182) days. Forty three percent of the entire cohort had LGA infants. Maternal age at pregnancy was similar between LGA and AGA groups (26.5 [16-37] versus 27.0 [15-38] years).

For the semiparametric mixed model fitting, a total of 8 knots were selected, ranging from 76.67 to 247.99 gestational days, using the quantile method described by Ngo and Wand²⁹. The fitted maternal glycemic profiles reflected differences in glucose changing across gestational days for

LGA and AGA births (Figure 2.2). The mean glucose concentration for both groups declined, with a higher rate of decrease for LGA group. Across the gestational time, AGA had lower rate of change or velocity in glucose levels compared to LGA births (Figure 2.3). Greater velocity in glucose recordings for both AGA and LGA occurred between 50 and 120 gestational days (AGA: -1.62 mg/dl/day to 0.09 mg/dl/day; LGA: -2.02 mg/dL/day to 0.03 mg/dL/day). The rate of change occurred rapidly between 150 and 250 gestational days, with a more rapid decline in glucose readings for LGA births (-0.21 mg/dL/day to -2.50 mg/dL/day) compared to AGA births (-0.002 mg/dL/day to -1.54 mg/dL/day). The most rapid rate of change in glucose levels occurred between 230 and 250 gestational days for both AGA and LGA with a more rapid decline for the LGA births (AGA: -0.51 to -1.54 mg/dL/day; LGA: -1.20 to -2.50 mg/dL/day). The pointwise 95% confidence band for rate of change (or first order derivative) of glucose readings reflected non-overlapping regions between LGA and AGA groups for the first trimester, suggesting that the velocity of glucose readings were different for AGA and LGA births (Figure 2.3). However, for the second trimester, the confidence band overlapped, reflecting the rate of change in glycemic levels were similar for LGA and AGA groups. During the third trimester the confidence band did not overlap, meaning that the glycemic velocity levels were different, with an increasing rate of change in glucose reading for the LGA compared to the AGA group.

The extent of rapid changes in glucose was quantified using second derivative estimates over gestation (Figure 2.4), which reflected distinct patterns between LGA and AGA profiles. Although both groups' exhibited sharp decelerations between 50 and 120 gestational days, there was less change in the AGA group. Glucose velocity appeared to be increasing for roughly 5 weeks thereafter, and then decreased for three weeks. There is a slight acceleration in glucose for both groups between 175 and 230 gestational days (over a period of 8 weeks) following a sharp decline

in glucose concentration for both groups. With regard to trimesters, acceleration in BG concentration between 120 and 150 gestational days (second trimester) is an indicator of hyperglycemic triggering events. There was a sharp decline in glucose level for both groups during trimester 1 and 3, depicting risk of hypoglycemia associated with AGA and LGA births. As shown by the 95% confidence bands, the deceleration in glucose during trimester 3 was significantly higher for women with LGA compared to AGA births, suggesting higher risk of hypoglycemia for the former group. For rate of acceleration in glucose control, the non-overlapping confidence band during first trimester suggested that mothers with LGA birth outcome had different rate of progression in glucose levels compared to AGA birth. There was a short period during trimester 2 where the 95% pointwise CIs overlapped, suggesting similar rate of progression in glucose readings for both LGA and AGA. From middle of trimester 2 till end of trimester 3, there was a difference in acceleration in glucose readings, with a higher rate of progression in glycemic levels for LGA compared to AGA group. The narrow confidence bands for rate of change and acceleration were due to relatively large number of observations in our study.

The COV, SD, and MAGE estimates by each trimester and group are presented in Table 2.1. For all three measures, there was a decline in glycemic variability stratified by LGA and AGA groups. The comparisons between LGA and AGA groups for each trimester did not yield statistically significant differences for any variability measures suggesting similar glucose concentration between the groups. Thus, it may be that the results from the standard methods are misleading or yield insufficient information, as the findings do not differentiate rates of progression in glucose intake for women delivering an LGA versus AGA infant.

2.4 Discussion

Using a novel statistical approach, we have demonstrated that time-specific fluctuations in velocity and acceleration/deceleration (or change in velocity) of glucose levels differed across gestational age and between women delivering LGA and AGA infants. These findings were not indicated in the conventional summary measures of glucose variability. There was a steeper rate of change (or velocity) in glucose concentration between 150 to 250 gestational days for LGA group compared to AGA group. The measurement of acceleration by the second order derivative with respect to gestation time provided instantaneous measures of glucose fluctuations over the entire pregnancy period. The acceleration in glucose levels displayed sharp decline in the first and third trimester for both LGA and AGA groups. During the first trimester, the acceleration on glucose levels was lower for AGA compared to LGA group, suggesting a potential risk of developing hypoglycemia for the former. The rebounds from hypoglycemia occurred for both groups in the second trimester, with a lower rate for LGA group. After that, there was a decline in glucose concentration for both groups, following a steady state and a sharp trough in glucose levels for LGA, suggesting risk of developing hypoglycemia during the third trimester.

Several studies have demonstrated that glycemic fluctuations is a deterministic factor for hyperglycemic and hypoglycemic excursions. In recent years continuous glucose monitoring systems have emerged as an effective technology with an ability to monitor glucose trends over time. Involving large amount of data, continuous glucose monitoring systems provide information about frequency of fluctuations about BG levels³¹. However, the unbalanced nature of BG data results in large magnitude of within and between patient variabilities. The semiparametric regression approach in our study improved our understanding of velocity and acceleration of maternal glycemic profiles on pregnancy outcomes. The estimation of derivatives provided more

accurate information on periods in which glucose has higher variation, compared to use of summary measures. An interesting feature of acceleration/deceleration in glucose levels was detecting gestation specific recurrence of risk of developing hypoglycemia associated with mother's glycemetic profiles.

Intervening when data indicate glucose levels are outside of the target area of normal ranges to control the glycemetic fluctuations delays several complications associated with diabetes. The acceleration, which measures the variability in glucose concentration, may provide a timely warning of severe hyperglycemia or hypoglycemia. The estimates could serve as a prognostic aid to monitor glucose across the gestational period, thus increasing patients' and clinicians' abilities to take timely actions in a more accurate manner, and hence improving pregnancy outcomes. The velocity and acceleration in glucose fluctuations may guide the clinicians to detect the time of intervention for continuous glucose monitoring. By precisely identifying specific time points with continuous modeling, the first and second order derivatives of glucose fluctuations will yield additional information that might contribute to developing risk model at different stages of pregnancy. Our approach showed that at the early stage of pregnancy, glucose monitoring will be useful for both LGA and AGA, and that at the end stage of pregnancy it is very important to monitor the glucose levels, especially to minimize the risk of an LGA birth. Such information was not evident from any of the traditional glucose variability indices.

There are some limitations to the current study. Although the original protocol directed multiple daily measurements of glucose, it is difficult for a subject to check glucose at the precise moment of a hypo- or hyperglycemic excursion. Rapid changes in BG levels may not be reflected in these measurements, which often occur at the end of an excursion. Thus, the measurements may not always reflect the precise level of fluctuations and may show lower readings instead. In that case,

the rate of change and acceleration (or deceleration) estimates may involve underestimation of change in glucose fluctuations. This study did not include covariate adjustment. Future work should aim to estimate the effects of fixed and time-varying predictors on the glucose readings to explain the causal inference and improve our understanding of fluctuation in maternal glucose profile on pregnancy outcome. Also, use of multilevel models to analyze cases of multiple pregnancies from the same women could be investigated using semiparametric regression modeling.

2.5 Conclusions

We showed that first and second order derivatives of glycemic levels with respect to gestational time, termed as velocity and acceleration in glucose readings, are useful to explain variability in maternal glycemic profiles with LGA and AGA births and detect gestational periods where they are different. We were able to identify specific gestational periods where rate of progression in maternal glucose concentration differed between LGA and AGA groups. The time-specific detection of velocity and acceleration (or deceleration) in glucose levels can be effective in monitoring risk for hyperglycemic (or hypoglycemic) events. Responding in a timely manner to high- and low-glucose alerts, the clinicians can significantly reduce glucose fluctuations and keeping the glucose readings within target range, thereby improving pregnancy outcome.

Figure 2.1: The glucose readings of 5 randomly selected women throughout pregnancy measured in mg/dl/day. These profiles show large variation both between and within subject during pregnancy.

Figure 2.2: Fitted glucose curves from semiparametric regression modeling using penalized regression splines and cubic truncated power basis function. The solid line represents fitted maternal glycemic profiles of LGA group; the dash line represents fitted maternal glycemic profiles of AGA group.

Figure 2.3: The rate of change (first order derivative with respect to gestation time) in glucose recordings measured by mg/dl/day for women with LGA and AGA births outcomes. The overlapping region in 95% pointwise CI indicates the gestational time range where both LGA and AGA have the same rate of change in glucose readings.

Figure 2.4: The acceleration/deceleration (second order derivative with respect to gestation time) in glucose recording of LGA and AGA groups across gestational time. The overlapping region in 95% pointwise CI indicates similar rate of progression in glycemic levels for both groups.

Table 2.1. Measure of glucose variabilities compared between LGA and AGA babies across trimesters

Measures of glucose variability	AGA EST(SE)	LGA EST(SE)	<i>p</i> -value
SD			
Trimester 1	80.13 (2.46)	74.64 (3.12)	NS
Trimester 2	75.93 (2.29)	70.25 (3.00)	NS
Trimester 3	67.19 (2.29)	62.72 (2.97)	NS
Coefficient of variation			
Trimester 1	55.66 (1.25)	52.36 (1.50)	NS
Trimester 2	52.49 (1.14)	48.35 (1.39)	NS
Trimester 3	49.72 (1.13)	44.14 (1.36)	NS
MAGE			
Trimester 1	175.38 (6.70)	166.89 (9.15)	NS
Trimester 2	165.93 (6.45)	162.65 (8.94)	NS
Trimester 3	158.73 (6.50)	155.31 (8.97)	NS

Note: EST: Estimate, SE: Standard error; MAGE: Mean Amplitude of Glycemic Excursion, Excursions>1 SD from the mean

Appendix A. SAS implementation

Table 2.A1.

Variable	Description
Glucose Cincatwt Id Gestday 11-18	Response variable : Glucose reading Group variable : LGA and AGA Subject identification number Gestational day (time) Columns of Z matrix (cubic) for smoothing (K=8 knots)

A total of 8 knots were selected using the algorithm from Ngo and Wand (2004), ranging from 76.67 to 247.99 gestational time.

```
PROC HP MIXED DATA=finaldata ;  
CLASS id cincatwt;  
MODEL glucose = gestdaytime gestdaytime*gestdaytime  
                gestdaytime*gestdaytime*gestdaytime cincatwt  
                cincatwt*gestdaytime  
                cincatwt*gestdaytime*gestdaytime  
                cincatwt*gestdaytime*gestdaytime*gestdaytime/s;  
  
random intercept gestdaytime/subject=id type=cs;  
random 11-18 / type=toep(1) s ;  
Title 'Semiparametric mixed regression model with cubic polynomial basis';  
run ;
```


Chapter 3

A Joint Modelling Approach in Characterizing Maternal Glycemic Profiles and Risk of Preterm Birth in Type 1 Diabetes

Assessing the Relationship between Gestational Glycemic Control and Risk of Preterm Birth in Women with Type 1 Diabetes: A Joint Modeling Approach

Abstract

Characterizing and estimating rate of maternal glucose sampling over the course of the entire pregnancy is an important task to improve prediction of adverse birth outcome, such as PB, in T1DM. We used a univariate JM approach in a shared-parameter framework involving a longitudinal outcome process (daily BG sampling) and an event process (PB) to characterize the relationship between the two processes. An LME model using natural cubic splines was fitted to predict the longitudinal sub-model. Covariates, selected through one-at-a time modeling, included mother's age at LMP, age at diabetes onset, BMI, white classification, hypertension, retinopathy, and nephropathy. Various association structures (value, value plus slope, AUC) were tested before selecting the final JM. We compared the JM approach to commonly used TDCM. There were 16,480 glucose readings over gestation (range: 50-260 days); 32 women (28%) had PBs. Based on BIC, a total of 22 knots was selected to fit individual glycemic profiles. Mother's age at LMP and age at diabetes onset were statistically significant ($\beta_{age} = 1.29$, 95% CI: 1.10, 1.72, $p=0.02$; $\beta_{ageatdiabetesonset} = 0.84$, 95% CI: 0.62, 0.98, $p<0.01$) for the longitudinal sub-model, reflecting that older women tended to have higher mean BG and those with later diabetes onset tended to have lower mean BG level. Presence of nephropathy was statistically significant in the event sub-model ($RR_{neph} = 2.29$, 95% CI: 1.05, 4.48, $p<0.01$). Cumulative association parameterization provided the best JM fit based on DIC. When comparing the JM approach to commonly used TDCM, the former was a better fit ($DIC_{JM} = 19895$; $DIC_{TDCM} = 19932$) and characterized association between

risk of PB and glucose measurements over the entire pregnancy period to provide additional insights compared to TDCM.

