Evaluating Hypoglycemic Risk of Two Basal Insulin Delivery Methods for Type 1 Diabetes: A Comparison of Tresiba Insulin against Continuous Subcutaneous Insulin Infusion using Data from Continuous Glucose Monitors

By

Benjamin Lane

Thesis

Submitted to the Faculty of the

Graduate School of Vanderbilt University

In partial fulfillment of the requirements

for the degree of

Master of Science

In

Biostatistics

March 31, 2020

Nashville, Tennessee

Approved

Jonathan Schildcrout, PhD

Bryan Shepherd, PhD

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Clinical Introduction

Type 1 Diabetes and Insulin Management

Type 1 diabetes mellitus is a form of diabetes that results from an autoimmune attack on the insulin-producing beta cells in the pancreas. The exact cause of this autoimmune attack is not known, though it is generally thought to be a result of environmental factors acting on a genetically susceptible individual. The loss of the beta cells results in the failure to produce enough insulin required to keep blood sugar levels appropriately regulated. Insulin is a hormone that allows sugar, primarily in the form of glucose, to enter the cells in the body. Without insulin, blood sugar levels rise, which can lead to both microvascular (diabetic retinopathy, neuropathy, and nephropathy) and macrovascular (coronary artery disease, peripheral arterial disease, and stroke) complications, including death (Fowler). Currently, there is no cure for type 1 diabetes, so the disease must be managed by replacement of insulin. Controlling blood sugar levels by supplementing insulin is known as insulin man- agement. Typically, this is accomplished in one of two ways: insulin is either injected multiple times per day (e.g. before meals and at bedtime), or it is continually infused at varying rates. For individuals using multiple daily injections (MDI), two types of insulin are needed. First, the individuals need a long acting insulin, called basal insulin, which releases over the course of the day. This helps to keep blood sugars from naturally rising during fasting states. The second insulin needed is a short acting insulin, known as bolus insulin, which is used to correct high blood sugars and cover mealtime insulin requirements. For individuals using continuous insulin infusion (specifically, continuous subcutaneous

insulin infusion, CSII), only short acting insulin is required. Perhaps surprisingly, one of the primary concerns with insulin replacement therapy is minimizing low blood sugars, hypoglycemia. A low blood sugar occurs when the glucose concentration drops below a specific threshold. Common threshold values include 70 mg/dL and 54 mg/dL. The adverse effects are variable, but they can range from hunger, dizziness, and confusion to seizures, coma, and death. (Cryer 2007; Weinstock 2019; Fowler 2008)

Risk of Hypoglycemia

Once individuals are using insulin replacement therapy, one of the most important clinical considerations is minimizing time spent in low blood sugars ranges. The recommended target for good glycemic control is an average glucose of 154 mg/dL, so individuals must inject enough insulin to prevent blood sugars from rising, but not so much that they drop their blood glucose too low. This balancing act is difficult because a host of factors affect blood sugars on a daily basis, e.g. food, exercise, sleep, stress, medications, etc. Further, insulin always reduces blood sugar levels, so hyperglycemia could safely and easily be avoided if risk of hypoglycemia were not a concern. Unfortunately, not only do low blood sugars cause long term risk, but they are also of immediate concern because they can quickly cause seizures and death. As an example, consider an individual at a safe blood glucose level: 80 mg/dL. Increasing their blood glucose by 50 mg/dL puts them just inside the two-hour postprandial target range, which is defined by the American Diabetes Association to be less than 140 mg/dL. However, a 50 mg/dL drop puts them at a blood glucose of 30 mg/dL, where they'd be experiencing functional brain failure and be dangerously close neuronal brain death (Cryer 2007).

Per the American Diabetes Association (ADA) Workgroups, a value of 70 mg/dL is meant to approximate the lower limit of the physiologic fasting non-diabetic blood glucose range, and is further defined to be an "alert" value that forewarns the patient of possibly developing clinically relevant hypoglycemia (Weinstock 2019).

Insulin pump therapy (CSII)

An insulin pump is a device that is used for administering insulin continuously at user-specified rates. The slow, continuous rate is what replaces the basal insulin for injection therapy. Because the insulin pump is administering insulin continually, it must be worn at all times, but it also offers greater flexibility than multiple daily injections due to the ability to vary the basal rate throughout the day. For example, users can administer less insulin at nighttime than during daytime if they fear nocturnal hypoglycemic events. Challenges of insulin pump therapy include, but are not limited to, optimizing insulin pump settings, affording the device and all necessary supplies, and taking care not to damage the device during recreational activities. Pickup 2002; Schaepelynck et al. 2011

Continuous Glucose Monitoring (CGM)

Traditionally, people with type 1 diabetes check their blood sugars by pricking the tips of their fingers and applying a droplet of blood to a test strip, which is inserted into a blood sugar meter. Blood sugar checks are typically performed in the morning, before each meal, and at bedtime. Checking at 3:00 a.m. is also common to assess nocturnal values. A continuous glucose monitor (CGM) is a device that checks an individual's blood sugar every

five minutes. The device is worn under the skin, similar to an insulin pump, and the blood sugar values are wirelessly communicated to a transmitter or smart phone, where users can view their blood sugar value in real time (within 5 minutes). Under ideal circumstances, continuous glucose monitors therefore provide 288 blood sugar readings per day. While CGMs offer convenience to users who do not wish to prick their fingers continuously, they are required to be worn at all times and can be cost prohibitive. (Battelino et al. 2019; Rodbard 2017; Thomas, Revital, Tadej, et al. 2017)

1.2 Study Motivation and Objectives

For type 1 diabetes management, the "gold standard" of care for achieving optimal glucose control and minimizing adverse health-related outcomes has come from administering insulin via continuous subcutaneous insulin infusion (CSII), also known as insulin pump therapy. (Pickup 2002) The primary alternative to insulin pump therapy is to use multiple daily insulin injections (MDI) throughout the day. Numerous studies have confirmed the superiority of insulin pump therapy over MDI in terms of glycemic control (as measured by HbA1c), incidence of hypoglycemia, and quality of life (QOL) questionnaires. How- ever, many of the studies that affirmed the use of CSII over MDI compared the use of rapid-acting insulin analogs (e.g. insulins: lispro, aspart, glulisine) in the insulin pump to the first-generation analogue, u100 insulin glargine (Lantus), which was used as the basal insulin in the MDI study arms. With the development of newer, more stable (i.e. less variable) basal insulins, however, comes the need to reassess whether insulin pump therapy continues to be the "best" way to administer insulin to individuals with type 1 diabetes.

Insulin degludec (Tresiba), is a new, uniquely designed insulin that has desirable properties such as once-daily dosing which can be administered at any time and "highly predictable, gradual dissociation" (Tambascia and Eliaschewitz 2015), resulting in low intra-individual variation (Tambascia and Eliaschewitz 2015). Further, when compared to insulin glargine, it has shown reduced risk of nocturnal (midnight to 6:00am) hypoglycemia and non-inferiority in average glucose (measured by HbA1c) reduction. (Kalra and Gupta 2015). Therefore, Tresiba is an ideal basal insulin to use when comparing glucose control between insulin pump and injectable insulin therapies.

In addition to the development of newer insulins, however, came the advancement of the available technology for monitoring blood sugar values throughout the day. Historically, studies rely on fingerstick data, whereby study participants actively check and record their blood sugar (ideally) when they first wake up, before and after each meal, before they go to sleep, and at 3:00 a.m. Beyond the impracticalities and difficulties of this for study participants in an out-patient setting, this small amount of information is far from ideal. However, with continuous glucose monitors (CGMs), study participants have their blood sugars recorded every five minutes for as long as they wear the device. Thus, the available data go from eight actively recorded, unevenly spaced blood sugar readings to 288 passively recorded, evenly spaced blood sugar readings per day. This massive amount of additional data offers a much clearer picture of what a study participant's blood sugar is doing over 24 hours. An additional benefit of the quantity of information comes with the fact that CGMs passively monitor blood glucose levels. This makes them especially useful for tracking

hypoglycemia because subjects who do not immediately feel the effects of low blood sugar values (i.e. individuals who are "hypoglycemic unaware") might frequently miss recording if they had to actively monitor with finger sticks. Alternatively, participants who have frequent lows are saved the inconvenience of constantly 'sticking' their finger. Ultimately, this leads to more accurate data capture, and an ideal tool for examining hypoglycemia in a population.

Taking advantage of advancements in insulin therapy and continuous glucose monitoring, this study sought to compare two different insulin treatment regimens. The first treatment is standard insulin pump therapy using Novolog for both basal and bolus insulin, and the second treatment administers Novolog via insulin pump for boluses but uses Tresiba (which must be injected) as the basal insulin. The study was an investigator initiated trial written by Mountain Diabetes and Endocrine Center out of Asheville, NC. The study is a randomized, cross-over, open label, single-center trial where subjects were randomized to treatment order. They remained on each treatment for twenty weeks. The first four weeks consisted of an insulin optimization period to ensure correct insulin dosing, then a 14 week maintenance period to ensure stable insulin doses, and finally concluded with a two week observation period. Subjects then crossed-over to the second treatment and repeated the twenty week process.

Because a primary limiting factor in insulin management is hypoglycemia, it is therefore of great clinical import to understand how different insulin regimens can affect the risk and duration of hypoglycemia in a population of interest. That is the focus of this analysis.

Given Tresiba's impressive reduction in hypoglycemia as compared to older insulins, the study team seeks to compare the 24 hour risk profile of developing hypoglycemia using Tresiba for basal insulin treatment versus using insulin pump therapy. Additionally, because so many different factors affect blood glucose, many of which cannot be easily quantified or controlled for (e.g. stress, eating habits, quality of sleep, etc.), understanding how latent variables and unmeasured confounding affect the analysis becomes essential context for interpreting study results. So, the team seeks to build this context by simulating populations with different amounts of unmeasured variability in their hypoglycemic risk profiles. Lastly, because many blood glucose measurements per subject are taken, the study data are correlated within a subject. Therefore, the final question this analysis seeks to answer is how correlation affects the study results. Specifically, using a cut point of 70 mg/dL as a threshold value, the study team aims to:

1. Describe hypoglycemic risk over the course of the day in each treatment arm, individually.
2. To compare CSII and Tresiba risk profiles over the course of the day and at crucial time periods of clinical interest.
3. To calculate model summaries of the analyses that are of clinical interest.
4. To explore the impact that population heterogeneity has on study summaries via simulation.
5. To examine the impact of dependence structure misspecification on statistical inferences.

Chapter 2

Mathematical Review

Before moving on to describe the study data and statistical analysis in detail (Chapter 3: Materials and Methods), this chapter serves as a mathematical introduction to the type of statistical analysis the study team uses. As mentioned at the end of chapter 1, the study team assesses hypoglycemia at a cut-point of 70 mg/dL. This has the effect of dichotomizing the data, making it fit a binomial distribution. Further, because blood glucose is measured mul-tiple times per day, the data have repeated measurements that make the values correlated.