3.1 Introduction

Self-monitoring of BG plays a significant role in reduction of perinatal mortality and morbidity in pregnant women with T1DM^{32,33}. Tracking BG over the entire pregnancy has provided insight into the effects of poor maternal glucose control on various neonatal outcomes, including PB³⁴. Recent studies have shown that the rate of both spontaneous and indicated PB is increased in women with T1DM³⁵. The rate of PB ranged from 22% to 45% in women with T1DM, which is greater by four- to eightfold over the frequency of preterm delivery in pregnancies which are not complicated by diabetes³⁶. Preterm babies have been shown to have a higher risk of morbidity, mortality and poor neurological outcome later in life³⁷. In addition to the overall increased rate, poor glycemic control was associated with both spontaneous and indicated PB, for mothers with T1DM³⁸. Hence, characterizing the glucose profile together with estimating the rate of maternal glucose sampling over the course of the entire pregnancy is an important step along the path to predict the risk of PB outcome and inform T1DM care management practices.

Regulation of glucose is a dynamic process that varies in response to meals, insulin dosage as well as gestational period³⁹. Recently, advancement in technology has improved diabetes self-management practices allowing monitoring of BG on both a programmed and continuous basis. The maternal glucose profile may be considered as measurements of an almost continuous-time monitoring process whose outputs are samples of functions or curves. Each curve accounts for BG oscillations that occur throughout the day, and perhaps at the same time on different days throughout pregnancy. The longitudinal measurements of BG level are subject to measurement

error as most other self-reported and machine recorded data, but measurements may be viewed as a noisy version of the true glucose profile.

A common aim in follow-up studies is to characterize the relationship between longitudinal measurements and the event outcomes to gain a better understanding of the risk of an event such as preterm delivery. The TDCM⁴⁰ is prevalent in the literature for modeling this kind of association. However, TDCM assumes that the covariates are external and, are independent of the event mechanism^{41,42}. Moreover, another strong assumption in TDCM is that the time dependent covariates are measured without error.

Recently, JM has been shown to be an efficient approach to simultaneously model longitudinal data and an observed outcome, often yielding more accurate, informative prediction than traditional models^{43,44}. The JM approach constructs two sub-models, longitudinal and event time, and links the two models by subject specific random effect terms. The early development of JMs for longitudinal and survival data was primarily motivated from HIV/AIDS clinical trials to model longitudinal CD4 counts and survival data. Classical models such as the LME for longitudinal data and the Cox proportional hazards model for time-to-event data do not consider dependencies between these two different data types (longitudinal and time-to-event data). By simultaneously modeling these two data types into a single model, JMs takes into account the dependence and association between the longitudinal and time to event outcomes, in order to better assess the effect of a treatment, quantify uncertainty and to provide accurate predictions of outcomes. Excellent expositions of JMs have been provided by Wulfsohn and Tsiatis⁴⁵, Tsiatis and Davidian⁴⁶, De Gruttola and Tu⁴⁷, Wang and Taylor⁴⁸, Henderson et al⁴⁹, and Brown and Ibrahim et al⁵⁰. More recent work on JMs includes Rizopoulos et al⁵¹, Wu et al⁵², Ye et al⁵³, Huang et al⁵⁴, and Wu et al⁵⁵, among others.

In addition to being able to characterize the association between longitudinal and time-to-event outcomes using JM, dynamic prediction has also gained increasing momentum in clinical research. Because dynamic predictions are individualized and have the ability to appropriately account for possible nonlinearity in each individual longitudinal trajectory, better prospective treatment decisions may become available. In contrast to LME, JM could provide individualized risk prediction based on the availability of longitudinal information up to that point with the assumption that the individual is event-free until that time.

In the current study, we utilize the aforementioned advantages of JM in the shared-parameter framework to characterize the relationship between the maternal glucose profile in the longitudinal sub-model and PB outcome in the event sub-model. Our overall objective is to improve the prediction using the JM approach while characterizing the relationship between maternal glycemic profiles and PB among women with T1DM. We hypothesize that JM would provide better model fit compared to the conventional approach, the TDCM. The final JM model will be chosen based on the model fit statistics for different association structures between the maternal glycemic profiles and risk of PB. Additionally, following Rizopoulos proposed Monte Carlo approach⁵⁶, we will estimate the risk of PB and illustrated how it can be dynamically updated, given that the subject-specific glycemic profiles were available up to the time of prediction.

3.2 Methods

3.2.1 Study Design and Participant Characteristics

The study methods and cohort characteristics have been described in detail elsewhere⁵⁷. Briefly, women with confirmed diagnosis of T1DM who were either pregnant or planning a pregnancy

were recruited and enrolled in our Diabetes in Pregnancy study. The women were prospectively followed over gestation as part of a 17-year interdisciplinary program of diabetes in pregnancy between 1978 and 1995 conducted at the University of Cincinnati Medical Center. The study subjects were managed with intensive insulin therapy, involving a split mixed-dose regimen of three to four injections per day using short- and intermediate-acting insulin combined with dietary regulation. After 1981, women were instructed to check BG concentrations 6–8 times a day: fasting, pre-prandial (before each meal), 90 min postprandial (after each meal), at bedtime, and occasionally at 3:00 AM. The current analysis included women who used a reflectance meter through pregnancy and delivered a singleton live fetus beyond 20 weeks of gestation. The glucose measurements recorded between gestational days 50 and 260 were included in the study. Data from profiles corresponding to neonatal death within 28 days of delivery or presence of a major congenital malformation were excluded from this analysis. Birthweight was measured within the first hour of delivery using an electronic scale (Toledo Scale, Worthington, Ohio). Preterm delivery was defined as delivery prior to 37 weeks' gestation as a result of either spontaneous preterm labor or preterm premature rupture of membranes, or any other cause classified by a study perinatologist as indicated preterm delivery. Glycosylated hemoglobin A1c concentration was measured every four weeks throughout pregnancy and prior to delivery using Isolab column chromatography. Mother's age at LMP, age at diabetes onset, BMI, White classification (widely used to assess maternal and fetal risk complicated by diabetes), chronic hypertension, retinopathy, and nephropathy were included as predictors in the initial longitudinal (glucose trajectory) and event (PB) model. Bivariate analyses indicated that mother's age, age at diabetes onset, and presence of nephropathy were statistically significant in longitudinal and event models and we included only these variables in the JM and TDCM modeling phase.

3.2.2 Statistical Methods

Each JM consists of two linked sub-models, a mixed effects model to fit longitudinal BG monitoring and a survival model to fit PB data. The longitudinal outcome, glucose recordings (log transformed scale) from the i^{th} woman and j^{th} pregnancy observed at gestation time t_{ij} ($i = 1 \dots n, j = 1 \dots n_{ij}$) can be expressed as:

$$(\text{Glucose Recording})_{ij} = X_{ij}^T(\theta) + f(t_{ij}) + Z_{ij}^T b_i + \varepsilon_{ij}$$

Where parametric effects are represented for traditional covariates X_{ij} and $p \times 1$ parameter vector θ ; $f(\cdot)$ is a non-parametric smooth function evaluated at time t_{ij} ; Z_{ij} and b_i correspond to the design matrix and the subject specific $q_i \times 1$ parameter vectors for random effects; ε_{ij} represent the measurement error corresponding to the observation at time t_{ij} . The mean response function $f(t_{ij})$ in equation (1) can be estimated using a natural cubic spline which can be expressed as:

$$f(t_{ij}) = \sum_{r=1}^R B_r \lambda_r(t_{ij}) \text{ where } B_r \text{'s are the parameter coefficients of the expansion with basis}$$

functions $\lambda_r(t_{ij})$. In order to capture the non-linear subject specific fluctuations in glucose recordings, we included a natural cubic spline that expands the time effects into a B-spline basis matrix. Based on the BIC⁵⁸, we selected a total of 22 knots, which was necessary to estimate the individual dynamic prediction for risk of PB (discussed in the next section). Taking into account the vector of fixed-effect covariates, maternal age at LMP, age at diabetes onset, and spline functions, the above equation can be written as:

$$(\text{Glucose Recording})_{ij} = m_i(t) + \varepsilon_i(t) =$$

$$(\beta_0 + b_{i0}) + (\beta_1 + b_{i1})B_r((\text{gestational day})_{ij} + \lambda_1) + (\beta_2 + b_{i2})B_r((\text{gestational day})_{ij} + \lambda_2) + \dots + (\beta_{22} + b_{i22})B_r((\text{gestational day})_{ij} + \lambda_{22}) + \beta_{23}(\text{age})_i + \beta_{24}(\text{age at diabetes onset})_i + \varepsilon_{ij} \quad (1)$$

We assumed that the random effects are mutually independent and distributed as $b_{ik} \sim N_{23}(0, \sigma_{ik})$; $k=0,1,...,22$; and measurement error $\varepsilon_{ij} \sim N(0, \sigma_{\varepsilon}^2)$.

We linked the event model, probability of PB to the glucose measurements through the random effects. Let h_i denote the event endpoint, which is the binary indicator of PB. The probability of PB is linked to the longitudinal process of glucose measurements as:

$$h_i [t] = lt^{l-1} \exp\{\Omega_i^T(\phi) + \alpha m_i(t)\} \quad (2)$$

where Ω_i^T denotes the vector of fixed effects covariates, such as, presence of nephropathy at baseline; $m_i(t)$ is the true values of the underlying glucose monitoring processes, with α being the parameter characterizing the association between longitudinal glucose recording process and the PB outcome. Equation (2) monotonically increases with time if $l > 1$, decreases if $l < 1$, and the exponential hazard remains constant if $l = 1$.

In equation (2), the risk of PB depends on the current value of BG level. However, since for each patient the BG follows a trajectory in time, it is also reasonable to consider a JM in which the risk depends on both the current true value of the BG trajectory and the slope of the true BG trajectory at that time. The event model then becomes

$$h_i [t] = lt^{l-1} \exp\{\Omega_i^T(\phi) + \alpha_1 m_i(t) + \alpha_2 m_i(t')\} \quad (3)$$

where $m_i(t') = \frac{d}{dt}(m_i(t))$. Additionally, we fitted the association structure to allow the whole history of the BG trajectory to be associated with the risk of the PB. Then the event model takes the form

$$h_i[t] = lt^{l-1} \exp\{\Omega_i^T(\phi) + \alpha \int_0^t m_i(s) ds\} \quad (4)$$

For any particular time point, α in (4) measures the strength of association between the risk of PB at time t and the area under the longitudinal trajectory of BG level up to the same t . We tested all three association structures (value, slope, and AUC) and selected the best model based on the DIC⁵⁹. The DIC measure balances the fit of a model to the data with its complexity. A smaller value of DIC indicates the preferred model.

We compared the JM with the TDCM where we fitted the PB event model with a cox model that included glucose measurements as a regular time dependent covariate (ignoring measurement error), i.e.

$$h_i[t] = lt^{l-1} \exp\{\gamma_1(\text{glucose})_{ij} + \gamma_2(\text{gestational day})_{ij} + \gamma_3(\text{glucose})_{ij} * (\text{gestational day})_{ij} + \gamma_4(\text{age})_i + \gamma_5(\text{age at diabetes onset})_i + \gamma_6(\text{presence of nephropathy})_i\} \quad (5)$$

Model comparison between TDCM and JM was conducted by DIC. All data analyses were conducted using SAS v9.4.4 (Cary, NC) and *JMbayes* package in RStudio (2015).

3.2.3 Individual Dynamic Prediction

In this paper, we used Rizopoulos proposed approach⁵⁶ to dynamically update the risk of PB based on the updated glucose trajectory for an individual subject. A key feature of these dynamic prediction frameworks is that the predictive measures for risk of PB can be dynamically updated as additional longitudinal measurements of BG level become available for the target subjects, providing instantaneous risk assessment.

3.3 Results

Descriptive statistics are presented in Table 3.1. A total of 114 women with T1DM with 16,480 glucose readings were included in this analysis. Median (IQR) number of glucose recordings was 157 (6-211); 32 (28%) of the women had preterm delivery. Median age of entry was 27 (22-31) years; median age of diabetes onset was 12 (8-17) years; median baseline BMI of the cohort was 23 (21-25) kg/m². A total of 32 (28%) were diagnosed with preeclampsia, 14 (12%) with chronic hypertension, 40 (35%) with white classification (B/C vs ≥D), 22(19%) with nephropathy, and, 16 (14%) with retinopathy.

In order to allow for flexibility in the nonlinear BG profiles, we included a natural cubic splines in both the fixed and random effects part of the mixed effects model. A range of knots was used to fit the longitudinal BG profiles. On the basis of BIC, a total of 22 knots was selected to fit individual profiles. Figure 3.1 corresponds to the overall model fit for the glycemic profiles and 95% CI for the period of 50-260 gestational days. Mother's age at LMP ($\beta_{age} = 1.29, 95\% \text{ CI}: 1.10, 1.72, p=0.02$) and age at diabetes onset ($\beta_{age \text{ at diabetes onset}} = 0.84, 95\% \text{ CI}: 0.62, 0.98, p<0.01$) were statistically significant for the longitudinal sub-model, reflecting that older women tended to have higher mean BG level, and those with later diabetes onset tended to have lower mean BG level (Table 3.2). Presence of nephropathy was statistically significant for the event sub-model, reflecting that, probability of PB significantly increases with presence of nephropathy among women with T1DM ($RR_{neph} = 2.29, 95\% \text{ CI}: 1.05, 4.48, p<0.01$). Based on DIC, JM provided better model fit in comparison to TDCM ($DIC_{JM} = 19895; DIC_{TDCM} = 19932$). Unlike JM, mother's age, and age at diabetes onset were not significant in the TDCM. Presence of nephropathy was statistically significant in TDCM ($RR_{neph} = 2.21, 95\% \text{ CI}: 1.01, 3.32, p=0.04$).

Based on DIC, the association link with cumulative glucose monitoring provided the best fit

($DIC_{\alpha_{Value}} = 19990$; $DIC_{\alpha_{Value \text{ plus slope}}} = 19905$; $DIC_{\alpha_{AUC}} = 19895$) compared to value, or rate of glucose level at time t (Table 3.3). The cumulative association (AUC) indicated that the cumulative effects of the glucose monitoring outcome up to time point t had the strongest association ($\alpha_{AUC} = -0.01$, $p < 0.01$) between the risk of PB and maternal glycemic profile.

The assumption for dynamic prediction is that the woman was event free (i.e., did not give birth) up to the time point of the last glucose reading available. Using the MCMC algorithm, the estimates of the risk of PB were computed for a new subject knowing her glucose readings up to a given gestation day. Figure 3.2 depicts the change in the risk of PB of a woman with increasing number of glucose readings over time. For glucose readings up to 107 gestational days, the woman had roughly a 25% risk of having premature delivery. The risk decreases by 5% with her available glucose readings out to 212 gestational days in pregnancy.

3.4 Discussion and Conclusion

In this study, we demonstrated that characterizing and estimating rate of maternal glucose sampling profile over the course of the entire pregnancy is an important task to predict PB outcome. Recent studies have shown that the rate of both spontaneous and indicated PB is increased in women with T1DM^{60,61}. Preterm babies go on to have a higher risk of increased morbidity, mortality and poor neurological outcome⁶². In mothers with T1DM, poor glycemic control was associated with both spontaneous and indicated PB⁶³. As we have shown in our previously published work³⁹, nonlinear curvature of BG monitoring data makes it analytically challenging. In recent years, JM has been shown to be an efficient approach to simultaneously

model longitudinal data and an observed outcome, often yielding more accurate, informative prediction than traditional models⁴³. In this paper, we describe a JM approach aimed at analyzing long sequences of longitudinal and time-to-event data and used it to simultaneously characterize the glycemic profile for women with T1DM and assess risk of PB. On the population level, we identified demographic characteristics associated with longitudinal and event outcomes. When comparing the JM approach to commonly used TDCM, we found that the former was a better fit based on the model fit statistics and characterized association between risk of PB and glucose measurements over the entire pregnancy period to provide additional insights compared to TDCM. While maternal age at LMP and age at diabetes onset were not significant in the TDCM, it was significant in the JM providing additional information regarding the relationship between maternal glycemic profile over the course of entire pregnancy and mother's age at LMP, and, age at diabetes onset. We found that older women tended to have higher mean BG, and those with later diabetes onset tended to have lower mean BG. Presence of nephropathy was statistically significant in both JM and TDCM.

We tested various association structures to share the parameters between the longitudinal and event sub-model within the JM framework. Based on DIC, the association structure based on AUC of the glycemic profile provided the best fit for the model in comparison to the value, and, value and rate association structures. The cumulative association structure, as defined by the AUC, is clinically more appealing as it includes the cumulative effects of the glucose monitoring outcome up to a certain time point to determine the risk of PB of a subject. Further, our approach, which includes a LME sub-model, accounts for missing glucose data. This is not possible in summary calculations, which are prone to measurement error⁶⁴.

A novel use of JMs, leveraged in this study, is to obtain dynamic personalized prediction of future longitudinal outcome trajectories and risks of an event at any time, given the subject-specific outcome profiles up to the time of prediction. Rizopoulos proposed a Monte Carlo approach²⁵ to estimate risk of a target event and illustrated how it can be dynamically updated. We followed this approach to show how dynamic prediction for risk of PB can be obtained as a function of maternal glycemic profile over the course of an entire pregnancy for a woman in this T1DM cohort. Based on the fitted JM, we derived the prediction for a new subject from the same population that provided a set of longitudinal glucose measurements up to a specified time. As an example we considered one maternal glycemic profile, and based on a Monte Carlo procedure, showed the mean and median estimates over the Monte Carlo samples along with the 95% pointwise credible intervals. We calculated the conditional full term birth (i.e. event-free) probabilities for an individual patient, starting from her first glucose measurements and adding glucose at each time point as an extra measurement.

There are some limitations to the current study. We conducted variable selection based on a one-at-a time model selection process. Although this approach is suitable for prediction modeling, it cannot be used to infer causality; therefore, association estimates from this study must be viewed cautiously⁶⁵. With respect to missing data in maternal glycemic profiles, our approach assumes that data are missing at random⁶⁶ in the sense that missingness can be explained by observed information. By using the existing R package, *JMbayes*, we modeled observed birth outcome as a survival process, with no censoring in the data. Our further research aims to rectify this problem by including the sub-model of PB as a binary process. The JM approach proposed in this study may be used to model continuous glucose monitoring data, thus incorporating the cutting-edge technology in diabetes care. Moreover, future work should aim to separately model the irregular

glucose sampling pattern associated with the BG level for women with T1DM to predict risk of PB.

In conclusion, we have developed a JM approach that, once applied to motivating data, provided more realistic estimates of the maternal glycemic profile over the entire pregnancy and enabled individualized assessment of PB risk in women with T1DM. Through a Bayesian MCMC algorithm implemented in open-source software, this model could be used to aid clinicians in estimating risk of preterm delivery from early pregnancy than later based on the glucose trajectory and other baseline covariates, potentially allowing for intervention strategies to be applied. We demonstrated that overall model fit improved with our JM, compared to the conventional TDCM. Some important covariates, mother's age at LMP, age of diabetes onset were significant in the JM unlike using the conventional approach, TDCM. Accurate characterization association between BG level and PB may serve as a prognostic aid for clinical decision making allowing for research focused on personalized, preemptive treatment and monitoring of pregnancies impacted by T1DM.

Tables and Figures

Table 3.1. Patient characteristics (n=114; 16480 glucose readings over gestation 50-260 days)

	Median (IQR)	<i>n</i> (%)
Mother's age at LMP	27 (22-31)	
Age at diabetes onset	12 (8-17)	
Baseline BMI	23 (21-25)	
PB		32 (28)
Preeclampsia		32 (28)
White classification		
- B, C		40 (35)
- >=D		74 (65)
Presence of nephropathy		22 (19)
Presence of hypertension		14 (12)
Presence of retinopathy		16 (14)
Number of glucose recordings	157 (6-211)	

IQR: Interquartile range; BMI: body mass index (kg/m²)

Table 3.2. Parameter estimates and model fit statistics from JM and TDCM

	<u>TDCM</u>			<u>JM</u>		
	EST (CI)	OR (CI)	M-Val	LSM EST (CI)	ESM OR (CI)	M-Val
Mother's age at LMP	1.02 (0.95-1.10)			1.29 (1.10-1.72)*		
Age at diabetes onset	0.97 (0.92-1.03)			0.84 (0.62-0.98)*		
Presence of nephropathy		2.21 (1.01-3.32)*			2.29 (1.05-4.48)*	
Association						-0.01*
DIC			19932			19895
Log-likelihood			-9231.81			-9717.11

*: $p < 0.05$; LSM: Longitudinal sub-model; ESM: Event sub-model; EST: Estimate; OR: Odds ratio, M-Val: Model value

Table 3.3. Value, value plus slope, and, AUC association links and model fit statistics for the JMs

Association link	DIC	pD
Value	19990	2182
Value and Slope	19905	2143
AUC	19895	2130

D: effective number of parameters. Lower values indicate better model fit.

Figure 3.1: Fitted glucose model (on log scale) with natural spline over 50-260 gestational days

Figure 3.2: Individual dynamic prediction of risk for PB with updating BG readings (on log scale)

Chapter 4

A Bayesian Multivariate Joint Model for Medical Monitoring Data

Flexible Multivariate Joint Model of Longitudinal Intensity and Binary Process for Medical Monitoring of Frequently Collected Data

Abstract

A frequent problem in longitudinal studies is that subjects may be assessed at self-selected, irregularly spaced points in time, resulting in highly unbalanced outcome data. This induces bias especially if the availability of data is directly related to the outcome. The aim of this study was to develop a MVJM in a mixed outcomes framework to minimize irregular sampling bias in longitudinal studies. We demonstrate our approach on data involving BG monitoring throughout pregnancy and risk of PB among women with T1DM. Time points of BG measurements were unequally spaced and the IOGS varied between and within individual over time. A multivariate LME sub-model for the longitudinal outcome (BG), a Poisson model for the IOGS, and a logistic regression model for binary process (PB) is specified. The association between models is captured through shared random effects. MCMC methods were used to fit the model. MVJM provided better prediction, compared to JM with multivariate LME sub-model (ignoring (IOGS)) and a TS model ($AUC_{MVJM} = 0.65$, 95%CI: 0.61-0.72, $AUC_{JM} = 0.56$, 95% CI: 0.52, 0.64, $AUC_{TS}=0.59$, 95%CI: 0.48-0.65). The majority of the association parameters were significant in the PB outcome model, signifying that the predictive ability of the binary endpoint had been improved by sharing the random effects between glucose monitoring and PB outcome. Higher IOGS relates to higher risk of PB (OR= 2.26; 95% CI: 1.16, 3.12). Women with preeclampsia were at higher risk of PB (OR = 3.85; 95% CI (2.85, 5.27)); women with later diagnosis of diabetes had lower risk of PB outcome (OR = 0.51, 95% CI (0.39, 0.59)). A simulation study is presented to illustrate the effectiveness of the MVJM approach.

4.1 Introduction

Medical monitoring plays a significant role in modern clinical practice and has become a powerful disease management tool that can inform treatment decisions for chronic diseases and disorders, such as diabetes, yielding improved quality of life and longevity in affected patients^{67,68}. In addition to static variables collected on individuals, such as certain demographics, a longitudinal sequence of BG measurements is collected to act as a surrogate for disease state; this sequence often yields a complex, unbalanced data structure. Advancements in technology and clinical studies have improved diabetes self-management practices, allowing monitoring of BG on a continuous basis⁶⁹⁻⁷². For women with T1DM, frequent testing of BGs is performed to safely and effectively manage blood sugar levels. Tracking maternal glucose control throughout pregnancy typically relies on self-report or capture of monitoring by a self BG monitor, where timing of glucose measurements may be unequally spaced and the numbers of measurements may vary between women and within a given woman over time. The irregularities in BG sampling may be attributed to disease severity of the subjects, but epidemiologic studies to date haven't assessed this potential relationship. It is possible that women who have poor glucose control actively monitor and check glucose levels more frequently over the course of the entire pregnancy, compared to those who are able to adequately control glucose levels or vice versa. These types of irregular sampling patterns may involve a missing data mechanism which may be embedded in a subject's glycemic monitoring profile; in fact, the frequency of monitoring glucose may be correlated with adverse neonatal outcomes, such as PB, potentially giving further insight into her diabetes control/management.

Recent studies have shown that the rate of both spontaneous and indicated PB is increased in women with T1DM, compared to women whose pregnancies are not complicated by diabetes⁷³. Preterm babies go on to have a higher risk of increased morbidity, mortality and poor neurological outcome⁷⁴. In mothers with T1DM, poor glycemic control was associated with increased risks of both spontaneous and indicated PBs⁷⁵. Hence, characterizing and estimating rate of maternal glucose sampling over the course of the entire pregnancy is an important task to improve prediction of PB outcome and inform T1DM care management practices. As we have shown in our previously published work⁷⁶ and in preliminary studies provided in this proposal, nonlinear curvature of BG monitoring data, coupled with irregularities in both the number and timing of glucose measurements, makes it analytically challenging. Developing a prediction model to characterize the relationship between PB and BG monitoring using common statistical modeling techniques could mischaracterize the glycemic profile and thus may misrepresent the association with risk of adverse outcome. Too often, in analysis of irregularly sampled data, the stochastic nature of the sampling process is ignored, and simple regression of the observed responses onto the time of observation and any relevant covariates using GEE or an LME model is used. It is well known that, in general, GEE analysis is biased when the measurement process is dependent on the longitudinal outcome process, although it has been shown that the bias varies according to the choice of working-correlation matrix to model longitudinal correlation. Lipsitz et al⁷⁷ showed that LMEs work well when the observation process is ignorable and the correlation structure of the outcomes is correctly specified, and biased otherwise. Robins et al^{78,79} proposed a weighted GEE technique that yields consistent and asymptotically normally distributed estimators when the dependence of missingness based on the previously observed data is correctly specified. However, this method is not applicable to unbalanced longitudinal data. Lin et al⁸⁰ proposed inverse-intensity

weighting in generalized linear models in a way that attains the advantages of LME models for handling missing data. The weights are determined according to their frequency of occurrence per unit of time. However, these models are biased in the presence of non-monotone missing data⁸¹. In recent years, Molenberghs et al⁸² outlined a pseudolikelihood approach for missing data, which uses a simpler function assembled from pairwise likelihoods, and the missingness is accounted for by using inverse probability weighting. However, this method requires either modeling the observation time process parametrically or having the correct specification of the covariance structure of the repeated measure process⁸³.