2.1 GLMMs for Binomial Data

For binomial data, a generalized linear model with a logit link function is appropriate to use for making inferences. Generalized linear models assume that the dependent random variable, Y_i, is part of an exponential family with mean μ_i. A generalized linear mixed effects model does the same, but is characterized by random (subject-specific) effects in addition to fixed (population level) effects. Just as in generalized linear regression, responses are not modeled directly, but instead are modeled via a link function. (Everitt and Howell 2005; N. E. Breslow and Clayton 1993) Therefore, a generalized linear mixed effects model with a logit link function can be described as:

$$\text{logit}(\mu_i) = \log\left(\frac{\mu_i}{1-\mu_i}\right) = \boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{Z}_i\boldsymbol{\gamma}_i \qquad (2.1)$$

where:

$$\begin{aligned}
\boldsymbol{\beta} &= (\beta_1, \beta_2, \ldots, \beta_p)^T && \text{Fixed effects} \\
\boldsymbol{X}_i &= (\boldsymbol{x}_{i1}, \boldsymbol{x}_{i2}, \ldots, \boldsymbol{x}_{im})^T && \text{Fixed effects design matrix} \\
\boldsymbol{\gamma} &= (\gamma_1, \gamma_2, \ldots, \gamma_q)^T && \text{Random effects} \\
\boldsymbol{Z}_i &= (\boldsymbol{z}_{i1}, \boldsymbol{z}_{i2}, \ldots, \boldsymbol{z}_{im})^T && \text{Random effects design matrix}
\end{aligned} \tag{2.2}$$

and subjects are indexed from $i = 1, \ldots, n$. Assuming an equal number of observations per subject, the observations within a subject are indexed as $j = 1, \ldots, m$, the covariates are indexed from $k = 1, \ldots, p$, and the random effects are $k^* = 1, \ldots, q$ where $q \leq p$.

Further assumptions are $\gamma_i \sim N_q(\mathbf{0}, \boldsymbol{G})$, and the response variables, $\boldsymbol{Y}_i = (Y_{i1}, \ldots, Y_{im})^T$ are independent and follow a Binomial distribution:

$$Y_{ij} | \beta, \gamma_i \sim \text{Bernoulli}(\mu_{ij}) \tag{2.3}$$

for $i = 1, \ldots, n$ and $j = 1, \ldots, m$.

The expected value and variance are similarly conditional on the random effects:

$$E[Y_{ij} \,|\, \gamma_i] = \mu_{ij} \quad \text{and} \quad \text{Var}[Y_{ij} \,|\, \gamma_i] = \phi \mu_{ij}(1 - \mu_{ij}) \tag{2.4}$$

where $\phi > 0$ is a scale parameter.

Then, for maximum likelihood estimation, the random effects are treated as unobserved

nuisance parameters, and integrating over their assumed distribution yields the marginal likelihood for $\boldsymbol{Y}_i$. Letting α parameterize the covariance matrix $\boldsymbol{G}$ for $\gamma_i \sim N_q(\boldsymbol{0}, \boldsymbol{G})$, the likelihood is a mixture distribution of the product of m bernoulli trials per subject and the multivariate normal distribution of the random effect, which is then multiplied across n subjects:

$$\begin{aligned}\mathscr{L}_Y(\beta,\alpha) &= \prod_{i=1}^{n} \int f_{Y|\gamma}\big(\boldsymbol{Y}_i|\gamma_i,\beta\big) \times f_\gamma\big(\gamma_i|\boldsymbol{G}\big)\, d\gamma_i \\ &= \prod_{i=1}^{n} \int \prod_{j=1}^{m_i} \mu_{ij}^{Y_{ij}} (1-\mu_{ij})^{1-Y_{ij}} \times (2\pi)^{-1/2} |\boldsymbol{G}|^{-m/2} \exp\left(\frac{-1}{2}\gamma_i \boldsymbol{G}^{-1} \gamma_i^T\right) d\gamma_i \end{aligned} \tag{2.5}$$

To perform the integration required in maximum likelihood estimation, however, requires algorithms that effectively estimate the solution because it does not exist in closed form. Analytical approximations includes penalized quasi-likelihood (PQL) while numerical integration can be achieved via Gaussian quadrature. (Wood 2017; Pinheiro and Chao 2006; N. Breslow 2004)

2.2 Correlation Structure

Because the random effects are assumed to be perturbations from the population level effects, they are typically not estimated. Instead, they're assumed to be normally distributed with mean 0 and variance-covariance matrix $\boldsymbol{G}$: $\boldsymbol{\gamma}_i \sim N(\mathbf{0}, \boldsymbol{G})$. The normality assumption makes intuitive sense because random effects can further be conceptualized as the sum of other factors that are unknown or cannot be measured. Then, because the fixed effect is the estimated population average, and the mean of the random effect is assumed to be 0, what's left to estimate is the variance-covariance matrix of the random effect, $\boldsymbol{G}$. $\boldsymbol{G}$ will be $q \times q$, where q is the number of random effects. For example, if a random intercept and slope is included in the model, then:

$$\begin{bmatrix} G_{11} & \\ G_{12} & G_{22} \end{bmatrix}$$

Here, $\sqrt{G_{11}}$ is the standard deviation in the level of the response, i.e. the amount of heterogeneity among subjects in the population that remains after the population level effects have been estimated. $\sqrt{G_{22}}$ is the standard deviation of the change of the response above and beyond the population level effects. Lastly, $\sqrt{G_{12}}$ is the correlation between the random intercept and slope. If $\sqrt{G_{12}} < 0$ then subjects with a higher level response have a lower rate of change. If $\sqrt{G_{12}} > 0$ then subjects with a higher level in the response have a higher rate of change. So, the matrix $\boldsymbol{G}$ quantifies the random variation in the response that occurs between subjects, but it doesn't capture the variation that occurs within subjects over time.

Variation within subjects over time is captured in the variance-covariance matrix of the

residuals. If a random intercept model is used, and if the assumption that there is no variation over time within subjects is valid (i.e. $Cov[\varepsilon_{ij}, \varepsilon_{ij'}] = 0$), then the within-subject correlation structure is exchangeable (also known as compound-symmetric). Specifically:

$$\begin{aligned} Cov[Y_{ij}, Y_{ij'}] &= Cov\big[E_\gamma[Y_{ij} \mid \gamma_{0i}], E_\gamma[Y_{ij} \mid \gamma_{0i}]\big] + E_\gamma\big[Cov[Y_{ij}, Y_{ij'}] \mid \gamma_{0i}\big] \\ &= Cov_\gamma[\boldsymbol{X\beta} + \gamma_{0i}, \boldsymbol{X\beta} + \gamma_{0i}] + E_\gamma[0] \\ &= Var_\gamma[\gamma_{0i}] + 0 = G_{11} \end{aligned} \tag{2.6}$$

Intuitively, this means that any for any two observations within a subject, irrespective of how closely related they are in time or some other distance measure, they will have the same correlation. If, however, this assumption is not valid, and instead assuming $Cov[\varepsilon_{ij}, \varepsilon_{ij'}] \neq 0$ is more appropriate, then the mixed effects model can be combined with a specification for serial dependence. That is, if the more appropriate assumption is that observations that are closer to each other in time, for example, are more correlated than observations that are farther away from each other, then that additional correlation can be captured and estimated. For a random intercept model, this is specified as:

$$Y_{ij} = \boldsymbol{x}_{ij}\boldsymbol{\beta} + \gamma_{0i} + W_i(t_{ij}) + \varepsilon_{ij} \tag{2.7}$$

Where $\gamma_{0i} \sim N(0, \sigma^2_\gamma)$ and is the random intercept for the ith subject, $W_i(t_{ij})$ is the serial dependence, and $\varepsilon_{ij} \sim N(0, \sigma^2_\varepsilon)$ is the measurement error.

There are numerous correlation structures that can be used for time-series modeling, but one common structure is auto-regressive correlation (AR(1)). Using a model with AR(1)

correlation structure assumes:

$$Cov[\varepsilon_{ij}, \varepsilon_{ij'}] = \sigma^2 \rho^{|t_{ij} - t_{ij'}|} \tag{2.8}$$

In other words, the covariance between the error terms at two different time points is proportional to how far apart in time (or other distance measure) those time points are. Under this model, the variance of the response for a specific subject and time has been further decomposed:

$$Var[Y_{ij}|\beta] = \sigma_\gamma^2 + \sigma_W^2 + \sigma_\varepsilon^2 = \sigma^2 \tag{2.9}$$

Where σ_γ^2 accounts for between subject variation, σ_W^2 accounts for within subject variation (in this case, with AR(1) correlation), and σ_ε^2 is the variation due to measurement error. Further, the covariance between two responses on the same subject is now:

$$\text{Cov}[Y_{ij}, Y_{ij'}] = \sigma_\gamma^2 + \sigma_W^2 \rho^{|t_{ij} - t)ij'|} \tag{2.10}$$

Then, letting $\boldsymbol{R}_i$ quantify random variation within subjects over time and $\rho(\alpha) = \alpha^{|j-j'|}$, the auto-regressive structure can be written as:

$$\boldsymbol{R}_i(\alpha) = \begin{bmatrix} 1 & \alpha & \alpha^2 & \dots & \alpha^{m-1} \\ \alpha & 1 & \alpha & \dots & \alpha^{m-2} \\ \alpha^2 & \alpha & 1 & \dots & \alpha^{m-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \alpha^{m-1} & \alpha^{m-2} & \alpha^{m-3} & \dots & 1 \end{bmatrix} \tag{2.11}$$

If responses within a subject are not expected to vary with time, the correlation structure is exchangeable, and can be written as:

$$\boldsymbol{R}_i(\alpha) = \begin{bmatrix} 1 & \alpha & \alpha & \dots & \alpha \\ \alpha & 1 & \alpha & \dots & \alpha \\ \alpha & \alpha & 1 & \dots & \alpha \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \alpha & \alpha & \alpha & \dots & 1 \end{bmatrix} \tag{2.12}$$

2.3 Relaxing Linearity with Natural Cubic Splines

Part of specifying the correct type of model involves the functional form of the relationship between the covariates and the response variable. In circumstances where overfitting is not a concern (due to availability of data), there is no need to assume relationships are linear. Instead, the relationships between the response and covariates should be flexible, allowed to fit curves if applicable. This is especially true for the data in this study because glucose values are known to rise and fall over the course of the day.

One of the simplest ways to achieve flexibility is to include transformations of a pre- dictor variable in the model. For example, to relax linearity in the time variable, X_t, the term X_t^2 might be included to allow time to have a quadratic relationship with the outcome.

Higher order terms (e.g. X_t^3, X_t^4, etc.) can be used to achieve more flexible fits. The downside of using polynomials, however, is that if the exact relationship is unknown, then the data might be overfit in some regions because the function is being specified across the whole domain of the predictor. In other words, it's not a piece-wise fit. Further, including higher order terms does not fit all functional forms, such as logarithmic functions (Harrell Jr 2015).

A better way to achieve a flexible fit is not just to use a polynomial, but rather to use a piece-wise polynomial, specifically a spline. To create the "pieces" of the polynomial, the X-axis is divided into different sections, and the points where sections begin and end are called knots. Then, a function $f : \mathbb{R} \to \mathbb{R}$ is a kth order spline with knot points at $t_1 < \cdots < t_m$ if:

1. f is a polynomial of degree k on each of the intervals $(-\infty, t_1], [t_1, t_2], \ldots, [t_m, \infty)$.
2. $f^{(j)}$, the jth derivative of f, is continuous at $t_1, \ldots, t_m$ for each $j = 0, 1, \ldots, k-1$.

(Tibshirani 2014)

One particularly useful feature is that because splines are polynomials, they can be expressed as a sum of finite, non-zero terms:

$$f(X) = \sum_{i=0}^{k+m+1} X_i^* \beta_i \tag{2.13}$$

Where X_i^* is a variable constructed from the covariate X, and β_i is the coefficient for the constructed variable. The index $k+m+1$ comes from having to specify terms at each of

the m knots, for each degree of the kth degree polynomial, and for a single additive constant. Therefore, the polynomials are linear in β_i, and the theory of linear or generalized linear models can be applied (Wood 2017).