In recent years, JM has been shown to be an efficient approach to simultaneously model longitudinal data and an observed outcome, often yielding more accurate, informative prediction than traditional models^{84,85}. Although recent joint modeling developments have demonstrated improved performance, compared to traditional models, joint modeling had its start in characterizing intermittent missing patterns in clinical trials, in order to alleviate bias in findings on the longitudinal outcome; however, these models are so-called univariate JMs, as they only include the intermittent missing pattern and longitudinal outcome. There still remains a knowledge gap concerning the role of the MVJM^{86,87}. Although the area remains relatively unexplored, MVJMs could feature a longitudinal model and subject-specific irregular sampling intensity simultaneously, in order to develop a better-fitting and more accurate predictive model for a binary event process.

The purpose of the current study is to develop an efficient, model-based method to predict a particular adverse neonatal outcome, PB, and characterize its association with a monitored longitudinal outcome, maternal BG, where this outcome forms a trajectory that is characterized by

irregular sampling resulting in missing data. In our proposed model, we focus on these mixed outcomes, incorporating longitudinal, count and binary outcome processes. To the best of our knowledge, no study has implemented a MVJM for mixed outcome processes of this nature, specifically when intensity or rate of sampling is embedded into a longitudinal process like BG monitoring, which in turn may have an effect on the event of interest, PB. Using a MVJM would allow these outcome processes to be modelled simultaneously through the shared random effects framework.

We present our proposed MVJM formulation in Section 2. Sections 3 and 4 include the estimation technique and application and results of our proposed method. Comparison of results between the proposed MVJM approach and other popular techniques, such as TS and univariate joint modeling, are presented in Section 5. The simulation design and results are presented in Section 6. Section 7 includes discussion and conclusion.

4.2 MVJM Formulation for Mixed Outcomes

The MVJM consists of three sub-models. For the i^{th} ($i = 1, \dots, n$) subject at gestation time j ($j = 1 \dots m$), we define the following three outcomes of each sub-model:

- (i) $Y_i = (y_{i1}, y_{i2} \dots y_{im})$ denotes the $1 \times m$ vector of the continuous longitudinal outcome (daily average glucose measurement in this application) for subject i ;
- (ii) $\lambda_i = (\lambda_{i1}, \lambda_{i2} \dots \lambda_{im})$ signifies the $1 \times m$ vector of discrete longitudinal outcome (IOGS per day) for subject i ;
- (iii) W_i denotes the binary endpoint (specifically PB in this application) for the event process.

Sub-model (i) of the continuous longitudinal outcome process can be written in terms of a LME model as

$$Y_i = X_i\beta + Z_ib_{1i} + \varepsilon_i \quad (1)$$

where X_i and Z_i are the design matrices of dimension $m \times p$ and $m \times q$ for fixed and random effects, respectively. These design matrices may involve splines basis functions to capture the nonlinear relationship between gestational day and Y_i . In Equation (1), β is the $p \times 1$ fixed effects, and, b_{1i} is the $q \times 1$ random effects parameter vector. ε_i denotes the vector of measurement error for subject i . We assume, $\varepsilon_i \sim N(0, \sigma_\varepsilon^2 I_m)$ and $b_{1i} \sim N(0, \Sigma_{b_1})$, where I_m is the $m \times m$ identity matrix; Σ_{b_1} is the $q \times q$ variance-covariance matrix for the random effects. We assume that ε_i and b_{1i} are independent.

Sub-model (ii) can be expressed in terms of a Poisson mixed effects regression for a discrete longitudinal outcome:

$$\lambda_i \sim \text{Poisson}(\psi_i) ; \text{ and}$$

$$\log(\psi_i) = X_i'\gamma + \tau_i b_{2i} \quad (2)$$

where γ and τ_i are the fixed and random effects parameter vectors respectively. In Equation (2), b_{2i} is the subject specific random effect with $b_{2i} \sim N(0, \Sigma_{b_2})$, where Σ_{b_2} is the variance-covariance matrix for the random effects of equation (2). We assume that the random effects of equations (1) and (2) have a multivariate normal distribution:

$$b_i = \begin{bmatrix} b_{1i} \\ b_{2i} \end{bmatrix} \sim MVN(0, D)$$

where D , the variance-covariance matrix of the two random effects can be expressed as

$$D = \begin{bmatrix} \Sigma_{b_1} & \Sigma_{b_1} \Sigma_{b_2} \\ \Sigma_{b_2} \Sigma_{b_1} & \Sigma_{b_2} \end{bmatrix}.$$

Sub-model (iii) is a generalized linear model with logit link for the binary endpoint, where

$W_i \sim \text{Bernoulli}(\eta_i)$ and can be expressed as

$$g(\eta_i) = X_i' \theta + \alpha b_i \quad (3)$$

where $g(\cdot)$ is the logit link function; θ is the parameter to be estimated; b_i involves the random effects of the two longitudinal processes specified in Equations (1) and (2). The binary event process in equation (3) is linked to the continuous and discrete longitudinal processes through the shared random effects. The parameter vector α measures the strength of association between the three models. It is implied that, in the presence of strong association, the MVJM would improve the predictive ability for the binary outcome process.

4.3 Estimation

The conditional joint distribution of Y_i, λ_i, W_i , given b_{1i}, b_{2i} can be expressed as

$f(Y_i | X_i, b_{1i}) f(\lambda_i | X_i, b_{2i}) f(W_i | X_i, b_{1i}, b_{2i})$. The marginal density of (Y_i, λ_i, W_i) for subject i can

be obtained by integrating out the random effects, which can be expressed as

$$p(Y_i, \lambda_i, W_i | X_i, \beta, \gamma, \theta, \alpha, \sigma_\varepsilon^2, D) = \int f(Y_i | X_i, b_{1i}, \beta, \sigma_\varepsilon^2) f(\lambda_i | X_i, b_{2i}, \gamma) f(W_i | X_i, \theta, \alpha, D) db$$

Then the likelihood function for subject i can be written as

$$l(\beta, \gamma, \theta, \alpha, \sigma_\varepsilon^2, D) = \prod_{i=1}^n p(Y_i, \lambda_i, W_i | X_i, \beta, \gamma, \theta, \alpha, \sigma_\varepsilon^2, D)$$

The key assumption of this MVJM is the conditional independence of the outcome processes, such that, the continuous and discrete longitudinal processes relate to the binary event endpoint only through the subject-specific random effects.

The estimation of the MVJM was formulated using a Bayesian approach. Posterior distributions of the parameters are derived under the assumption that, given the shared random effects, the continuous and discrete longitudinal outcomes are independent of the final binary event. We use MCMC methods via Gibbs sampler⁸⁸ to simultaneously estimate all unknown model parameters. The MCMC method avoids the numerical complications of integrating over the shared random effects. The positive aspect of MCMC is that it iteratively draws the next values of parameters and unobserved response data points from their corresponding posterior and predictive distributions. We used the WinBUGS⁸⁹ software for estimating the posterior distribution of all model parameters. We specified non-informative prior distributions for all parameters in the model. By choosing non-informative priors, the posterior distribution means are asymptotically equivalent to the maximum likelihood estimators for the MVJM. A multivariate normal distribution with mean 0 and large diagonal elements variance-covariance matrix can be used for regression parameters $\beta, \gamma, \theta, \alpha$ in the model. For the residual error variance σ_ϵ^2 , an inverse gamma prior with a large scale parameter value can be specified. A noninformative multivariate Inverse-Wishart distribution can be used for the variance-covariance matrix of the random effects' distribution.

4.4 Application and Results

4.4.1 Cohort Data and Study Variables

We applied our proposed MVJM method by conducting a retrospective longitudinal cohort study using an established T1DM birth cohort. The cohort is comprised of mothers with T1DM who underwent

daily glucose monitoring throughout the course of their pregnancies⁹⁰. This cohort of women with pre-gestational diabetes was enrolled in an NIH-funded Diabetes in Pregnancy Program Project Grant (DiPP) [NICHD 11725]. The information prospectively collected detailed maternal and neonatal data, including monthly glycohemoglobin A1, units of insulin administered, blood pressure, weight, urinalysis, hematology, renal status and nutrition, plus extensive information on maternal and neonatal outcomes. In the current study, we included observations between 50 to 260 days of gestation. Mother's age at diabetes onset and preeclampsia were included as fixed effects in all the three sub-models under the MVJM framework.

Table 4.1 has the descriptive statistics of the variables included in the models. A total of 114 women with 16,480 observations over 50-260 days of gestations were analyzed in this study. The median (range) level of glucose concentration was 131 (25-400) mg/dL, with significant variations in the average days monitored across women (median: 157 days range: 6-211 days). The IOGS varied across women, with median (range) number of recordings over gestation of 200 (62-370). A random sample of five women showed large within-subject and between-subject variations in IOGS over the gestational period (Figure 4.1). Twenty eight percent of the entire cohort had a PB. Median (range) maternal age at diagnosis of diabetes was 12 (8-17) years, and twenty eight percent of women was diagnosed with preeclampsia until after 20 weeks gestation.

4.4.2 MVJM Application

Within the LME model framework, the glucose recordings from the i^{th} women ($i=1,2,\dots,114$) at gestation t_{ij} ($j=1,2,\dots,16,480$) can be expressed as

$$\begin{aligned}
(\text{Glucose reading})_{ij} = & \beta_0 + b_{0i} + (\beta_1 + b_{1i})\text{Gestational day}_{ij} + (\beta_2 + b_{2i})\text{Gestational day}_{ij}^2 + \\
& (\beta_3 + b_{3i})\text{Gestational day}_{ij}^3 + \sum_{k=1}^K u_k (\text{Gestational day}_{ij} - K_k)_+^3 + \\
& \beta_4 \text{Preeclampsia}_i + \beta_5 \text{Age diabetes onset}_i + \varepsilon_{ij}
\end{aligned} \tag{4}$$

Equation (4) represents the longitudinal continuous outcome process sub-model under MVJM framework. The terms of the polynomial coefficients, $\beta_0, \beta_1, \beta_2, \beta_3$ are for the intercept, slope, quadratic and cubic terms to model global curvature of the maternal glycemic profile. The summed expression is comprised of the knots at distinct locations $K_k (k = 1 \dots K)$ evaluated at gestational day t_{ij} with corresponding coefficients $u_1, u_2 \dots u_k$. The expression $(\text{Gestational day}_{ij} - K_k)_+^3$ is the positive part of the function (ie., $\max\{0, \text{Gestational day}_{ij} - K_k\}$).

A total of 22 knots were selected, ranging from 59.1 to 250.8 gestational days, using the quantile method described by Ngo and Wand⁹¹. The summed expression in equation (4) becomes

$$\sum_{k=1}^K u_k (\text{Gestational day}_{ij} - K_k)_+^3 = (u_1 + b_{4i})s_{1ij} + (u_2 + b_{5i})s_{2ij} + \dots + (u_{22} + b_{25i})s_{22ij}$$

Where $u_1, u_2 \dots u_{22}$ represent the fixed effects part of the nonlinear glycemic recording process. In equation (4), we have 26 random effects parameters with b_{0i} (intercept), b_{1i} (slope), b_{2i} (quadratic), b_{3i} (cubic), and $b_{4i}, \dots, b_{25i}$ (nonlinear glucose trajectory). The measurement error is distributed as $\varepsilon_{ij} \sim N(0, \sigma_\varepsilon^2)$. While β_4 represents regression parameters associated with presence of preeclampsia, β_5 denotes age at onset of diabetes for the i^{th} women. A random sample of six women's fitted glucose trajectories with 22 knots are presented in Figure 4.2, reflecting that the LME were able to capture the non-linear curvature in maternal glycemic profiles.

Within the Poisson regression framework, representing the second sub-model, we define ψ_{ij} as the count of glucose readings for the i^{th} women ($i=1,2,\dots,114$) at gestation t_{ij} ($j=1,2,\dots,16,480$). Then, using the log link function, IOGS for the i^{th} women can be expressed as

$$\log(\psi_{ij}) = \alpha_0 + b_{26i} + \alpha_1 \text{Pr } e eclampsia_i + \alpha_2 \text{Age at diabetes onset}_i + \alpha_3 \text{Day}_{ij} \quad (5)$$

where $\alpha_0, \alpha_1, \alpha_2$ are the regression parameters associated with the intercept, and fixed effects (presence of preeclampsia for the i^{th} women, and, age on diabetes onset at baseline for the i^{th} women) terms. α_4 is the estimated coefficient for the offset parameter, day_{ij} , and b_{26i} represent subject specific random effect term.

We assume that random effects of equation (4) and (5) are jointly distributed as:

$[b_{0i}, b_{1i}, \dots, b_{26i}]^T \sim MVN(0, D)$, with variance-covariance matrix is given by

$$D = \begin{bmatrix} \sigma_{b_{0i}}^2 & \sigma_{b_{0i} b_{1i}} & \sigma_{b_{0i} b_{2i}} & \dots & \sigma_{b_{0i} b_{27i}} \\ & \sigma_{b_{1i}}^2 & \dots & \dots & \sigma_{b_{1i} b_{27i}} \\ & & \sigma_{b_{2i}}^2 & \dots & \dots \\ & & & \dots & \dots \\ & & & & \sigma_{b_{26i}}^2 \end{bmatrix}$$

For the third sub-model, representing the event outcome process, we used the generalized linear regression with logit link, expressed as

$$\text{logit}(\pi_i) = \gamma_0 + \gamma_1 \text{Pr } e eclampsia_i + \gamma_2 \text{Age at diabetes onset}_i + a_1 b_{0i} + a_2 b_{1i} + \dots + a_{27} b_{26i} \quad (6)$$

Where π_i denotes the probability of the i^{th} women having PB outcome; $\gamma_0, \gamma_1, \gamma_2$ are the regression parameters for intercept, presence of preeclampsia, and, age at diabetes onset for the i^{th} women.

The coefficients $a_1, a_2, \dots, a_{27}$ in equation (6) are the strength of association between the binary event,

PB, and continuous and discrete longitudinal processes represented by glucose recordings, and IOGS. This allowed the simultaneous estimation of parameters from all three models (LME, Poisson, and, Binary) processes in one model, accounting for the association between processes through the incorporation of shared random effects.

For regression coefficients $\beta_0, \beta_1, \beta_2, \beta_3, \beta_4, \beta_5, u_1, u_2, \dots, u_{22}, \alpha_0, \alpha_1, \alpha_2, \alpha_3, \gamma_0, \gamma_1, \gamma_2, a_1, a_2 \dots a_{27}$ we used a normal prior distribution with mean 0, and variance 1000 for each parameter. An inverse Wishart prior with 27 degrees of freedom was used for the variance of random effects. An inverse Gamma prior, $IGamma(0.01, 0.01)$ was used for σ_ϵ^2 . We ran the MCMC with a single chain for 120,000 iterations, thinning by 4, and we discarded the first 100,000 iterations as burn-in. Thus, we obtained 20,000 samples of targeted posterior distributions of the unknown parameters for statistical inference. Convergence was monitored by trace plot and the Geweke diagnostic test⁹² in the *coda* package in *R2Winbugs*, version 2.1-21.

The results of the proposed MVJM are presented in Table 4.2. The majority of the association parameters ($a_1, a_2 \dots a_{26}$) were significant in the PB outcome model, signifying that the predictive ability of the binary endpoint had been improved by sharing the random effects between glucose monitoring and PB outcome (discussed in detail in Section 5). The significant association between random effects from Poisson regression and PB event model (a_{27}) reflected that higher IOGS relates to higher risk of PB (OR= 2.26; 95% CI: 1.16, 3.12). In the shared random effects framework with polynomial time and spline effects, separate interpretation of each of the association parameters for each time and spline coefficients is not recommended⁷⁴. Hence, the association parameters ($a_1, a_2 \dots a_{27}$) joining the random effects between both longitudinal processes (glucose monitoring and IOGS) and binary endpoint (PB) were examined to check if the

predictive ability of MVJM is improved, which is the focal point of the current study. For the fixed effects, the MVJM result reflected that women with preeclampsia were at higher risk of PB (OR = 3.85; 95% CI (2.85, 5.27)). The result also showed that women with later diagnosis of diabetes had lower risk of PB outcome (OR = 0.51; 95% CI (0.39, 0.59)).

4.5 Comparison with Other Models

We wanted to compare the proposed MVJM approach with the traditional TS regression approach, and a JM ignoring the Poisson sub-model. The TS approach involves (i) fitting the longitudinal outcome using an LME model, and then (ii) including the subject specific random effects from the LME model as covariates in the binary event outcome model. In comparison to the JM approach, the advantage of TS regression is that it involves fewer parameters to be estimated with simpler convergence techniques. The standard software package can be used for the TS regression. However, TS regression is biased in the presence of unbalanced irregular longitudinal measurements, mainly due to bias in the best linear unbiased predictor, and increasing uncertainty in the parameter estimate^{93,94}.

We used the TS regression method to model the longitudinal measurements of glucose concentration using the same sub-model described in Equation (4), and then the posterior expectations of all the random effects ($\tilde{b}_{0i}, \dots, \tilde{b}_{25i}$) are used in the binary event sub-model of PB as:

$$\text{logit}(\pi_i) = \omega_0 + \omega_1 \text{Preeclampsia}_i + \omega_2 \text{Age at diabetes onset}_i + \omega_4 \hat{b}_{0i} + \omega_5 \hat{b}_{1i} + \dots + \omega_{27} \hat{b}_{26i} \quad (6)$$

$i = 1, 2, \dots, 114$

Additionally, we compared the proposed MVJM to a JM ignoring the IOGS process. To maintain consistency, we used the same sub-models for longitudinal and event outcome as mentioned in Section 4.2. The prior distributions for TS and JM are used in a similar fashion to the proposed

MVJM. The parameters were estimated using MCMC methods in the similar fashion as mentioned in Section 4.2.

The results of the TS and JM ignoring sampling intensity are presented in Table 4.2. Even though the TS model detected the same significant effect in age at diabetes onset as the MVJM, it did not yield significant effects for the random slope estimates in the binary event sub-model as in the MVJM approach. Figure 4.3 shows the predictive accuracy of the three models by comparing the AUC⁹⁵. The proposed MVJM produced a higher AUC (0.65, 95%CI: 0.61-0.72), compared to the TS model (0.59, 95%CI: 0.48-0.65) and univariate JM ignoring sampling intensity (0.56, 95% CI: 0.52, 0.64). By simultaneously modeling the sampling intensity mechanism using Poisson regression, the MVJM approach produced higher predictive ability compared to TS and JM ignoring IOGS in estimating risk of PB event outcome.

4.6 Simulation

We performed a set of 50 simulations to investigate the proposed method under the MVJM framework.

4.6.1 Design and Analysis

We used N=90 subjects for each simulated data set. To simulate the continuous outcome for the longitudinal sub-model, we used the following LME model per data set, i.e.

$$y_i(t) = f_i(t) + \varepsilon_i(t) = \beta_0 + \beta_1 t + \beta_2 (\text{preeclampsia})_i + b_{0i} + b_{1i} t + \varepsilon_i(t)$$

where $\varepsilon_i \sim N(0, \sigma_\varepsilon^2)$. For simplicity, we adopted a linear mean response function for both fixed and random parts of the model and adjusted for one binary variable, preeclampsia.

To simulate the discrete outcome for the longitudinal sub-model, we used the following Poisson regression with one random part, the subject specific random intercept, i.e.,

$$\log(\lambda_i(t)) = \eta_i(t) = \alpha_0 + \alpha_1(\text{preeclampsia})_i + b_{2i}$$

$$\text{with } b_i = (b_{0i}, b_{1i}, b_{2i})' \sim N_3(0, \Sigma_b), \text{ where } \Sigma_b = \begin{bmatrix} \sigma_{b_{0i}}^2 & \sigma_{b_{0i}} \sigma_{b_{1i}} & \sigma_{b_{0i}} \sigma_{b_{2i}} \\ & \sigma_{b_{1i}}^2 & \sigma_{b_{1i}} \sigma_{b_{2i}} \\ & & \sigma_{b_{2i}}^2 \end{bmatrix}$$

For the event sub-model, we assumed the following model. In particular,

$$\logit(p_i) = \gamma_0 + \gamma_1(\text{preeclampsia})_i + a_1 b_{0i} + a_2 b_{1i} + a_3 b_{2i}$$

We linked the binary event model by the three subject specific random effects from the longitudinal processes. We used the MCMC technique via Gibbs sampler to estimate all the parameters of three above specified simulated model. The software package WinBUGS was used to estimate the posterior distribution of the parameters. We specified the prior distribution of all parameters as follows:

$$\begin{aligned} \beta_k &\sim N(0, 1000) \quad k = (0, 1, 2); \quad \alpha_k \sim N(0, 1000) \quad k = (0, 1) \\ \gamma_k &\sim N(0, 1000) \quad k = (0, 1); \quad a_k \sim N(0, 1000) \quad k = (1, 2, 3) \\ \sigma_\varepsilon^2 &\sim \text{IGamma}(0.01, 0.01); \quad \Sigma_b^{-1} \sim \text{Wishart}(I_3, 3) \end{aligned}$$

We ran the MCMC using a single chain with 50,000 iterations, 45,000 burn-in and 5 thinning for each dataset. The true values of each parameter are presented in table 4.3. These true values for the parameters $\beta_s', \alpha_s', \gamma_s'$ are taken by fitting our application data by the simplified version of LME, Poisson and logistic regression as mentioned above.

To test whether the association parameters (shared random effects in the binary event sub-model) contribute to explaining the variability in the event sub-model, and hence improving the predictive ability for the binary endpoint, we specify the global null hypothesis as

$$H_0 : a_1 = a_2 = a_3 = 0$$

$$H_a : a_1 \neq 0 \text{ or } a_2 \neq 0 \text{ or } a_3 \neq 0$$

In order to check for predictive ability of the binary endpoint, we computed the number of times the null hypothesis was rejected for the association parameters in our proposed MVJM approach. Rejection was based on whether or not the 95% credible interval included zero.

4.6.2 Simulation Results

The results of the simulation study are presented in Table 4.3. In particular, we had more than 80 percent coverage probability for all the parameters (shown on log scale) in the longitudinal sub-model. The same held for the binary event sub-model. However, the coverage probabilities of the Poisson parameters were low compared to the other two sub-models. This may be attributable to the higher variation factor in the Poisson model from the empirical data generation process. For all parameters, bias and root mean squared error (a measure of predictive accuracy) were of reasonable magnitude. For testing the association parameters, a_1 , a_2 , a_3 , the proportion of time the global null hypothesis got rejected was 0.98, 0.96, and 0.76 respectively. This indicated the improvement of prediction ability of the binary event endpoint by our proposed MVJM approach.

4.7 Discussion and Conclusion

We developed and applied a MVJM for predicting PB that also fit BG trajectories and accounted for differing intensity within which BG levels were sampled. Within the Bayesian framework we

specified the continuous longitudinal response process, glucose monitoring over the entire pregnancy by LME which was flexible and adaptive to reflect patient-specific temporal behaviors in mother's glycemic profiles. Instead of using a known shape for curve fitting, we used cubic penalized regression splines to estimate each woman's glycemic profile. This helped us avoid a loss of information by providing a more flexible modeling structure for curvature in the data. By using Ngo and Wand's quantile based approach⁹¹ we selected 22 knots to reflect the nonlinearity in glucose data. Our second sub-model was based on modeling the IOGS process. Under the generalized linear model framework, we used Poisson regression with log link to model mother's IOGS over the entire pregnancy. While in LME, the correlation between the repeated measurements in the longitudinal outcome is captured by the random intercepts and slopes parameters, for the Poisson sub-model, it is captured by random intercept only. Our final event outcome model, PB was set up within logistic regression framework. The association between the continuous and discrete longitudinal processes with the binary event endpoint is captured by shared random effects.

Results from the proposed MVJM indicate that presence of preeclampsia and early onset of diabetes increased the risk of PB. Additionally, it reflected that women with higher IOGS tended to have higher risk of PB outcome. This finding was interesting as it was expected that in this T1DM cohort, those who had poor glycemic control actively monitored and check glucose levels less frequently over the course of the entire pregnancy, compared to those who were able to adequately control glucose levels. These additional insight was not captured when we compared the proposed model with the TS regression and univariate JM excluding separate modeling for IOGS mechanism.

PBs in pregnancies complicated by T1DM have been associated with poor glycemic control. Tracking maternal glucose control throughout pregnancy historically relies on self-reports of monitoring. Advances in technology have the capacity to promote improved glycemic management strategies during pregnancy; however, improvements have not been matched by statistical analysis and implementation science needed to aid the healthcare provider in interpreting the massive amounts of information available from glucose recordings. A frequent problem in longitudinal studies with self-monitoring data is that the subjects may be assessed at self-selected points in time. As a result, observed outcome data may be highly unbalanced. This induces bias if the availability of the data is directly related to the outcome status, and current analytical methods are unsuitable to combat this type of bias⁹⁶. In recent years, JM has been used to characterize associations primarily between longitudinal and event processes in the presence of missing data¹⁸. These models originated as a means to account for drop-out in longitudinal studies, following the missing not at random assumption⁹⁷. Univariate JM allows the simultaneous estimation of parameters from both the longitudinal and survival processes in a single, parsimonious model, typically through the incorporation of shared random effects. Developing a prediction model to characterize the relationship between PB and BG monitoring using common statistical modeling techniques could mischaracterize the glycemic profile and risk of adverse outcome. Using currently available methods such as GEE may lead to biased inference since it requires the correct specification of either the observed time process or the repeated measure process with a correct covariance structure.

In recent years, the MVJM is gaining momentum to model multiple longitudinal outcomes and possibly multiple event processes⁸⁷. However, no previous study focused on simultaneous estimation of a mixture of outcomes, incorporating longitudinal, count and binary outcome processes. The

development of a MVJM involving mixed outcomes was motivated by a real life example data where intensity or rate of sampling is embedded into a longitudinal process like BG monitoring, which in turn may have an effect on the event of interest, PB. The proposed method allowed these three outcome processes to be modelled simultaneously through the shared parameter framework. From the estimation perspective, the proposed JM can be implemented via MCMC techniques and fitted using the readily available WinBUGS software.