A common type of spline that has useful properties is a cubic spline, which means that it is a piece-wise polynomial of degree 3 and fits the definition of a spline above. To construct the variables X_i^* for a cubic polynomial with knots given at points $t_1, \ldots, t_m$, however, requires a basis for functions that are continuous at their first and second derivatives. One such basis is the *truncated power basis*, $g_1, \ldots, g_{m+k+1}$, defined to be:

$$g_1(X) = 1,\ g_2(X) = X,\ \ldots\ ,\ g_{k+1}(X) = X^k,$$
$$g_{k+1+j}(X) = (X - t_j)_+^k,\ \ \text{for } j = 1,\ \ldots\ , m$$

where $(X - t_j)_+^k$ is defined as:

$$(X - t_j)_+^k = \begin{cases} 0 & (X - t_j)^k < 0 \\ (X - t_j)^k & (X - t_j)^k \geq 0 \end{cases}$$

With the study data, for example, a cubic spline, $f_{cs}(X_t)$, with 4 total knots, one knot at 3:00am, 9:00am, 3:00pm, and 9:00pm each, would be specified as follows. Note, the time variable is in minutes over the course of the day, so the knots are placed at minutes 180, 540, 900, and 1260:

$$g_1(X_t) = 1 \tag{2.14}$$

$$g_2(X_t) = X_t \tag{2.15}$$

$$g_3(X_t) = X_t^2 \tag{2.16}$$

$$g_4(X_t) = X_t^3 \tag{2.17}$$

$$g_{3+1+1}(X_t) = (X_t - 180)_+^3 \tag{2.18}$$

$$g_{3+1+2}(X_t) = (X_t - 540)_+^3 \tag{2.19}$$

$$g_{3+1+3}(X_t) = (X_t - 900)_+^3 \tag{2.20}$$

$$g_{3+1+4}(X_t) = (X_t - 1260)_+^3 \tag{2.21}$$

$$\implies \quad f_{cs}(X_t) = \beta_0(1) + \beta_1 X_t + \beta_2 X_t^2 + \beta_3 X_t^3 \tag{2.22}$$

$$+ \beta_4(X_t - 180)_+^3 + \beta_5(X_t - 540)_+^3 + \beta_6(X_t - 900)_+^3 + \beta_7(X_t - 1260)_+^3 \tag{2.23}$$

$$= \beta_0 + \beta_1 X_1^* + \beta_2 X_2^* + \beta_3 X_3^* + \beta_4 X_4^* + \beta_5 X_5^* + \beta_6 X_6^* + \beta_7 X_7^* \tag{2.24}$$

$$= \mathbf{X}^* \beta \tag{2.25}$$

Then, from linear model theory:

$$\hat{\beta} = (\mathbf{X}^{*T}\mathbf{X}^*)^{-1}\mathbf{X}^{*T}\mathbf{y} \tag{2.26}$$

A problem with cubic splines, however, is that they have high variance in the tails, which are the piece-wise sections that occur before and after the first and last knot, respectively.

One way to rectify this is to use a natural cubic spline (also called a restricted cubic spline). Specifying that the spline is *natural* imposes an additional constraint: f is a polynomial of degree $(k-1)/2$ on $(-\infty, t_1]$ and $[t_m, \infty)$. A natural cubic spline, then, is defined to be linear in the tails. Further, the number of estimated parameters drops from $m+3$ (disregarding the intercept) to $m-1$ (where m is the number of knots) because in each tail, there is neither a quadratic nor cubic component to estimate, so four degrees of freedom are saved (Harrell Jr 2015). Due to the new constraint, however, a variant of the above truncated power basis is required (Tibshirani 2014).

The natural cubic spline, $f_{ncs}(X)$, is then given by:

$$f_{ncs}(X) = \beta_0 + \beta_1 X_1^{**} + \beta_2 X_2^{**} + \cdots + \beta_{m-1} X_{m-1}^{**} \tag{2.27}$$

where $X_1^{**} = X$, and for $j = 1, \ldots, m-2$:

$$X_{j+1}^{**} = (X - t_j)_+^3 - (X - t_{m-1})_+^3 \frac{t_m - t_j}{t_m - t_{m-1}} + (X - t_m)_+^3 \frac{t_{m-1} - t_j}{t_m - t_{m-1}} \tag{2.28}$$

(Harrell Jr 2015).

Then, on the study data, specifying a natural cubic spline, $f_{ncs}(X_t)$, with 4 total knots on the time variable with the same knot placement as before (minutes 180, 540, 900, 1260) would be:

$$f_{ncs}(X_t) = \beta_0 + \beta_1 X_1^{**} + \beta_2 X_2^{**} + \beta_3 X_3^{**} \tag{2.29}$$

$$= \mathbf{X}^{**}\beta \tag{2.30}$$

Where: (2.31)

$$X_1^{**} = X_t \tag{2.32}$$

$$X_{1+1}^{**} = (X_t - 180)_+^3 - (X_t - 900)_+^3 \frac{1260-180}{1260-900} + (X_t - 1260)_+^3 \frac{900-180}{1260-900} \tag{2.33}$$

$$X_{2+1}^{**} = (X_t - 540)_+^3 - (X_t - 900)_+^3 \frac{1260-540}{1260-900} + (X_t - 1260)_+^3 \frac{900-540}{1260-900} \tag{2.34}$$

(2.35)

X_t^{**} represents the variable X_t transformed by the variant of the truncated power function, and applying linear model theory:

$$\hat{\beta} = (\mathbf{X}^{**T}\mathbf{X}^{**})^{-1}\mathbf{X}^{**T}\mathbf{y} \tag{2.36}$$

Before specifying a natural cubic spline (or any spline), however, both the number of knots and the location of the knots need to be chosen. Per Harrell, the number of knots chosen is the more important concern for model fit, and for location he recommends simply placing knots at equally spaced quantiles if prior-experience is non-informative or not applicable. Doing so helps ensure enough data in each piece-wise section and guards against outliers having too much influence on knot location. For large sample sizes (e.g. $n \geq 100$), Harrell Jr

2015 recommends using 5 knots, but states that using Akaike's Information Criterion (AIC) can be used for a data-driven approach by maximizing the model likelihood ratio, $\chi^2 - 2m$, where m is the number of knots.

For modeling the study data, ten knots were specified, and they were spaced in two hour increments from 4am to 10pm. This spacing was chosen because blood sugar values are expected to peak approximately two hours after eating, so the two hour gaps were in hopes of capturing those peaks. The initial knot at 4am was chosen because people are assumed to be sleeping from midnight to 6am, so blood sugar values are expected to be moving linearly, assuming no nocturnal meals or changes in basal insulin rate. Lastly, because the splines are fitted to the time variable, the sample size is large ($n > 250,000$) and can accommodate the degrees of freedom required.

Chapter 3

Materials and Methods

3.1 Data Collection

The data used in this study consist of the first 35 completed subjects from the investi- gator initiated trial: "A Comparison of Insulin Degludec to Continuous Subcutaneous Infusion of Insulin Aspart for Basal Insulin Delivery in Type 1 Diabetes." Blood sugar data were recorded via Dexcom CGM for both treatment periods, and the treatment order was randomized for each subject. Because Dexcom CGMs record blood sugars values every five minutes, the expected number of blood sugar readings per day for each subject is 288. Additionally, the study team collected demographic data and medical history that the investigators deemed clinically relevant. These data included, but were not limited to age, sex, duration of type 1 diabetes, and type of insulin pump.

The Dexcom CGM devices were downloaded at patient visits, and the blood sugar readings were exported to .csv files to be used for statistical analysis.

3.2 Inclusion/Exclusion Criteria

Inclusion criteria were:

Study Inclusion Criteria

• Male and female patients > 18 years of age with type 1 diabetes using CSII with any pump for > 12 months.

• Females must be using adequate contraception, defined as oral contraceptive pill,

barrier method of contraception, or surgical method (tubal ligation or hysterectomy).

- Good glycemic control (HbA1c < 8.0%).
- Patients are experienced in carbohydrate counting, evidenced by pump downloads showing frequent meal boluses with realistic carbohydrate entries, few over-rides of the pump bolus calculator, few to no omitted boluses (at least 3 boluses per day), and post-meal glucose levels generally below 200 mg/dl indicating accurate carbohydrate assessment.
- Patients are regular (>85% of time) users of the Dexcom G5 or Dexcom G6 CGM.
- Pump download confirms correct use of insulin pump features, including appropriate use of bolus calculator with minimal overrides, entering carbohydrate content of meals, at least 3 boluses taken per day, appropriate use of correction boluses, and infusion set changes every 2 to 3 days.
- No serious comorbidities including: retinopathy requiring active intervention, eGFR < 30, CV event within the previous 6 months, active malignancy with ongoing treatment, any condition requiring chronic use of systemic glucocorticoids, or any other condition which in the opinion of the investigator would interfere with the subject's ability to comply with the study protocol or acutely affect insulin requirements.
- Able to comply with study protocol.
- Ability to provide written informed consent prior to any study-related procedures.

Study Exclusion Criteria

- Subjects with type 2 diabetes.
- Subjects with HbA1c > 8.0%
- Subjects not using CSII and CGM (ie, on MDI)

- Subjects inexperienced in the use of CSII, or whose pump download shows poor utilization of bolus calculator features, i.e. fewer than 2 boluses per day, lack of correction boluses, frequent overrides of the recommended boluses, unrealistic carbohydrate entries (suggestive of under-bolusing), not changing infusion set at least every 3 days, or other evidence of poor insulin pump usage.
- Subjects inexperienced in or not regular users (>85% of time) of Dexcom G5 or Dexcom G6 CGM
- Subjects who are using a Medtronic pump with low blood glucose suspend who are unwilling to use the Dexcom CGM or to disengage the low blood glucose suspend feature of the pump.
- Use of any other CGM than Dexcom.
- Serious concomitant illness.
- Females unwilling to use adequate contraception, intending to become pregnant, or breastfeeding.
- Known or suspected allergy to study products, their excipients, or related products.
- Previous participation in this trial. Note: subjects who screen fail because of A1c may rescreen once if, in the opinion of the investigator, the HbA1c was explainable (i.e., recent steroid injection or illness, etc) and atypical for the subject.
- Hypoglycemic unawareness.
- Episode of severe hypoglycemia (requiring assistance for treatment) within the previous 90 days.

3.3 Description of Data and Cleaning

To obtain the data, the study team downloaded subjects' CGMs, and at the end of the study all CGM data, which consisted of all recorded blood sugars identified by a date-time stamp, were exported to a .csv file. All 20 weeks of data for both observation periods were available, but only the two observation periods were used in the analysis. Though Dexcom CGMs record blood sugars in a wide range, they are limited to 40 mg/dL and 400 mg/dL, inclusively. Outside of these values, the device either records "Low" or "High". (Dexcom G6 User Guide, p112). The total number of observed "High" values was 395 (0.15% of the data), and the total number of "Low" values was 143 (0.05% of the data). These values were set to 401 and 39 for "High" and "Low", respectively. Because the threshold used to dichotomize the data was 70 mg/dL, the primary outcome is not affected. Finally, because Dexcom CGMs only record blood sugar values every five minutes, the glucose readings were grouped by five-minute increments.

Once the Dexcom data were cleaned, demographic and covariate data were joined in by subject ID. Specifically, age, sex, duration of type 1 diabetes, and type of insulin pump were included. These data required little cleaning beyond two considerations: missing data and sparse categories. Only two subjects had missing data for duration of type 1 diabetes, and predicted mean imputation based on subject characteristics (age, insulin pump, and sex) was used to fill in the missing values. Types of insulin pump were grouped by manufacturer, so where there were originally eight different categories for insulin pumps, with the regrouping there were only four, including an "other" category. The "other" category included pumps manufactured by the company Animas as well as one pump with a missing

value. Finally, while the type of insulin pump ideally remained static, one subject changed the type of insulin pump used from the first to the second observation period. The subject was on an "Animas" insulin pump for the first observation period and a "Medtronic" insulin pump for the second observation period. This subject was categorized as "other" because they started on an Animas.

3.4 Statistical Analysis

For modeling, the data were grouped by subject, five minute interval, and regimen. Specifically, for each treatment, every subject has a single observation for every five minute period over 24 hours. These five minute increments were aggregated across the 14 days for each treatment, and the total number of glucose readings as well as total number of glucose readings below 70 mg/dL were included. To assess hypoglycemic risk over the course of the day and at crucial times for each treatment, as well as to profile the risk factors for hypoglycemia, the data were modeled using a generalized linear mixed effects model with a logit link and a random intercept plus AR(1) correlation to capture the dependence structure. In addition to treatment, the model controlled for age, sex, duration of type 1 diabetes in years, insulin pump, treatment order, and allowed for the interaction of time and treatment. Finally, the time variable (five minute intervals over the course of the day) was fit with a natural cubic spline with 10 knots to capture meals and snacks, insulin correction doses, and changes in physical activity, all of which cause changes in blood glucose. All data cleaning and statistical analysis were performed in R 3.6.0, and the function used to fit

the model was “glmmPQL" the “MASS" package.

Chapter 4

Results

4.1 Study Sample

A total of 35 study participants were included in this study. The average age was 57.2 (SD = 14.0) years, and there were 18 (51.4%) males and 17 females (48.6%) (Table 4.1). Of the four different categories for insulin pumps, 14 (40.0%) were Medtronic, 11 (31.4%) were Omnipod, 6 (17.1%) were Tandem, and 4 (11.4%) were other. The average duration of type 1 diabetes was 34.5 years (SD = 22.9).

To examine differences between treatments, the data were aggregated at 5 minute intervals across days. Across the 35 subjects, we observed 484 person-days on CSII treatment and 479 person-days of Tresiba treatment. The average proportion of hypoglycemic events per day was 0.0214 and 0.0232, for CSII and Tresiba, respectively. This translates to 30.8 minutes/day of hypoglycemia for CSII and 33.4 minutes/day of hypoglycemia for Tresiba. The average number of readings per day in both groups was 275. Lastly, in this crossover study, 17 subjects were randomized to receive Tresiba during the first phase of the study and CSII for the second phase, while 18 were randomized to receive CSII during the first phase and Tresiba during the second phase. Table 4.2 provides further information regarding between treatment comparisons.

Table 4.1: Descriptions of Characteristic Differences Between Subjects

	Overall (n=35)
Age (years)	
Mean (SD)	57.2 (14.0)
Median [Min, Max]	59.0 [26.0, 80.0]
Quantiles [2.5, 97.5]	[31.95 , 80]
Duration of T1DM (years)	
Mean (SD)	33.0 (18.0)
Median [Min, Max]	29.6 [3.10, 59.0]
Quantiles [2.5, 97.5]	[4.04 , 58.32]
Sex	
Male	18 (51.4%)
Female	17 (48.6%)
Insulin Pump	
medtronic	14 (40.0%)
omnipod	11 (31.4%)
other	4 (11.4%)
tandem	6 (17.1%)

4.2 Exploratory Analysis

The focus of the exploratory analysis was on glucose over time. Ultimately, three plots were used to understand the time trend and guide statistical modeling. All plots aggregated the subjects' data over 24 hours and split it by treatment regimen.

Figure 4.1, examines the smoothed mean and median 24 hour glucose profile more closely. Both plots show a clear polynomial trend over time, with three distinct rises near typical

Table 4.2: Person-day Summaries of Glucose Data

	CSII (Novolog only) (n=484)	Novolog + Tresiba (n=479)
Average Glucose per Day (mg/dL)		
Mean (SD)	152 (26.6)	152 (30.2)
Median [Min, Max]	149 [90.4, 285]	147 [89.7, 337]
Quantiles [2.5, 97.5]	[112.012 , 211.114]	[106.339 , 218.755]
Readings per Day		
Mean (SD)	275 (27.8)	275 (27.8)
Median [Min, Max]	286 [55.0, 295]	286 [39.0, 297]
Quantiles [2.5, 97.5]	[195.075 , 288]	[199.85 , 288]
Proportion of Lows (Glucose < 70 mg/dL)		
Mean (SD)	0.0214 (0.0437)	0.0232 (0.0379)
Median [Min, Max]	0.00 [0.00, 0.319]	0.00353 [0.00, 0.239]
Quantiles [2.5, 97.5]	[0 , 0.142]	[0 , 0.126]
Days of Data		
Mean (SD)	13.9 (0.605)	13.9 (0.869)
Median [Min, Max]	14.0 [10.0, 14.0]	14.0 [3.00, 14.0]
Quantiles [2.5, 97.5]	[13 , 14]	[14 , 14]

meal times (i.e. 8:00am, 12:00pm, and 6:00pm), and an overnight (midnight to 6:00am) decline. The range is narrow, neither dipping below 120 mg/dL nor rising above 170 mg/dL at any point in time.

Figure 4.2, shows percentiles of glucose concentrations across subjects (light blue) over the course of days during the study. The dark blue trend line is a cubic spline smoother on the average, the black lines are cubic smoothers on the 2.5th and 97.5th percentiles, and the purple and blue shading show the middle 50% and 99% of the data. Lastly, the horizontal yellow

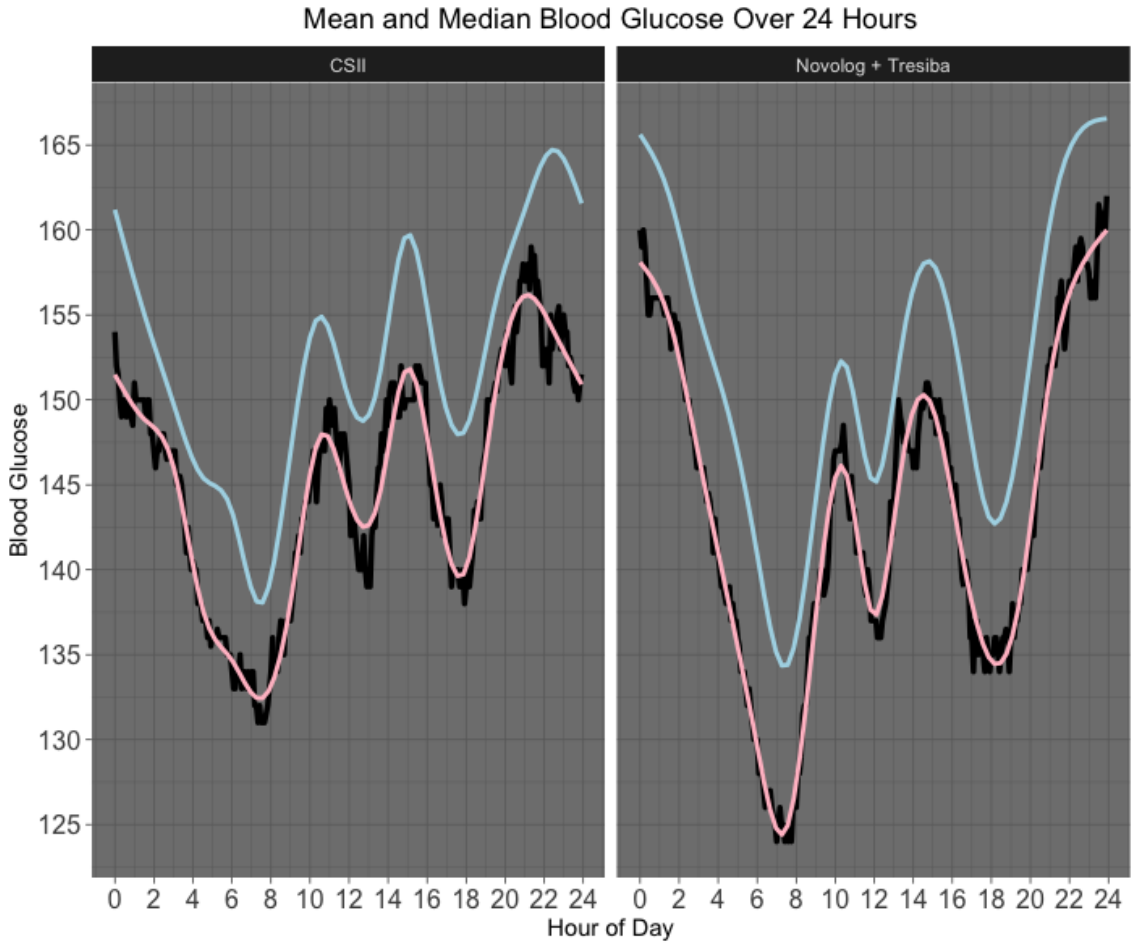


Figure 4.1: Smoothed graph of mean and median blood glucose over 24 hours. The blue line is the population average; the pink line is the smoothed population median; the black line is the non-smoothed population median. The smoothers used are both cubic splines with 17 knots.

line is at 70 mg/dL, the cutoff used to define hypoglycemia in the subsequent analyses. On CSII, subjects' blood glucose appears relatively stable from midnight to 6:00am at the 2.5 and 97.5 percentiles. Then, their blood glucose rises and falls with three distinct peaks at approximately 9:00am, 3:00pm, and 10:00pm. On Tresiba, the 97.5th percentile decreases from midnight to 6:00am before increasing again. There appears to be a slight peak at

9:00am, but the only clear peak appears at approximately 3:00pm. After 3:00pm, subjects' blood glucose falls until approximately 6:00pm before again increasing until midnight. At the 2.5th percentile, Tresiba appears to have a less variable profile, with blood glucose declining steadily until about 7:00am, rising to approximately 10:00am, declining again until about 6:00pm, and then increasing from 6:00pm to midnight. On CSII, the 2.5th percentile steadily rises and falls until roughly 6:00pm, where it appears to steadily rise until midnight.

To address study questions surrounding hypoglycemic risk, the study team defined blood sugar values of less than 70 mg/dL to be hypoglycemia, and time spent in that range as time that subjects were at risk of developing clinically relevant symptoms of hypoglycemia. Figure 4.3 shows two plots examining hypoglycemic risk. The top plot shows the proportion of hypoglycemic observations over the course of the day across study subjects. The green line is a cubic spline smoother, and the blue points are the observed data. The plot shows a similar trend to the previous figure for CSII, exhibiting a similar pattern around meal times with a slight increase in hypoglycemic risk around 10:00pm. The Tresiba arm, however, has two primary peaks for observed proportion of hypoglycemia: approximately 7:30am and 6:00pm.