Our proposed MVJM produced higher AUC compared to TS or univariate JM without modeling the glucose sampling intensity process. Moreover, the TS regression approach to estimate the relationship between a longitudinal process and a binary endpoint may result in bias and loss of efficiency, and is, therefore inadvisable. The MVJM yielded better predictive accuracy compared to TS and univariate JM ignoring rate of sampling, and hence improved the predictive ability for the final binary endpoint, PB outcome. While the frequentist JM likelihood formulation may be computationally expensive, the MVJM under the Bayesian framework was useful to capture the complex features in the association among multiple longitudinal processes and an binary event process. The posterior distribution of the model parameters given the observed data was derived under the assumption of conditional independence. Specifically, given the random effects, the longitudinal outcomes were independent from the event outcome process. A simulation study reflected good performance of the proposed model.

There are some weaknesses in the proposed MVJM approach. For the LME sub-model, we used a total of 22 knots to fit the nonlinear curvature of the glucose trajectory. Even though it's possible that there was model overfitting, we were able to capture the extreme nonlinearity in the maternal glycemic profile. Additionally, due to the large number of parameters, the convergence of Markov

chains was slow during the estimation phase. More adaptive approaches for knot selection that maintain reasonable computing speed represent important areas for future work.

Being able to accurately predict the risk of an adverse neonatal outcome, the proposed MVJM within a Bayesian framework increased our epidemiologic understanding of the link between glucose level, rate of sampling maternal glucose and odds of PB. The model was able to address fundamental analysis flaws currently hindering perinatal research and other areas of chronic disease research that rely on feedback from medical monitoring data. Our approach could be applied to other medical monitoring studies subject to irregular sampling bias.

Tables and Figures

Table 4.1. Participants' Characteristics (N=114)

	Median (Range)	<i>n</i> (%)
Mother's age at LMP	27(16-38)	
Mother's BMI at LMP	23 (16-31)	
Glucose (mg/dl)	131 (25-400)	
Average number of glucose sampling	200 (62-370)	
Average number of days of glucose sampling	157 (6-211)	
PB		32(28)
Presence of preeclampsia		32 (28)
Age at diagnosis of diabetes	12 (8-17)	

Note: BMI: Body mass index; PB is defined as birth outcome at less than 37 completed weeks of gestation.

Table 4.2. Model results of Posterior means, SDs, and the 95% credible intervals fitted for the DiPP data

Multivariate Joint Model					Two-stage Model					JM Ignoring Sampling Intensity				
Parameter	Mean	SD	2.50%	97.50%		Mean	SD	2.50%	97.50%		Mean	SD	2.50%	97.50%
β_0	20.64	73.07	-121.50	165.21		21.77855	71.76230	-118.50250	160.20250		19.97242	71.98373	-122.40250	161.30250
β_1	-1.92579	1.66471	-5.15423	1.31700		-2.03126	1.64454	-5.15040	1.22905		-1.86485	1.64433	-5.11443	1.34403
β_2	0.03443	0.02917	0.02234	0.09111		0.03657	0.02892	-0.02089	0.09171		0.03339	0.02881	-0.02272	0.09049
β_3	-0.00021	0.00017	-0.00054	0.00012		-0.00023	0.00017	-0.00055	0.00011		-0.00020	0.00017	-0.00053	0.00013
β_4	20.11074	72.22327	-120.22	160.11		21.36680	71.48066	-119.21	159.90		19.54201	71.97005	-122.90	161.00
β_5	-0.00076	0.00253	-0.00468	0.00449		-0.02235	0.00529	-0.03155	0.01209		0.00292	0.00259	-0.00115	0.00835
u_1	0.00035	0.00022	-0.00009	0.00079		0.00148	0.00070	0.00017	0.00264		0.00034	0.00022	-0.00009	0.00077
u_2	-0.00024	0.00010	-0.00045	-0.00004		-0.00414	0.00174	-0.00698	-0.00104		-0.00024	0.00010	-0.00045	-0.00004
u_3	0.00013	0.00008	-0.00003	0.00030		0.00686	0.00221	0.00281	0.01073		0.00013	0.00008	-0.00003	0.00030
u_4	-0.00002	0.00008	-0.00017	0.00013		-0.00706	0.00213	-0.01052	-0.00246		-0.00002	0.00008	-0.00017	0.00013
u_5	-0.00005	0.00007	-0.00019	0.00010		0.00254	0.00300	-0.00144	0.01001		-0.00005	0.00007	-0.00019	0.00010
u_6	0.00002	0.00007	-0.00012	0.00016		0.00249	0.00303	-0.00515	0.00589		0.00002	0.00007	-0.00012	0.00016
u_7	0.00006	0.00007	-0.00007	0.00020		-0.00210	0.00187	-0.00581	0.00079		0.00006	0.00007	-0.00007	0.00020
u_8	-0.00013	0.00007	-0.00027	0.00001		0.00006	0.00370	-0.00619	0.00612		-0.00013	0.00007	-0.00027	0.00001
u_9	0.00012	0.00007	-0.00001	0.00026		-0.00132	0.00321	-0.00658	0.00380		0.00012	0.00007	-0.00002	0.00026
u_{10}	-0.00004	0.00007	-0.00017	0.00009		0.00433	0.00237	-0.00065	0.00814		-0.00004	0.00007	-0.00018	0.00009
u_{11}	-0.00004	0.00007	-0.00017	0.00008		-0.00608	0.00286	-0.01020	-0.00023		-0.00004	0.00007	-0.00018	0.00009
u_{12}	0.00005	0.00007	-0.00007	0.00019		0.00456	0.00328	-0.00219	0.00989		0.00006	0.00007	-0.00008	0.00019
u_{13}	-0.00002	0.00007	-0.00015	0.00011		-0.00122	0.00344	-0.00806	0.00521		-0.00003	0.00007	-0.00016	0.00010
u_{14}	0.00001	0.00007	-0.00012	0.00014		-0.00103	0.00330	-0.00571	0.00534		0.00002	0.00007	-0.00011	0.00015
u_{15}	-0.00002	0.00007	-0.00015	0.00011		0.00155	0.00283	-0.00364	0.00674		-0.00003	0.00007	-0.00016	0.00010
u_{16}	0.00003	0.00007	-0.00010	0.00016		-0.00115	0.00299	-0.00685	0.00520		0.00002	0.00007	-0.00011	0.00015
u_{17}	-0.00002	0.00007	-0.00016	0.00011		0.00192	0.00364	-0.00494	0.00795		-0.00001	0.00007	-0.00014	0.00012
u_{18}	0.00005	0.00007	-0.00008	0.00018		-0.00323	0.00390	-0.00965	0.00360		0.00005	0.00007	-0.00008	0.00018
u_{19}	-0.00011	0.00007	-0.00026	0.00003		0.00128	0.00364	-0.00523	0.00705		-0.00013	0.00007	-0.00027	0.00000
u_{20}	0.00009	0.00008	-0.00006	0.00024		0.00293	0.00354	-0.00209	0.00952		0.00015	0.00008	0.00000	0.00030
u_{21}	-0.00019	0.00010	-0.00038	0.00000		-0.00434	0.00380	-0.01111	0.00183		-0.00017	0.00010	-0.00035	0.00002
u_{22}	-0.00038	0.00066	-0.00161	0.00069		-0.00330	0.00620	-0.01472	0.00691		-0.00502	0.00070	-0.00635	-0.00362

Multivariate Joint Model					Two-stage Model					JM Ignoring Sampling Intensity			
Parameter	Mean	SD	2.50%	97.50%	Mean	SD	2.50%	97.50%	Mean	SD	2.50%	97.50%	
α_0	1.03340	0.31696	0.60649	1.75902									
α_1	-0.13305	0.25279	-0.49911	0.34502									
α_2	-0.00306	0.00152	-0.00609	-0.00067									
α_3	0.11290	0.03321	0.05070	0.17420									
γ_0	4.72748	0.18829	4.31900	5.06600	-2.02213	0.38658	-2.50403	-1.09600	5.47731	0.42662	4.70395	6.23502	
γ_1	3.85636	0.73862	2.85200	5.27830	-0.02928	1.03356	-1.54500	1.73203	-1.00897	0.47044	-1.83703	-0.33580	
γ_2	-0.68937	0.10803	-0.93101	-0.52380	-2.24903	0.37581	-2.87000	-1.51398	-1.04178	0.55525	-1.92203	-0.00945	
a_1	-3.24270	0.36363	-3.72300	-2.43798	-2.46974	0.59844	-3.37103	-1.39598	2.00095	0.51896	0.91439	2.68600	
a_2	3.92538	0.47305	3.19300	4.77002	-0.75815	0.49513	-1.68100	0.03285	-0.25282	0.42077	-1.17200	0.48554	
a_3	-2.93019	0.78425	-4.38500	-2.03200	-0.38502	0.65317	-1.32500	0.70400	-0.43234	0.24408	-0.86842	0.13000	
a_4	2.73427	0.32173	2.05595	3.24902	0.52755	0.43667	-0.28570	1.15100	-2.26420	0.35636	-2.88700	-1.64100	
a_5	1.98766	0.21249	1.51500	2.39100	0.70187	0.57323	-0.22381	1.64505	-1.66203	0.55342	-2.63200	-0.81729	
a_6	1.24677	0.32944	0.56829	1.75000	-1.26022	0.34612	-1.88403	0.75099	2.67283	1.46691	-0.20088	4.67500	
a_7	1.87885	0.39792	1.18687	2.64402	1.87973	0.76064	0.68659	3.11102	3.83865	0.60935	3.00397	4.99600	
a_8	-0.94244	1.02199	-2.62003	1.06203	1.47880	0.47178	0.81829	2.33900	-0.06165	0.82962	-1.32000	1.23203	
a_9	0.36325	0.34490	-0.27333	1.04700	-0.02615	0.34738	-0.55010	0.81094	1.52090	0.54010	0.22217	2.17100	
a_{10}	1.19993	0.58058	0.32649	2.28400	2.23117	0.39575	1.28800	3.04502	0.08795	0.43557	-0.92010	0.70420	
a_{11}	3.58606	0.21784	3.20900	4.09702	-1.38978	0.37414	-1.93403	-0.54880	1.46613	0.70675	0.01625	2.39603	
a_{12}	2.31669	0.24071	1.75297	2.70500	0.32824	0.41079	-0.27140	1.17300	-0.27219	0.46353	-0.95450	0.90912	
a_{13}	1.71981	0.48087	1.01495	2.72202	-0.64885	0.55204	-1.65903	0.27270	-0.88009	0.80428	-2.28603	0.68847	
a_{14}	0.51829	0.15661	0.24319	0.80843	-3.51037	0.71117	-4.57203	-2.24400	2.00829	0.68390	0.87419	3.83502	
a_{15}	-2.44328	0.51142	-3.43305	-1.78000	-1.97696	0.55449	-2.75300	0.79189	-1.52966	1.10011	-3.39600	0.37211	
a_{16}	1.13129	0.33455	0.51953	1.63202	-0.38886	0.47874	-1.27400	0.71130	-1.57432	0.35920	-2.20500	-0.82880	
a_{17}	-6.87472	0.43655	-7.60003	-6.11700	0.24148	0.43768	-0.84661	0.97160	2.74335	0.41139	2.11300	3.50700	
a_{18}	-1.10388	0.49111	-1.98800	-0.24720	2.87194	0.41378	2.10500	3.56900	-0.57904	0.37745	-1.20803	0.08729	
a_{19}	-0.45924	0.38234	-1.05803	0.13831	0.29786	0.65434	-0.64990	1.20200	-0.96245	0.66941	-2.23503	-0.02380	
a_{20}	3.52111	0.27041	2.93100	3.87400	1.56765	0.60464	0.43159	2.37902	3.67450	0.42990	3.07897	4.64802	
a_{21}	0.20370	0.21480	-0.29631	0.61533	0.29109	0.29320	-0.25333	0.84631	-0.31138	0.55600	-1.31100	0.42161	
a_{22}	0.07718	0.44236	-0.68761	0.84591	-2.65233	0.41827	-3.48100	-2.06798	-1.80593	1.15562	-3.79500	-0.00290	
a_{23}	2.54982	0.24575	2.21300	3.11400	2.41412	0.47316	1.50100	3.36300	1.53538	0.75762	0.42736	2.78200	
a_{24}	-0.1304	0.4671	-1.224	-0.11561	0.97919	0.52262	-0.40580	1.93500	0.46236	0.36559	-0.19650	1.06600	
a_{25}	0.42824	0.53659	0.39390	0.99550	0.25746	0.41823	-0.31370	1.31600	5.89192	0.56218	4.63987	6.80602	
a_{26}	-3.51196	0.32363	-4.16000	-2.97698	0.03459	0.01725	-0.00062	0.06475	-0.77873	0.20077	-1.26900	-0.48100	
a_{27}	0.83755	0.19838	0.43398	1.15900									
σ_ϵ^2	0.09363	0.00105	0.09163	0.09571	0.11531	0.0019	0.11323	0.11682	0.10213	0.00103	0.101	0.10262	
* Bold values indicate significant effects at 0.05 significance level													

Table 4.3. Simulation results including the true parameter values, the bias (over the MCMC samples) for each parameter, the root mean squared error, and the 95% coverage probability (over the MCMC samples) for each parameter

Parameter	True values	Bias	RMSE	Coverage	Proportion (H ₀ rejected)
β_0^*	4.9	-0.39	0.51	0.87	
β_1^*	-0.005	-0.004	0.01	0.84	
β_2^*	0.02	0.28	0.53	0.81	
α_0^*	1.3	-0.25	0.35	0.78	
α_1^*	-0.04	0.33	1.06	0.66	
γ_0^{**}	-1.7	0.24	0.42	0.83	
γ_1^{**}	1.3	0.19	0.56	0.80	
a_1					0.98
a_2					0.96
a_3					0.76

Note: Bias = $E(\hat{\theta} - \theta)$; RMSE = $\sqrt{E([\hat{\theta} - \theta])^2}$; Coverage: proportion of samples for which the true values are contained in the 95% CI. True values are on log scale.