The lower plot is the difference (Tresiba - CSII) in the proportion of hypoglycemic events between the two treatments. The green points are the observed differences, the light blue line is a cubic spline smoother, and the yellow horizontal line is at 0. The plot shows that from approximately midnight to 5:00am, Tresiba has less hypoglycemia than CSII. From 5:00am to approximately 8:30pm, CSII has less hypoglycemia, but from approximately

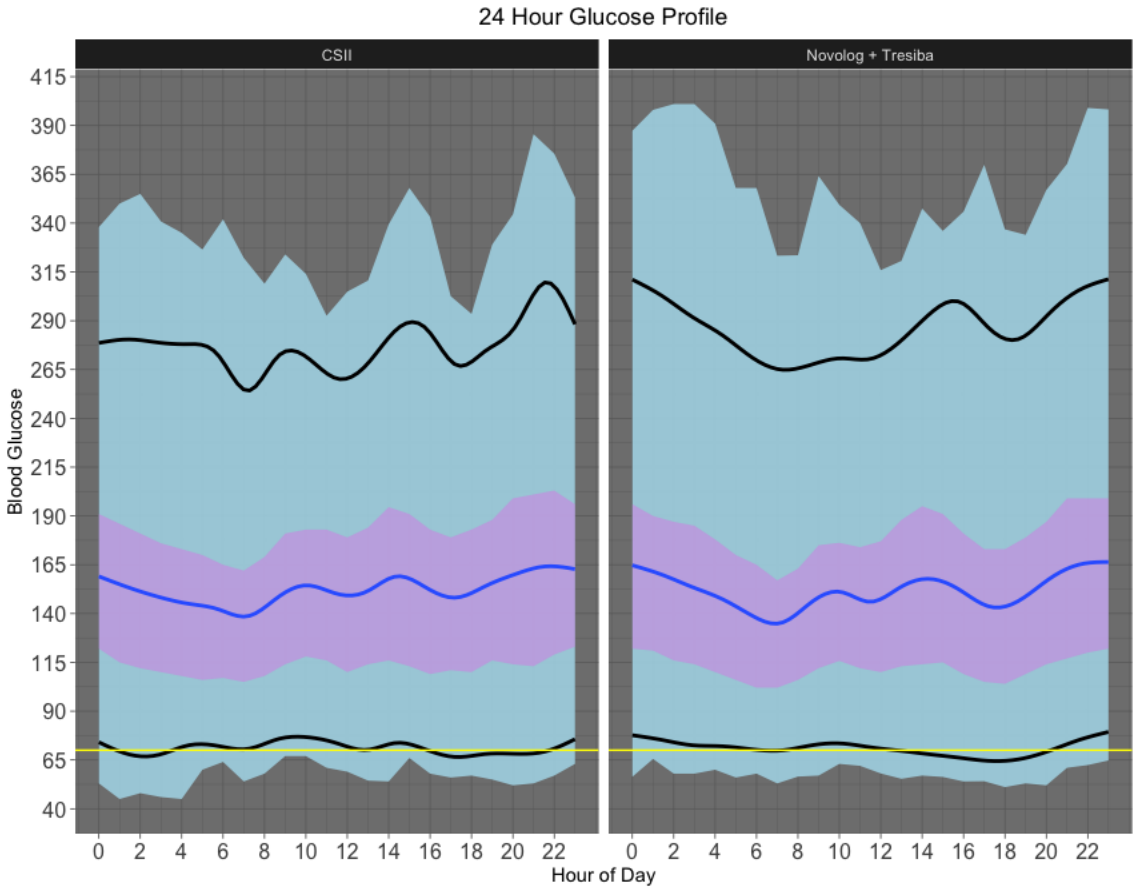


Figure 4.2: Aggregated 24 Hour Glucose Profile by Treatment Regimen. The light blue shading shows the middle 99% of the data, the purple shading is the middle 50% of the data, and the blue trend line is a cubic spline with 17 knots smoothing over the average. The black lines are cubic spline smoothers with 17 knots for the 2.5th and 97.5th percentiles, and the yellow line is at 70 mg/dL.

8:30pm to just before midnight, Tresiba again has less hypoglycemia.

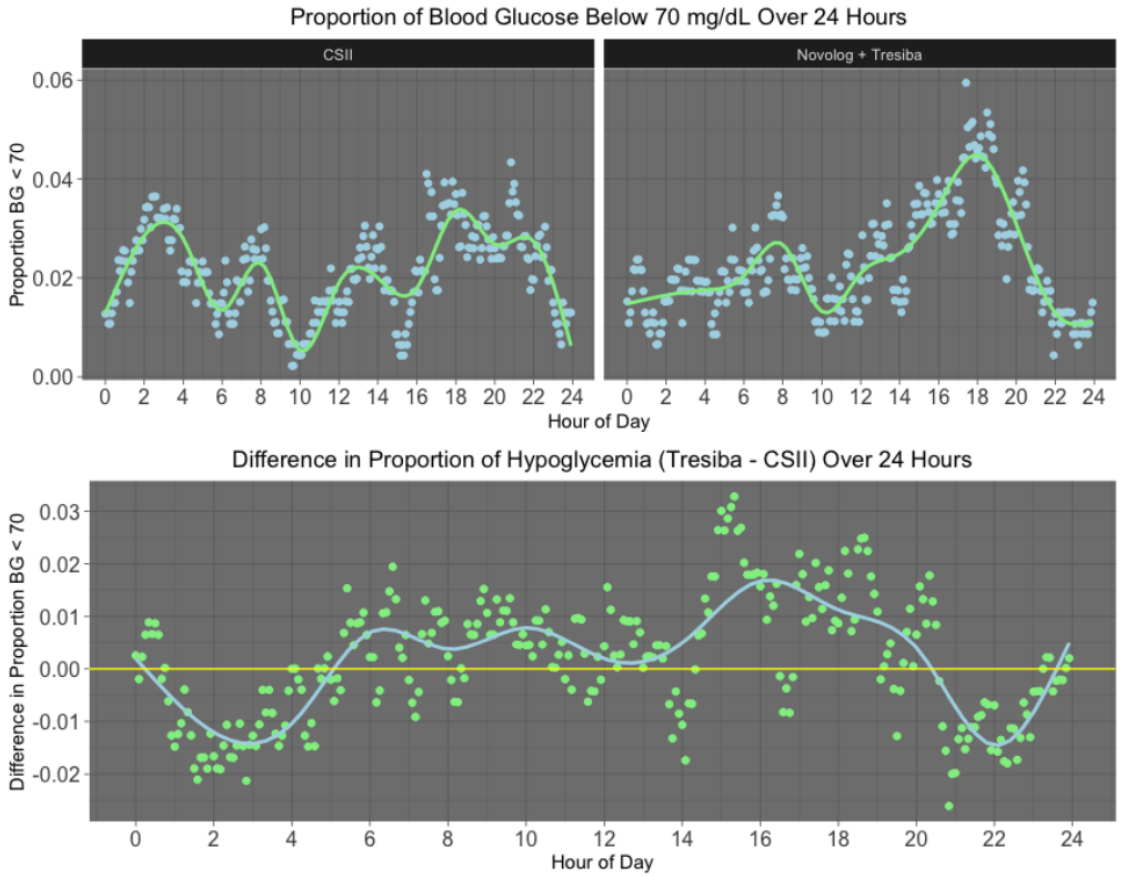


Figure 4.3:
Top: Smoothed graph of hypoglycemic events over 24 hours. Blue points are the observed data. The green line is a cubic spline smoother with knots placed every 2 hours from 4:00am to 10:00pm. Data are averaged across all subjects and plotted over time. **Bottom**: Plot of the difference in proportion of hypoglycemia between CSII and Tresiba. The green points are the observed differences (Tresiba - CSII), the blue line is a cubic spline smoother with knots placed every 2 hours from 4:00am to 10:00pm, and the yellow line is at 0.

4.3 Model Results

Figure 4.4 displays results from the generalized linear mixed effects model with AR(1) correlation analysis showing odds ratios and 95% confidence intervals. The data are on

the odds ratio scale, so values greater than 1 indicate an increase in risk of developing hypoglycemia, and values less than 1 indicate a decreased risk. Both the treatment order and Tandem insulin pump were statistically significant. The treatment order had a point estimate of 0.83 and 95% CI from 0.71 to 0.98, indicating that being randomized to receive Tresiba first was associated with less hypoglycemia. The Tandem insulin pump had a point estimate of 0.38 and a 95% CI from 0.17 to 0.82, indicating that subjects who used a Tandem insulin pump, as compared to a Medtronic, had a lower risk of developing hypoglycemia. Additionally, a triple interaction term between time, treatment, and treatment order was tested for significance using a Wald Chi-Squared test ($p = 0.082$), and because it was not statistically significant, it was removed from the model to aid interpretation and understanding of model results.

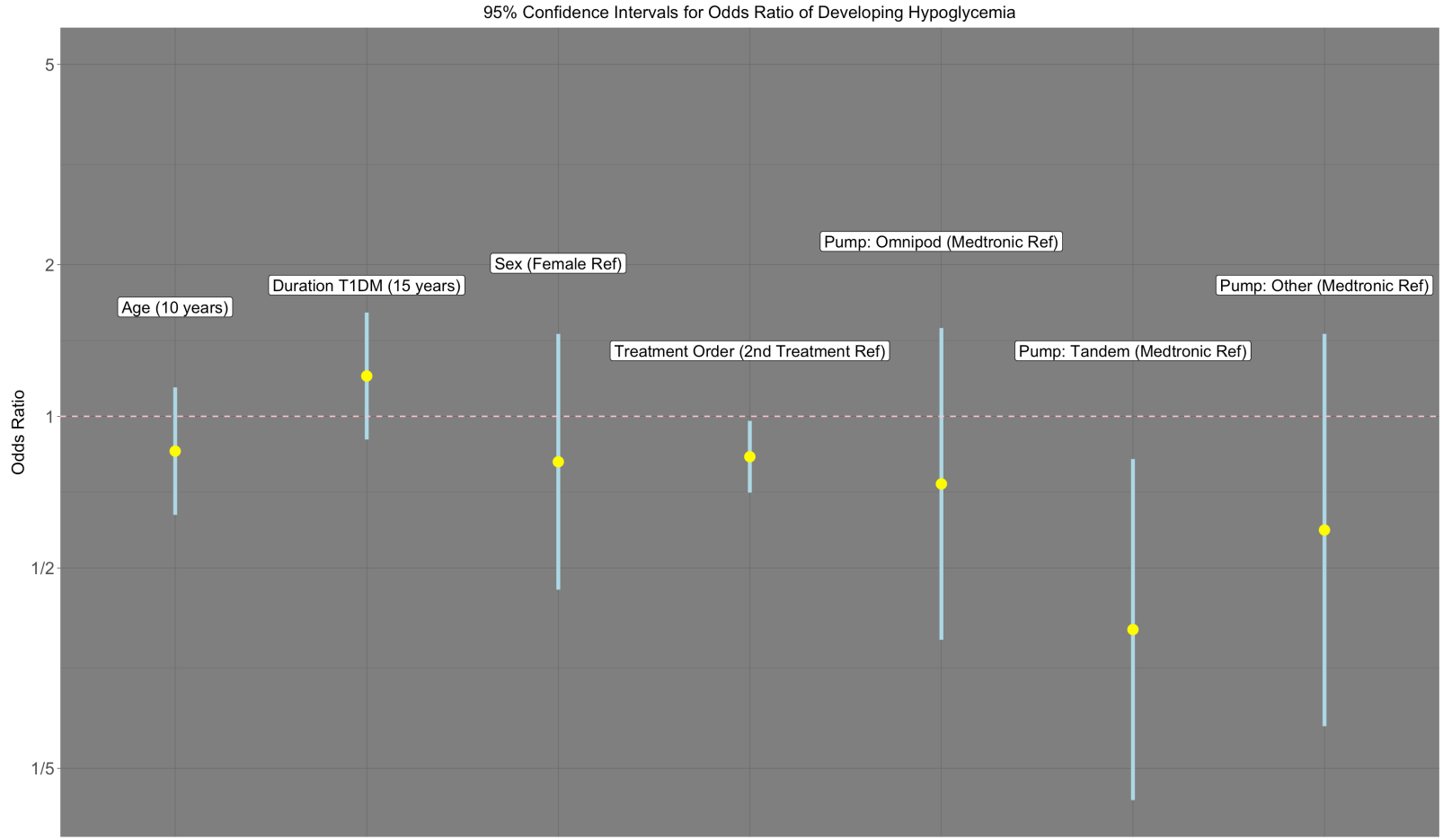


Figure 4.4: Dot-whisker plot of covariates controlled for in the model. Point estimates are in yellow; 95% confidence intervals are in blue, and a dividing line for no effect is in pink at 1.

The Tresiba effect was modeled flexibly over the course of the day using a treatment by (flexible) time interaction, and Figure 4.5 shows two plots summarizing the dynamic effects over the course of the day. The top plot shows 95% confidence intervals for the average Tresiba effect at different times of the day. The bottom plot shows the change in hypoglycemic risk of being on Tresiba versus CSII over the full 24 hours of the day. The confidence intervals in the top plot show that Tresiba is associated with lower risk of hypoglycemia from midnight to 5:00am (odds ratio: 0.65, 95% CI from 0.46 to 0.93) and from 8:30pm to midnight (odds ratio: 0.61, 95% CI from 0.38 to 0.99). Tresiba was associated with a higher risk of hypoglycemia, however, from 1:00pm to 8:30pm (odds ratio: 1.44, 95% CI from 1.11 to 1.88). Across the full 24 hours of the day, on average, no statistically significant differences were detected. The bottom graph shows how Tresiba's hypoglycemic risk profile changes over the course of the day. The early morning (midnight to 5:00am) and late evening (approximately 8:30pm to midnight) hours are when Tresiba appears to be most protective against being hypoglycemic, as compared to CSII.

Lastly, as the simulation study for population heterogeneity will show, the standard deviation for the random intercepts affects where hypoglycemic risk concentrates in a study population, so results must be interpreted in the context of that value. For this population, the standard deviation of the random intercepts was estimated to be 0.72.

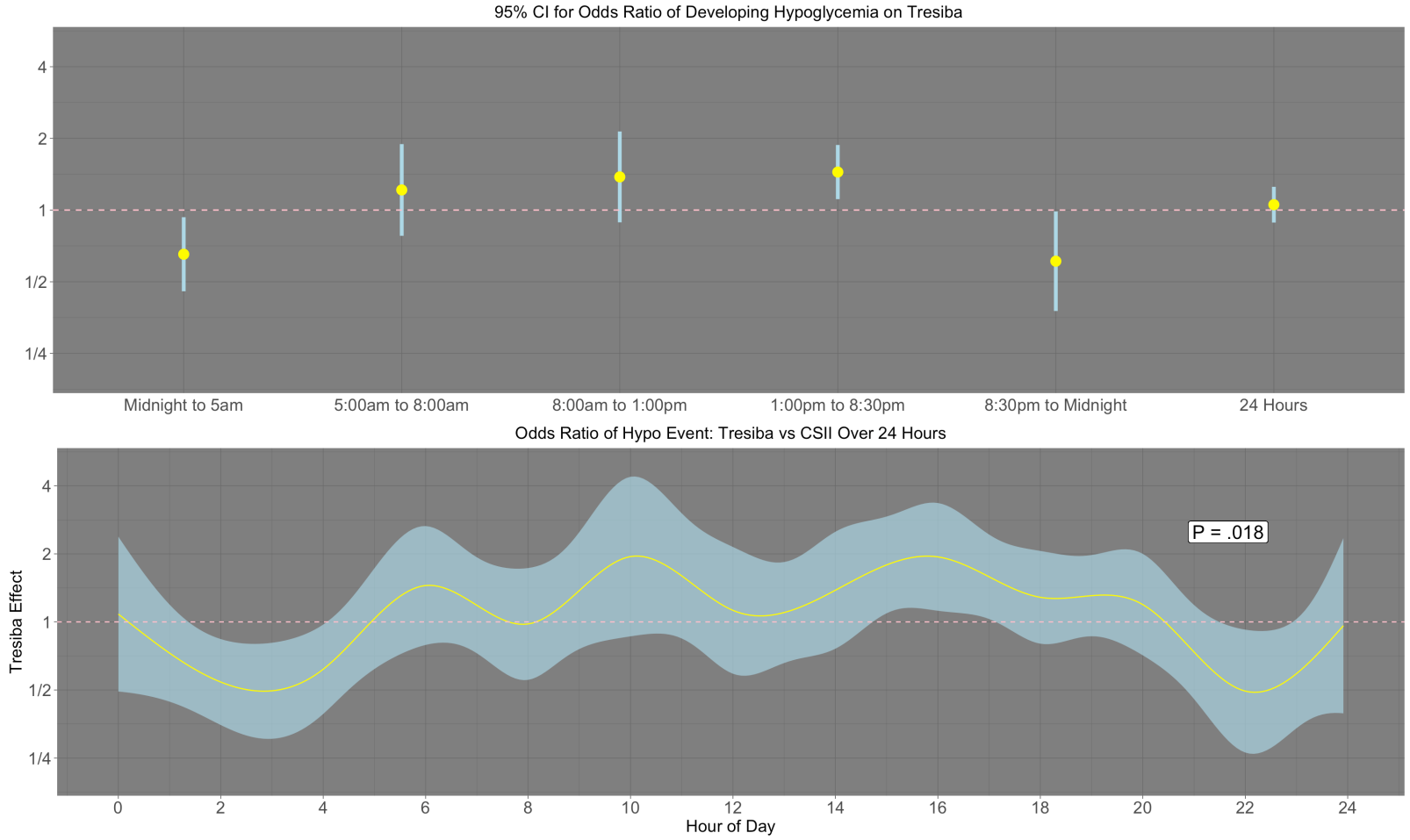


Figure 4.5:
Top: Confidence intervals for the effect of Tresiba vs CSII at different times of the day. Point estimate are in yellow and 95% confidence intervals are in blue.
Bottom: Odds of developing hypoglycemia over the course of the day for Tresiba vs CSII

Figure 4.6 shows the averaged model predictions for hypoglycemia over the course of the day. The average prediction is smoothed with a natural spline in yellow, and the observed averages from the data are plotted in blue. Specifically, the probability of hypoglycemia was predicted for all 5 minute intervals over the course of the day, then averaged across all subjects, and split by treatment.

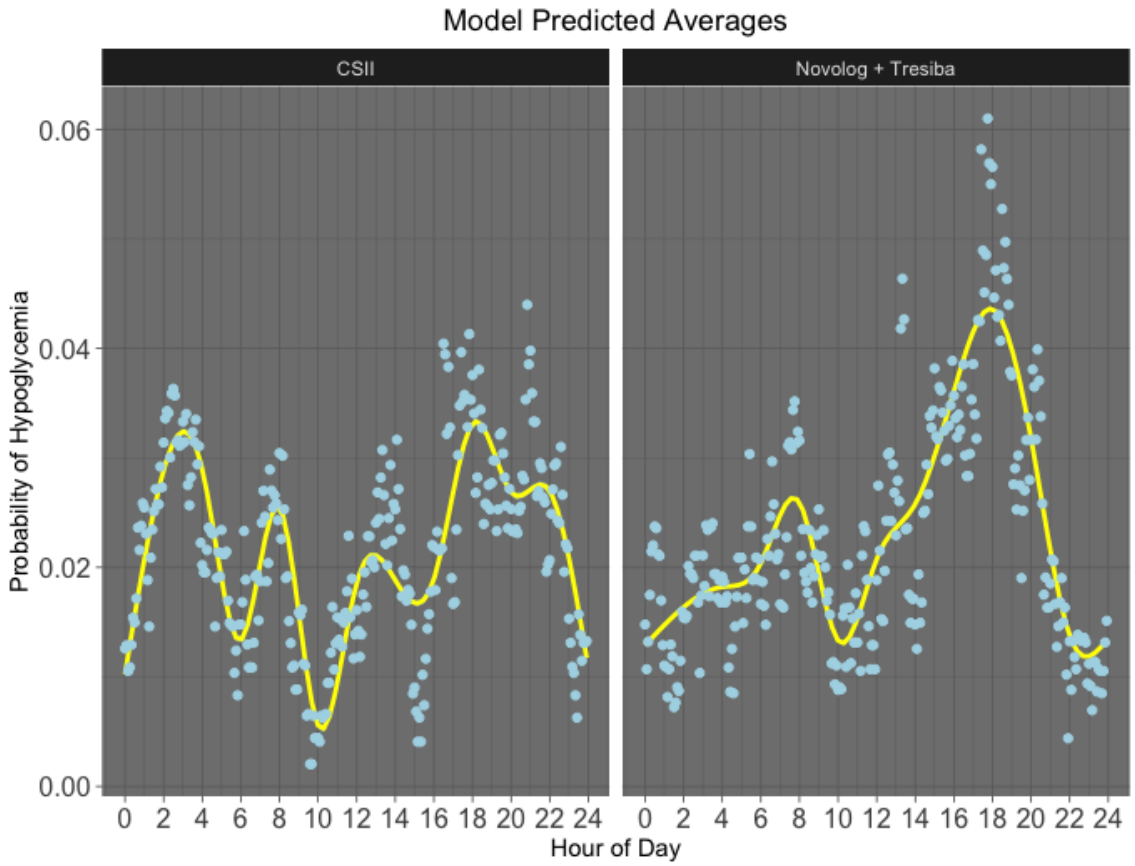


Figure 4.6: Averaged model predictions for experiencing hypoglycemia by treatment regimen. The yellow line is smoothed average of the model predictions, and blue points are the observed averages.

Chapter 5

Simulation Studies

To better understand model results and build context in which to interpret the findings, the study team conducted two simulation studies. The first study sought to answer how the heterogeneity of a population, defined by the variation in random intercepts, affected the risk of hypoglycemia. Would a more heterogeneous population increase the number of people experiencing 30 minutes or more of hypoglycemia per day, and if so, what does the risk concentration look like? The second simulation study sought to understand the operating characteristics of the models used, assessing how dependence structure misspecification and differences in population heterogeneity affect inference. Specifically, are the models biased, and if so, then how much does misspecification affect confidence interval coverage?

5.1 Population Heterogeneity

To examine the effect of population heterogeneity on time at risk, the study team simulated three study populations of 10,000 subjects all with the same fixed effects, but with random effects generated from a Normal distribution with mean 0 and standard deviations of 0.30, 0.60, and 1. In each population, the number of individuals who experienced greater than thirty minutes of hypoglycemia was calculated and compared. To ensure that each study population had the same baseline probability of experiencing hypoglycemia on average, however, the average probability of being hypoglycemic during the day for the study populations with random effect standard deviations of 0.60 and 1 were adjusted back to be the same as the population with 0.20 for the random intercept standard deviation by subtracting off the average difference.

Table 5.1 shows how the number of individuals who spend more than 30 minutes per day with a blood glucose below 70 mg/dL changes with the standard deviation of the random intercepts. Each simulated study population included 10,000 individuals with identical covariate values, and the average probability of hypoglycemia over the day was held constant across all three populations. As the level of population heterogeneity increases, the number of individuals who experience more than 30 minutes of hypoglycemia per day also increases, moving from 868, to 1933, to 2398 for levels of heterogeneity equal to 0.30, to 0.60, to 1, respectively.

SD of Random Intercepts	Number at Risk
0.30	868
0.60	1933
1	2398

Table 5.1: Count of simulated people who experience greater than 30 minutes of hypoglycemia (glucose < 70 mg/dL) per day for differing levels of population heterogeneity. The simulated population size is 10,000, and the daily average probability of experiencing hypoglycemia is held constant while the standard deviation of the random intercepts increased.

Figure 5.1 offers a visual representation of the Table 5.1 in terms of the empirical cumulative distribution functions. The graph shows that as the standard deviation of the random intercepts (g11) increases, the distributions shift left towards 0 and approach 1 more slowly. The yellow vertical line at 30 minutes of hypoglycemic risk helps to show how the percentile of 30 minutes changes as the population heterogeneity changes. The distributions also build up at 0 as g11 increases. Notice that for g11 = 1, the blue line, the value 0 is

over the 25th percentile. For both g11 = 0.3 and g11 = 0.6, however, there are far fewer people experiencing 0 minutes of hypoglycemia. Note, to aid in visualizing the change in trend, all individuals with greater than 150 minutes of risk per day were dropped from the visualization.