Figure 4.1: The IOGS (on log scale) of 8 randomly selected women throughout pregnancy. These profiles show large variation both between and within subject during pregnancy.

Figure 4.2: Fitted glycemic trajectory of six randomly selected women with 22 knots using LME model. The gray dots are observed glucose level (on log scale) and the black solid line is the fitted line by LME across gestational days.

Figure 4.3: ROC curves based on MVJM, TS model, and JM ignoring sampling intensity

Chapter 5

Conclusion

5.1 Summary

This dissertation focused on development and application of novel statistical modelling aimed at improving prediction ability for medical monitoring studies. The work was motivated by a retrospective longitudinal cohort study using an established T1DM birth cohort. In Chapter 2, we proposed a semiparametric mixed effects regression to examine the complex form of subject specific rates and acceleration/deceleration change in glycemic profiles and its association with adverse birth outcome, specifically for delivering a LGA baby in women with type 1 diabetes. Moreover, we wanted to examine the difference in velocity, acceleration and deceleration in glucose trajectory over gestation according to LGA versus AGA birth. In Chapter 3, we proposed a JM framework which involves repeated measurements of self-monitored glucose measurements over the entire pregnancy with a binary outcome, PB, using the aforementioned T1DM cohort. In Chapter 4, we introduced a novel MVJM approach under a Bayesian framework to estimate the association between BG level and risk of PB while accounting for intermittent missing pattern in the former in women with T1DM. Additionally, we compared our developed MVJM approach with other commonly used techniques, such as TS, to determine the predictive ability of our novel MVJM.

5.2 Discussion

The semiparametric regression approach (Chapter 2) improved our understanding of velocity, acceleration and deceleration of maternal glycemic profiles with respect to pregnancy outcomes. The estimation of derivatives provided more accurate information on periods in which glucose has higher variation, compared to use of summary measures. An interesting feature of acceleration/deceleration in BG levels was detecting gestation specific recurrence of risk of

developing hypoglycemia associated with the mother's glycemic profiles. Our approach showed that monitoring at both the early stage of pregnancy and at the end stage of pregnancy, is very important, especially to minimize the risk of an LGA birth. Such information was not evident from any of the traditional glucose variability indices. The estimates could serve as a prognostic aid to monitor glucose across the gestational period, thus increasing patients' and clinicians' abilities to take timely actions in a more informed manner, and hence improving pregnancy outcomes. In Chapter 3, we illustrated that a JM approach was a better model choice compared to popular epidemiologic methods, such as the TDCM, in providing additional insights and characterizing the association between a binary endpoint, risk of PB, and a longitudinal process, glucose measurements over the entire pregnancy in T1DM cohort. On testing various parametrizations of association structures between the two processes, we showed that the AUC association provided the best fit for the JM. By accurately characterizing association between BG level and PB, this model could be used to aid clinician decision making in estimating risk of preterm delivery. The clinical application could enable timelier decision making based on BG levels and baseline covariates/risk factors included in the model. In Chapter 4, we extended the univariate JM to an MVJM framework involving mixed outcomes by separately modeling the intermittent missing pattern coming from IOGS per day over the entire pregnancy in the aforementioned T1DM cohort. We defined three sub-models to illustrate each process, a continuous longitudinal process (glucose measurements), a count process (IOGS per day), and a binary event process (PB) and joined them by a shared random effects framework. By simultaneously estimating all three outcomes, our proposed MVJM yielded better predictive ability compared to popular approach, TS regression technique. Within a Bayesian framework, the MVJM approach increased our epidemiologic understanding and was able to address fundamental

analysis flaws that currently hinder perinatal research and other areas of chronic disease research that rely on feedback from medical monitoring data.

An important aspect of JM is the conditional independence assumption. That is, given the random effects the longitudinal outcomes are independent of the event outcomes process. Another key feature of JM is the correct specification of association structures between the correlated outcomes of interest. However, the choice of different parameterizations of association structures should be based on the clinical research question of interest.

5.3 Future Research

All of the above-mentioned studies can be extended for future research. For our semiparametric regression approach (Chapter 2), future work could involve estimation of fixed and time-varying predictors on the glucose readings to explain causal inference and improve our understanding of fluctuation in maternal glucose profile on pregnancy outcome. A time-dependent association structure could be introduced for the JM approach discussed in Chapter 3. For Chapter 4, we would like to test different parameterizations of association structures, such as slope or AUC, under the MVJM framework. Additionally, we would like to introduce dynamic prediction to estimate risk of PB and illustrate how it can be dynamically updated, given that the subject-specific glycemic profiles were available up to the time of prediction.

References

1. Kirk, J. K., & Stegner, J. (2010). Self-Monitoring of Blood Glucose: Practical Aspects. *Journal of Diabetes Science and Technology*, 4(2), 435–439.
2. Deiss D.; Bolinder J.; Riveline J-P.; Battelino T.; Bose E.; Tubiana-Rufi N.; Kerr D.; Phillip M. (2006). Improved glycemic control in poorly controlled patients with type 1 diabetes using real-time continuous glucose monitoring. *Diabetes Care*, 29 (12): 2730–2732.
3. Kaufman FR, Gibson LC, Halvorson M, Carpenter S, Fisher LK, Pitukcheewanont P (2001). A pilot study of the continuous glucose monitoring system: clinical decisions and glycaemic control after its use in pediatric type 1 diabetic subjects. *Diabetes Care*, (12):2030–2034.
4. Schaepelynck-Bélicar P, Vague P, Simonin G, Lassmann-Vague V (2003). Improved metabolic control in diabetes adolescents using the continuous glucose monitoring system (CGMS) *Diabetes Metab*, 29(6):608–612.
5. Halvorson M, Carpenter S, Kaiserman K, Kaufman FR (2007). A pilot trial in pediatrics with the sensor-augmented pump: combining real-time glucose monitoring with the insulin pump. *J Pediat*, 150(1):103–105.
6. Buckingham B, Beck RW, Tamborlane WV, Xing D, Kollman C, Fiallo-Scharer R, Maurus N, Ruedy KJ, Tansey M, Weinzimer SA, Wysocki T (2007). Continuous glucose monitoring in children with type 1 diabetes. *Journal of Pediatrics*, 151(4):388–393. Diabetes Research in Children Network (DirecNet) Study Group.
7. Kovatchev B, Clarke W (2007). Continuous glucose monitoring (CGM) reduces risks for hypo- and hyperglycemia and glucose variability in diabetes. *Diabetes*, 56 (Suppl 1):A23.

8. Itai Bar-Hava H, Barnhard Y, Scarpelli SA, Orvieto R, Ben-Rafaelb Z, Divan MY (1997). Gestational diabetes and preterm labor: is glycemic control a contributing factor?. *European Journal of Obstetrics & Gynecology and Reproductive Biology*, 73:111 – 114.
9. Hickey GL, Philipson P, Jorgensen A, Ruwanthi-Kolamunnage D (2016). Joint modelling of time-to-event and multivariate longitudinal outcomes: recent developments and issues. *BMC Medical Research Methodology*, 16:117.
10. Albert PS, and Shih JH (2010). An approach for jointly modeling multivariate longitudinal measurements and discrete time-to-event data. *The Annals of Applied Statistics*, 4(3): 1517-1532.
11. Hanson U, Persson B (1993). Outcome of pregnancies complicated by type 1 insulin-dependent diabetes in Sweden: acute pregnancy complications, neonatal mortality and morbidity. *Am J Perinatal*, 10:330– 333.
12. Casson IF, Clarke CA, Howard CV, McKendrick O, Pennycook S, Pharoah PO, Platt MJ, Stanistreet M, van Velszen D, Walkinshaw S (1997). Outcomes of pregnancy in insulin dependent diabetic women: results of a five year population cohort study. *BMJ*, 315:275–278.
13. Silverman BL, Rizzo T, Green OC, Cho NH, Winter RJ, Ogata ES, Richards GE, Metzger BE (1991). Long-term prospective evaluation of offspring of diabetic mothers. *Diabetes* 40(Suppl. 2), 121–125.
14. Kahn SE (2003). The relative contributions of insulin resistance and beta-cell dysfunction to the pathophysiology of type 2 diabetes. *Diabetologia*, 46 : 3 –19.
15. Fox CS, Coady S, Sorlie PD, Levy D, Meigs JB, et al (2004). Trends in cardiovascular complications of diabetes. *JAMA*, 292: 2495-2499.
16. McElvy SS, Miodovnik M, Rosenn B, Khoury JC, Siddiqi T, Dignan PS, Tsang RC (2000). A focused preconceptional and early pregnancy program in women with type 1 diabetes reduces

perinatal mortality and malformation rates to general population levels . *J Matern Fetal Med*, 9:14–20.

17. Kovilam, O., Khoury, J., Miodovnik, M., Chames, M., Spinnoto, J., & Sibai, B (2002). Spontaneous preterm delivery in the type 1 diabetic pregnancy: the role of glycemic control . *J Matern Fetal Neonatal Me*, 11(4), 245-248.

18. Langer O., Mazze R (1988). The relationship between large-for-gestational-age infants and glycemic control in women with gestational diabetes (1988). *Am J Obstet Gynecol*, 159(6):1478-83.

19. Clausen TD, et al (2008). High prevalence of type 2 diabetes and pre-diabetes in adult offspring of women with gestational diabetes mellitus or type 1 diabetes. *Diabetes Care*, 31(2): 340-346.

20. Kilpatrick ES, Rigby AS, Goode K, Atkin SL (2007). Relating mean blood glucose and glucose variability to the risk of multiple episodes of hypoglycaemia in type 1 diabetes. *Diabetologia*, 50:2553–2561.

21. Hirsch Irl B (2015). Glycemic Variability and Diabetes Complications: Does It Matter? Of Course It Does! .*Diabetes Care*, 38:1610–1614.

22. Service FJ, Molnar GD, Rosevear JW, Ackerman E, Gatewood LC, Taylor WF (1970). Mean amplitude of glycemic excursions, a measure of diabetic instability. *Diabetes*, 19:644–655.

23. Froslic KF, Roislien Jo, Qvigstad E, Godang K, et al (2013). Shape information from glucose curves: Functional data analysis compared with traditional summary measures. *BMC Medical Research Methodology*, 13:6; DOI: 10.1186/1471-2288-13-6.

24. Szczesniak RD, Li D, Duan LL, Altaye M, Miodovnik M, Khoury JC (2016). Longitudinal Patterns of Glycemic Control and Blood Pressure in Pregnant Women with Type 1 Diabetes Mellitus: Phenotypes from Functional Data Analysis. *Am J Perinatol*, 33(13):1282-1290.

25. Vandyke RD, Ren Y, Sucharew HJ, Miodovnik M, Rosenn B, Khoury JC (2012). Characterizing maternal glycemic control: a more informative approach using semiparametric regression. *J Matern Fetal Neonatal Med*, 25(1):15–19.
26. Sparacino G, Facchinetti A, Maran A, Cobelli C (2008). Continuous glucose monitoring time series and hypo/hyperglycemia prevention: requirements, methods, open problems. *Curr Diabetes Rev*, 4(3):181–92.
27. Monnier L, Mas E, Ginet C, et al (2006). Activation of oxidative stress by acute glucose fluctuations compared with sustained chronic hyperglycemia in patients with type 2 diabetes. *JAMA*, 295:1681–1687.
28. Ruppert D, Wand MP, Carroll RJ (2003). Semiparametric Regression. Cambridge series in statistical and probabilistic mathematics. Cambridge, NY: Cambridge University Press.
29. Ngo, L., & Wand, M. P. Smoothing with Mixed Model Software (2004). *Journal of Statistical Software*, 9(1), 1-54. Retrieved from <http://ro.uow.edu.au/eispapers/2649/>.
30. Littell RC, Milliken GA, Stroup WW, Wolfinger RD, Schabenberger O. SAS for Mixed Models, 2nd Edition. SAS Institute, Inc, Cary, NC, USA; 2006.
31. Chen R, Yogeve Y, Ben-Haroush A, Jovanovic L, Hod M, Phillip M (2003). Continuous glucose monitoring for the evaluation and improved control of gestational diabetes mellitus. *J Matern Fetal Neonatal Med*, 14(4):256-60.
32. Kirk, J. K., & Stegner, J. (2010). Self-Monitoring of Blood Glucose: Practical Aspects. *Journal of Diabetes Science and Technology*, 4(2), 435–439.
33. Langer O, Conway DL (2000). Level of glycemia and perinatal outcome in pregestational diabetes. *J Matern Fetal Med*, 9:35–41.