5.2 Dependence Structure Misspecification

Due to how closely the glucose values are measured in time, and because glucose can only change so quickly in five minutes, the data have serial correlation (correlation over time). In other words, the blood glucose value at time ti1 is informative of the blood sugar value at time ti2. Therefore, when specifying the model, this correlation should be accounted for. The difficulty, however, is that currently, no package in R is available for easily specifying the type of model the study team would like to use: generalized linear mixed effects model with a logit link and random intercepts plus AR(1) correlation dependence structure estimated using Adaptive Gaussian Quadrature (AGQ). Using the function ‘glmer’ in the “lme4" package does not allow for specifying serial correlation, but it does use AGQ to approximate the likelihood, which is known to be less biased than Penalized Quasi Likelihood (PQL) for approximating variance components of binary outcomes (N. Breslow 2004). However, the study team is able to include serial correlation with the function ‘glmmPQL’ from the “MASS" package, but the approximation method, as the name suggests, uses PQL rather than AGQ.

To investigate how these differences affect the model results, the study team simulated data with a similar correlation structure to the study data and investigated three different models.

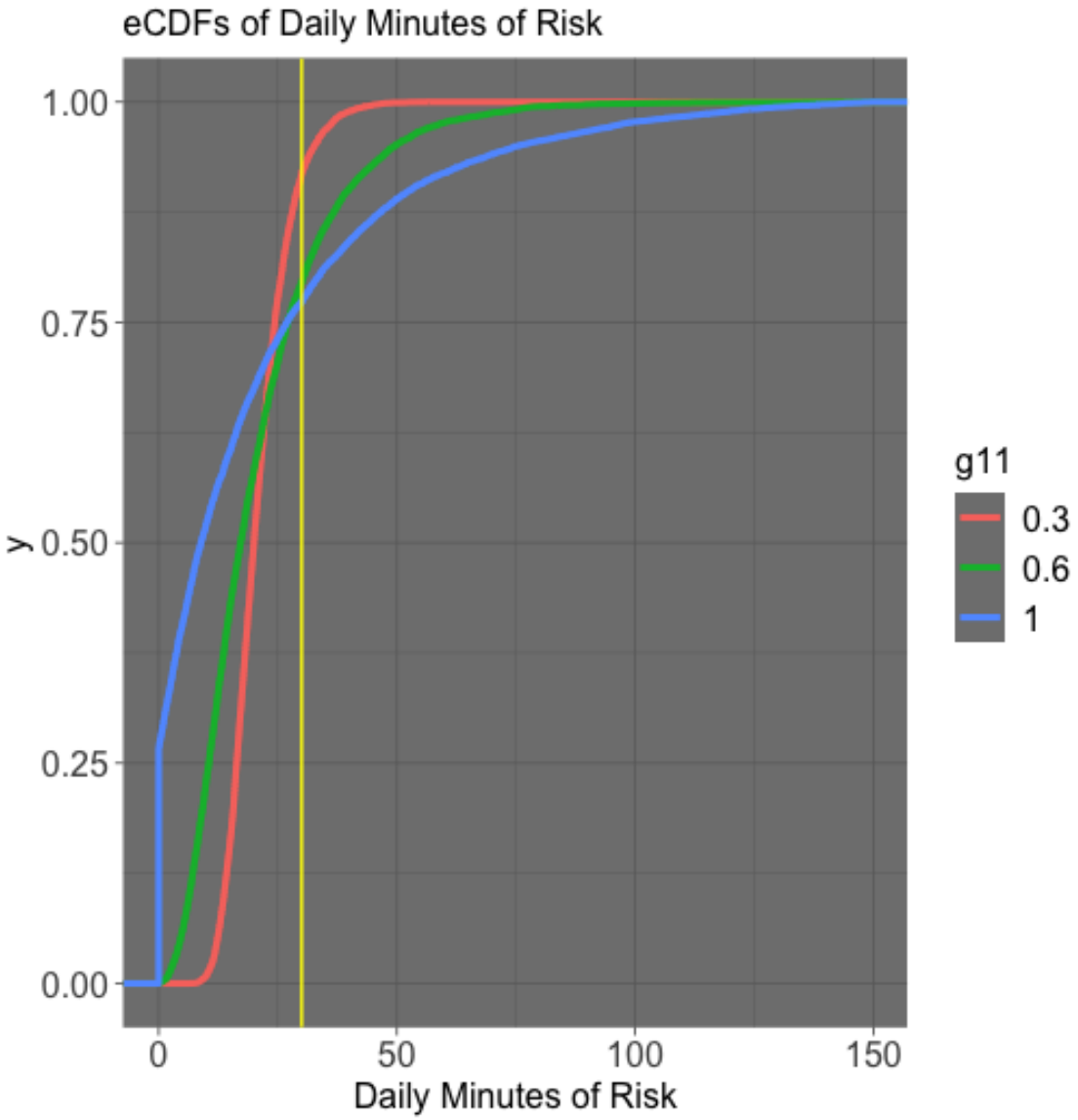


Figure 5.1: Empirical CDFs showing how changes in population heterogeneity also change the concentration of people at risk of developing hypoglycemia. The yellow vertical line is at 30 minutes of risk.

The first model uses the function 'glmer', the second model uses the function 'glmmPQL', and the third model uses 'glmer' but uses a non-parametric bootstrap to approximate the

standard errors. Then, the percent bias in the estimates and the coverage probabilities of the confidence intervals were calculated. To further understand operating characteristics, the study team varied the serial correlation and population heterogeneity, refitting the ‘glmer’ (without bootstrap) and ‘glmmPQL’ models and once more calculating bias and confidence interval coverage.

Simulating binomial data with serial correlation structure while maintaining linearity in the beta coefficients, however, is difficult due to the non-linearity of the logit transformation. To do this, data were modeled with a random intercept, and then those data were passed to marginal model software, which “thought" the mixed effect data were actually marginal mean data. Then, using a transition term, auto correlation was included.

To explore how dependence structure misspecification affects the fit of the ‘glmer’ (non-bootstrap) and ‘glmmPQL’ models, a single data set was simulated and then fit with each model. This was repeated 1000 times, so a true sampling distribution was produced from the beta coefficients of each model. For the ‘glmer’ model with a non-parametric bootstrap, a single data set was generated, subject IDs were sampled with replacement 100 times, and each bootstrap sample was fit with ‘glmer’. Then, the beta coefficients were averaged across all bootstrapped models, and the average was used for the sampling distribution. That process was repeated a total of 500 times.

Table 5.2 shows the operating characteristics for the three models. All three models had

95% confidence interval coverage that is too low, indicating inflated type 1 error rates. The glmmPQL model performed the best in terms having the least bias and the best confidence interval coverage. To calculate the percent bias in the parameter estimates for each model, the true parameter value was subtracted from the average estimated beta, then multiplied by 100:

$$\text{Bias}_{\beta} = 100 \times \left(\frac{1}{M} \sum_{m=1}^{M} \hat{\beta}_m - \beta \right) / \beta = \left(\bar{\hat{\beta}} - \beta \right) / \beta \tag{5.1}$$

where $\hat{\beta}_m$ is the β estimate from the sampling distribution, and M is the total number of observations in the sample distribution.

To calculate the bias in the variance estimates, the variance of the estimated betas in the sampling distribution is calculated, and that is subtracted from the averaged estimates of the variance (model estimates of the standard error):

$$\text{Bias}_{\text{Var}} = 100 \times \left(\frac{1}{M} \sum_{m=1}^{M} \hat{\text{Var}}(\hat{\beta}_m) - s^2_{\hat{\beta}} \right) / \, s^2_{\hat{\beta}} \tag{5.2}$$

where:

$$s^2_{\hat{\beta}} = \frac{1}{M-1} \sum_{m=1}^{M} (\hat{\beta}_m - \bar{\hat{\beta}})^2 \tag{5.3}$$

To further probe the effects of dependence structure misspecification, the glmer (non-bootstrap) and glmmPQL models were refit under varying conditions, with each condition simulated 2000 times. The standard deviations of the random intercepts were 0.50, 1, and 2, and the parameters for the amount of serial correlation included were 0, 2, and 4, where 0

model	**Intercept = −0.50**	**Treatment = 0.50**	**Time = 0.25**	**Treatment:Time = 0.25**
	$\hat{\beta}$ $s^2_{\hat{\beta}}$ Bias_{β} Bias_{var} CI coverage	$\hat{\beta}$ $s^2_{\hat{\beta}}$ Bias_{β} Bias_{var} CI coverage	$\hat{\beta}$ $s^2_{\hat{\beta}}$ Bias_{β} Bias_{var} CI coverage	$\hat{\beta}$ $s^2_{\hat{\beta}}$ Bias_{β} Bias_{var} CI coverage
glmer	-0.68	0.66	0.33	0.33
	0.009	0.012	0.008	0.021
	33.4	31.7	30.4	33.9
	-31.6	-82.2	-68.6	-74.6
	42.5	24.8	57.8	58.5
PQL	-0.48	0.47	0.24	0.25
	0.004	0.005	0.004	0.011
	-4.07	-4.65	-3.30	-1.83
	-2.22	-9.02	-7.02	-8.11
	92.4	92.5	93.8	93.2
bootstrap	-0.69	0.68	0.34	0.34
	0.010	0.013	0.010	0.022
	37.6	35.7	34.6	36.6
	-3.60	-83.06	-70.9	-75.5
	50.8	22.8	54.2	56.4

Table 5.2: Confidence interval coverage and bias of glmer, glmmPQL, and bootstrapped glmer models rounded to three decimal places. The true parameter values are listed at the top of the table: -0.50, 0.50, 0.25, 0.25.

indicates no serial correlation and 4 is approximately the level of serial correlation estimated in the data (0.82). Figure 5.3 shows the results graphically. The full tabular results of the simulation are shown in the appendix.

Figure 5.2: Plots showing change in bias of point estimates for intercept, time, treatment, and the treatment by time interaction as population heterogeneity and serial correlation change. The models fit are glmer and glmmPQL.