34. Deiss D.; Bolinder J.; Riveline J-P.; Battelino T.; Bose E.; Tubiana-Rufi N.; Kerr D.; Phillip M. (2006). Improved glycemic control in poorly controlled patients with type 1 diabetes using real- time continuous glucose monitoring. *Diabetes Care*, 29 (12): 2730–2732.
35. Eidem, I., Vangen, S., Hanssen, K.F. et al. (2011). Perinatal and infant mortality in term and preterm births among women with type 1 diabetes. *Diabetologia*, 54: 2771.
36. Lepercq J, Coste J, Theau A, Dubois-Laforgue D, Timsit T (2004). Factors Associated With Preterm Delivery in Women With Type 1 Diabetes. *Diabetes Care*, Vol:27(12).
37. Deryabina EG, Yakornova GV, Pestryaeva LA, Sandyreva ND (2016). Perinatal outcome in pregnancies complicated with gestational diabetes mellitus and very preterm birth: case–control study. *Gynecological Endocrinology*, 32(S2): S52–S55.
38. Lin SF, Kuo CF, Chiou MJ, Chang SH (2017). Maternal and fetal outcomes of pregnant women with type 1 diabetes, a national population study. *Oncotarget*, Vol. 8, (46), pp: 80679-;8068.
39. Gupta R, Khoury J, Altaye M, Dolan L, Szczesniak RD (2017). Glycemic Excursions in Type 1 Diabetes in Pregnancy: A Semiparametric Statistical Approach to Identify Sensitive Time Points during Gestation. *Journal of Diabetes Research*, Article ID 2852913.
40. Andersen PK, and, Gill RD (1982). Cox's Regression Model for Counting Processes: A Large Sample Study. *The Annals of Statistics*, Vol. 10, No. 4, pp. 1100-1120.
41. Prentice R (1982). Covariate measurement errors and parameter estimation in a failure time regression model. *Biometrika*, 69(2):331–42.
42. Kalbfleisch, J.D. and Prentice, R.L. (2002) *The Statistical Analysis of Failure Time Data*. 2nd Edition, John Wiley and Sons, New York.

43. Rizopoulos D (2012). Joint Models for Longitudinal and Time-to-Event Data, with Applications in R. Boca Raton: Chapman & Hall/CRC.
44. Rizopoulos D, Lesare E (2014). Introduction to the special issue on joint modelling techniques. *Statistical Methods in Medical Research*, 23, 3-10.
45. Wulfsohn, M. and Tsiatis, A. (1997). A joint model for survival and longitudinal data measured with error. *Biometrics*, 53:330 – 339.
46. Tsiatis, A. A. and Davidian, M. (2004). Joint modeling of longitudinal and time-to-event data: An overview. *Statistica Sinica*, 14, 809–834.
47. DeGruttola, V. and Tu, X. (1994). Modeling progression of CD-4 lymphocyte count and its relationship to survival time. *Biometrics*, 50:1003 – 1014.
48. Wang, Y. and Taylor, J. (2001). Jointly modeling longitudinal and event time data with application to acquired immunodeficiency syndrome. *JAMA*, 96:895 – 905.
49. Henderson, R., Diggle, P., and Dobson, A. (2000). Joint modelling of longitudinal measurements and event time data. *Biostatistics*, 1:465 – 480.
50. Brown, E. and Ibrahim, J. (2003). A Bayesian semiparametric joint hierarchical model for longitudinal and survival data. *Biometrics*, 59:221 – 228.
51. Rizopoulos, D. and Ghosh, P. (2011). A Bayesian semiparametric multivariate joint model for multiple longitudinal outcomes and a time-to-event. *Statistics in Medicine*, 30:1366 – 1380.
52. Wu L, Liu W, Grace Yi, Huang Y (2012). Analysis of Longitudinal and Survival Data: Joint Modeling, Inference Methods, and Issues. *Journal of Probability and Statistics*, vol. 2012, Article ID 640153.

53. Ye W, Lin XH, Taylor JMG (2008). Semiparametric modeling of longitudinal measurements and time-to-event data – a two-stage regression calibration approach. *Biometrics*, 64:1238–1246.
54. Huang, X., Li, G., Elashoff, R., and Pan, J. (2011). A general joint model for longitudinal measurements and competing risks survival data with heterogeneous random effects. *Lifetime Data Analysis*, 17:80 – 100.
55. Yu, M., Law, N., Taylor, J., and Sandler, H. (2004). Joint longitudinal-survival-cure models and their application to prostate cancer. *Statistica Sinica*, 14:835 – 862.
56. Rizopoulos, D. (2011). Dynamic predictions and prospective accuracy in joint models for longitudinal and time-to-event data. *Biometrics*, 67:819 – 829.
57. McElvy SS, Miodovnik M, Rosenn B, Khoury JC, Siddiqi T, Dignan JP, Tsang R (2000). A focused preconceptional and early pregnancy program in women with type 1 diabetes reduces perinatal mortality and malformation rates to general population levels. *The Journal of Maternal-Fetal Medicine*, 9:14–20.
58. Gelfand A, Dey D. (1994). Bayesian model choice: Asymptotics and exact calculations. (Series B). *Journal of the Royal Statistical Society*, 56:501–514.
59. Spiegelhalter, D. J., Best, N., Carlin, B., and van der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society, Series B*, 64: 583–640.
60. S. Li1, M. Zhang, H. Tian, Z.Liu4,X.Yin, B. Xi (2014). Preterm birth and risk of type 1 and type 2 diabetes: systematic review and meta-analysis.
61. Eidem, I., Vangen, S., Hanssen, K.F. et al. (2011). Perinatal and infant mortality in term and preterm births among women with type 1 diabetes. *Diabetologia*, 54: 2771.

62. Lepercq J, Coste J, Theau A, Dubois-Laforgue D, Timsit T (2004). Factors Associated With Preterm Delivery in Women With Type 1 Diabetes. *Diabetes Care*, Vol:27(12).
63. Feldman, A. Z., & Brown, F. M. (2016). Management of Type 1 Diabetes in Pregnancy. *Current Diabetes Reports*, 16, 76.
64. Garrett M. Fitzmaurice, Nan M. Laird, James H. Ware. Applied Longitudinal Analysis, 2nd Edition. Wiley, New York.
65. Lederar DJ, Bell SC, Brandosn RD et al. (2019). Control of Confounding and Reporting of Results in Causal Inference Studies. Guidance for Authors from Editors of Respiratory, Sleep, and Critical Care Journals. *Annals of the American Thoracic Society*, 16(1): 22-28.
66. Molenberghs, Geert; Fitzmaurice, Garrett; Kenward, Michael G.; Tsiatis, Anastasios; Verbeke, Geert, eds. (2015), *Handbook of Missing Data Methodology*, Chapman & Hall.
67. Kirk, J. K., & Stegner, J. (2010). Self-Monitoring of Blood Glucose: Practical Aspects. *Journal of Diabetes Science and Technology*, 4(2), 435–439.
68. Deiss D.; Bolinder J.; Riveline J-P.; Battelino T.; Bose E.; Tubiana-Rufi N.; Kerr D.; Phillip M. (2006). Improved glycemic control in poorly controlled patients with type 1 diabetes using real- time continuous glucose monitoring. *Diabetes Care*, 29 (12): 2730–2732.
69. Kaufman FR, Gibson LC, Halvorson M, Carpenter S, Fisher LK, Pitukcheewanont P (2001). A pilot study of the continuous glucose monitoring system: clinical decisions and glycaemic control after its use in pediatric type 1 diabetic subjects. *Diabetes Care*, (12):2030–2034.
70. Schaepelynck-Bélicar P, Vague P, Simonin G, Lassmann-Vague V (2003). Improved metabolic control in diabetes adolescents using the continuous glucose monitoring system (CGMS) *Diabete Metab*, 29(6):608–612.

71. Halvorson M, Carpenter S, Kaiserman K, Kaufman FR (2007). A pilot trial in pediatrics with the sensor-augmented pump: combining real-time glucose monitoring with the insulin pump. *J Pediatr*,150(1):103–105.
72. Buckingham B, Beck RW, Tamborlane WV, Xing D, Kollman C, Fiallo-Scharer R, Maurus N, Ruedy KJ, Tansey M, Weinzimer SA, Wysocki T (2007). Continuous glucose monitoring in children with type 1 diabetes. *J Pediatr*, 151(4):388–393. Diabetes Research in Children Network (DirecNet) Study Group.
73. Eidem, I.,Vangen, S., Hanssen, K.F. et al. (2011).Perinatal and infant mortality in term and preterm births among women with type 1 diabetes. *Diabetologia*, 54: 2771.
74. Lepercq J, Coste J, Theau A, Dubois-Laforgue D, Timsit T (2004). Factors Associated With Preterm Delivery in Women With Type 1 Diabetes. *Diabetes Care*, Vol:27(12).
75. Feldman, A. Z., & Brown, F. M. (2016). Management of Type 1 Diabetes in Pregnancy. *Current Diabetes Reports*, 16, 76.
76. Gupta R, Khoury J, Altaye M, Dolan L, Szczesniak RD (2017). Glycemic Excursions in Type 1 Diabetes in Pregnancy: A Semiparametric Statistical Approach to Identify Sensitive Time Points during Gestation. *Journal of Diabetes Research*, Article ID 2852913.
77. Lipsitz, S. R., Fitzmaurice, G. M., Ibrahim, J. G., Gelber, R. and Lipshultz, S. (2002). Parameter estimation in longitudinal studies with outcome-dependent follow-up. *Biometrics*, **58**,621–630.
78. Robins, J. M., and Rotnitzky, A (1995). Semiparametric Efficiency in Multivariate Regression Models with Missing Data. *Journal of the American Statistical Association*,90:122–129.
79. Robins, J. M., Rotnitzky, A. and Zhao, L. P (1994). Estimation of regression coefficients when some regressor are not always observed. *Journal of the American Statistical Association*, 89, 846–866.

80. Lin H, Scharfstein D, Rosenheck R (2004). Analysis of longitudinal data with irregular, outcome-dependent follow-up. *Journal of the Royal Statistical Society, Series B*; 66:791–813.
81. Diggle, Peter J.; Patrick Heagerty; Kung-Yee Liang; Scott L. Zeger (2002). Analysis of Longitudinal Data. Oxford Statistical Science Series.
82. Molenberghs, G., Kenward, M., Verbeke, G., Birhanu, T (2011). Pseudo-likelihood estimation for incomplete data. *Statistica Sinica*, 21(1), 187.
83. Chen Y, Jing N, Chunyan C (2015). Regression analysis of longitudinal data with irregular and informative observation times. *Biostatistics*, 16, 4, pp. 727–739.
84. Rizopoulos D (2012). Joint Models for Longitudinal and Time-to-Event Data, with Applications in R. Boca Raton: Chapman & Hall/CRC.
85. Rizopoulos D, Lesare E (2014). Introduction to the special issue on joint modelling techniques. *Statistical Methods in Medical Research*, 23, 3-10.
86. Hickey GL, Philipson P, Jorgensen A, Ruwanthi-Kolamunnage D (2016). Joint modelling of time-to-event and multivariate longitudinal outcomes: recent developments and issues. *BMC Medical Research Methodology*, 16:117.
87. Albert PS, and Shih JH (2010). An approach for jointly modeling multivariate longitudinal measurements and discrete time-to-event data. *The Annals of Applied Statistics*, 4(3): 1517-1532.
88. Faucett, C. and Thomas, D. (1996). Simultaneously modelling censored survival data and repeatedly measured covariates: A Gibbs sampling approach. *Statistics in Medicine*, 15:1663 – 1685.
89. Lunn, D.; Spiegelhalter, D.; Thomas, A.; Best, N. (2009). "The BUGS project: Evolution, critique and future directions". *Statistics in Medicine*, 28 (25): 3049–3067.

90. McElvy SS, Miodovnik M, Rosenn B, Khoury JC, Siddiqi T, Dignan JP, Tsang R (2000). A focused preconceptional and early pregnancy program in women with type 1 diabetes reduces perinatal mortality and malformation rates to general population levels. *The Journal of Maternal-Fetal Medicine*, 9:14–20.
91. Ngo, L., & Wand, M. P (2004). Smoothing with Mixed Model Software. *Journal of Statistical Software*, 9(1), 1-54. Retrieved from <http://ro.uow.edu.au/eispapers/2649/>.
92. J. Geweke (1992). *Evaluating the accuracy of sampling-based approaches to calculating posterior moments*, *Bayesian Statis*, (Bernardo, J. M, Berger, J.O).
93. P.S. Albert and J.H. Shi (2010). On estimating the relationship between longitudinal measurements and time-to-event data using a simple two-stage procedure, *Biometrics*, 66: pp. 983–987.
94. W.Ye, X. Lin, and J.M.G. Taylor (2008). Semiparametric modeling of longitudinal measurements and time-to-event data: a two stage regression calibration approach, *Biometrics*, 64: pp. 1238–1246.
95. D.W. Hosmer and S. Lemeshow (2000). *Applied Logistic Regression*, JohnWiley & Sons, NewYork.
96. Fitzmaurice, G. M., Laird, N. M. and Shneyer, L (2001). An alternative parameterization of the general linear mixture model for longitudinal data with non-ignorable drop-outs. *Statistics in Medicine*, 20, 1009–1021.
97. Little, R. and Rubin, D (2002). *Statistical Analysis with Missing Data*. Wiley, New York, 2nd edition.