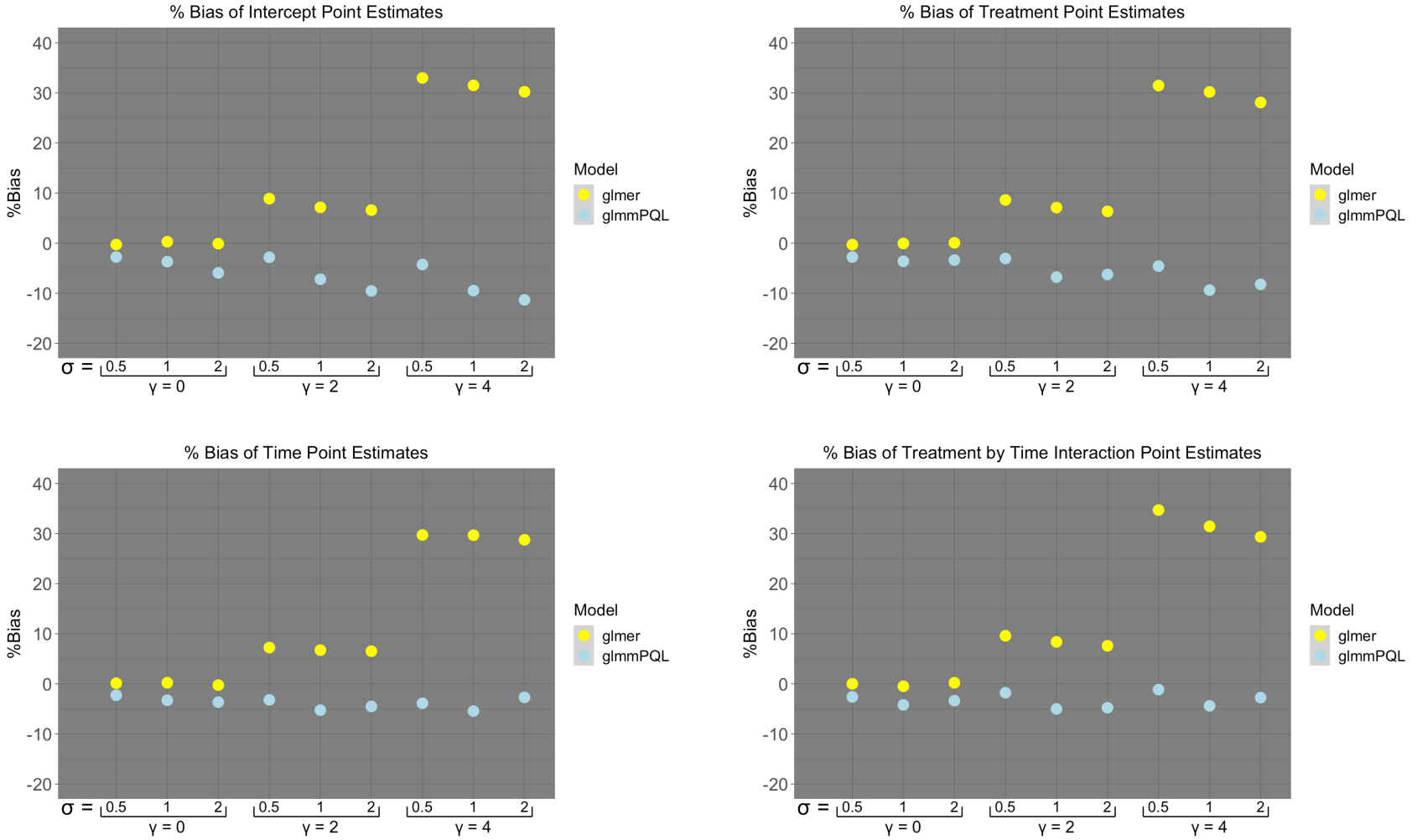
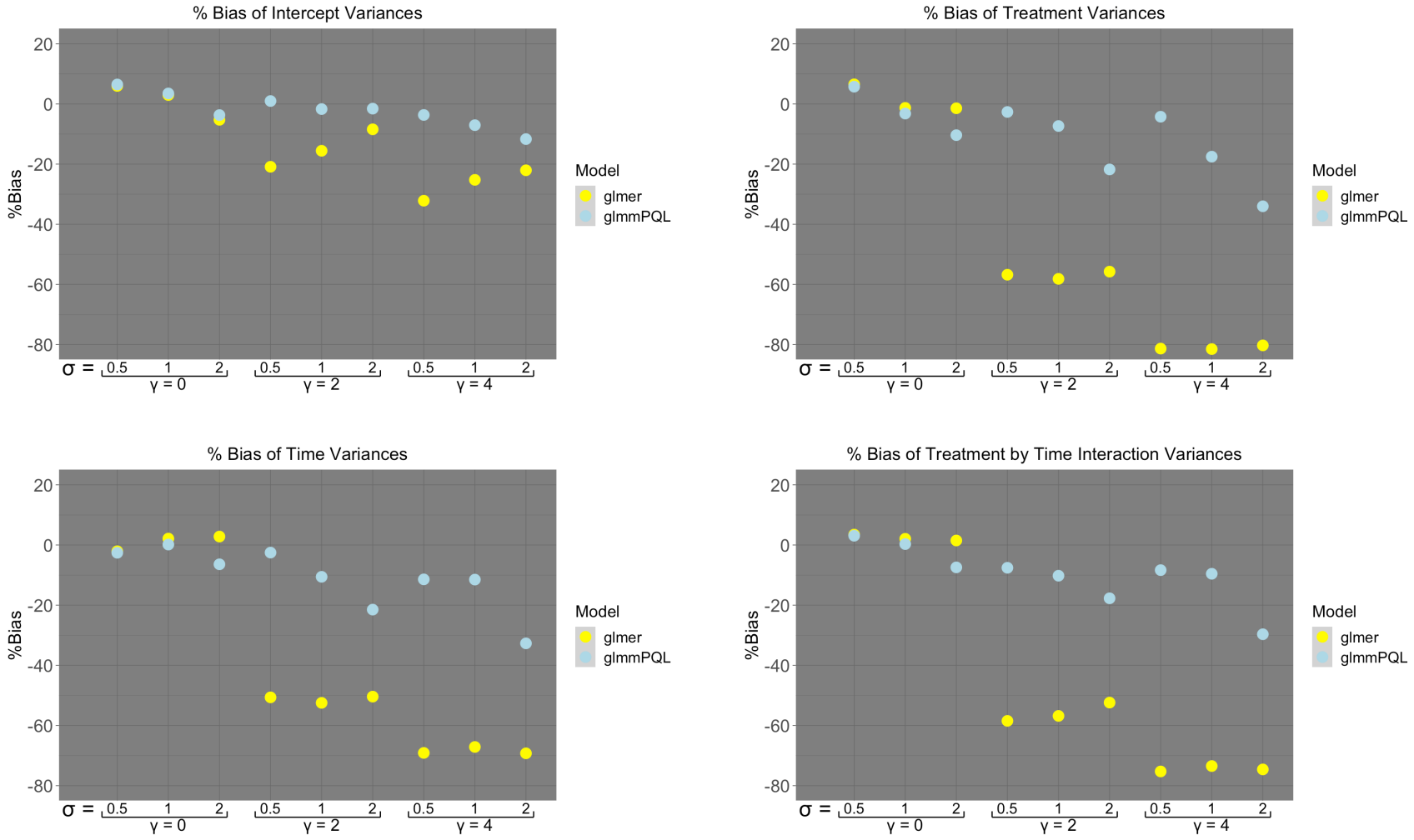


Figure 5.3: Plots showing change in bias of variances for intercept, time, treatment, and the treatment by time interaction as population heterogeneity and serial correlation change. The models fit are glmer and glmmPQL.



Ultimately, the study team observed that the glmer model did well when it was the correct model (no serial correlation), but even low amounts of serial correlation (gamma = 2), caused type 1 error rates to almost double, and as the serial correlation increased, the model performance continued to decline. The level of population heterogeneity, however, appears to have little effect on the glmer model's confidence interval coverage. For glmmPQL model, increasing gamma affects the model's performance far less than for glmer, but increasing population heterogeneity (σ) appears to have a detrimental effect on glmmPQL's performance. Looking at the simulated results for bias in the estimates, the same trend appears to hold, with the glmer model having estimates as high as 33% biased when gamma is 4. The glmmPQL model does better, but double-digit biases are still seen, such as -11.3% bias when gamma is 4 and population heterogeneity is 2.0. Both models, however, exhibit poor behavior for variance estimates. Even when the glmer model is correctly specified, the treatment bias for the variance is still 5.22. At worst, however, the glmer model has an estimated treatment bias of -81.5 (gamma = 4, RIntSD = 1.0). Again, glmmPQL does better, but the treatment bias for the estimate of the variance is still as high as -34.0.

Chapter 6

Discussion

The study team found that, on average, over the course of the full 24 hours of the day, no statistically significant differences in the hypoglycemic risk profile between Tresiba insulin and CSII were detected. However, blood glucose values vary considerably across the day, and different patients are at risk of hypoglycemia during different times. Therefore, looking only at the 24 hour effect does not allow clinicians to make as informed, personalized decisions about their patients' treatment plans as when the treatment effect is examined within a day.

Making more nuanced comparisons by examining the treatment effects for different times of the day reveals that the hypoglycemic risk profile between Tresiba insulin and CSII differs over the course of the day. During late night and early morning hours (8:30pm to midnight and 5:00am to 8:00am), Tresiba is associated with a lower risk of hypoglycemia than CSII, but from 1:00pm to 8:30pm, CSII has a lower associated risk of hypoglycemia. For a clinician trying to make personalized decisions about what treatment regimen to use, these differences are critical because different patients have different concerns and needs. As an example, active individuals doing recreational activities throughout the day might be much more concerned about minimizing their risk of low blood sugars during the hours they're awake or when they're performing than in the late evening or early morning. On the other hand, patients with hypoglycemic unawareness (a complication where an individual is no longer able to feel the symptoms of low blood sugars) might be much more concerned about minimizing nocturnal hypoglycemia. Therefore, averaging over the full 24 hours

inhibits the ability to make the most person-specific decisions about insulin treatment for different patients and patient populations.

Among the covariates controlled for in the study, the treatment order and Tandem insulin pump were statistically significant. For treatment order, this implies that there is a statistically significant difference in the hypoglycemic risk profile over the 24 hours of the day for individuals who received Tresiba insulin first as compared to those who did not. The study team attributes this to subject burnout. Though individuals are instructed not to change their behavior, subjects who were randomized to Tresiba first might have used the switch to a new treatment regimen (the inclusion criteria required subjects already be using CSII) as a "clean slate" to try to improve their diabetes control. If so, then paying closer attention to e.g. insulin dosing for meals could have reduced the amount of hypoglycemia observed. Regarding the Tandem insulin pump, because only six subjects used it, the result should be read with caution, and a larger sample size should be used to draw more reliable comparisons.

Once the study team had described the 24 hour risk profile for Tresiba, they sought to better understand how uncaptured variation in the study population might affect the outcome. Because so many factors affect blood glucose throughout the day, far fewer are able to be controlled for than not. A simple but important example is patient behavior. Two individuals can look identical in their demographic profiles and medication records, but their dietary choices can require that they have entirely different treatment regimens and hypoglycemic risk profiles. Therefore, exploring the effect that population hetero-

geneity (unmeasured, between-person variability) had on model results was an important clinical consideration. Specifically, is the hypoglycemic risk profile of a homogeneous population different than the hypoglycemic risk profile of a more heterogeneous population?

The simulation study examining population heterogeneity addressed this concern. The study team found that as population heterogeneity increases, the distribution of individuals experiencing hypoglycemia shifts left, becoming more right skewed. Intuitively, this makes sense, because increasing variation causes the distribution to spread out, but in this case, it cannot go below 0 (time spent below a blood glucose of 70 mg/dL cannot be negative). So, as heterogeneity increases, hypoglycemic risk spreads to the tails, which both increases the number of people experiencing no hypoglycemia and increases the number of people who experience extreme amounts of it.

This shows that it is essential to build context for study results evaluating the amount of time spent below a glucose threshold. The variation in the random intercept is, by definition, residual variability in the study population beyond what can be controlled for, and because it affects the risk of hypoglycemia, it should be considered when making statements about the generalizeability of study results. Beyond this study, it is also an important consideration when advisory boards make recommendations for treatment regimens for minimizing hypoglycemic risk. As this simulation study showed, the risk profile can be entirely different when only the population heterogeneity changes, so clinicians should know that the hypoglycemic risk profile of their patient population could be different from what a well-designed clinical trial shows. Conveniently, the study population considered for

this analysis was designed to be as homogeneous as possible (estimated standard deviation fo the random intercepts was 0.72

Once the study team had built context for interpreting the model results in terms of population heterogeneity, the final concern was to build context for understanding the accuracy of the model by examining its characteristics. Specifically, because the subjects' glucose measurements were taken so closely in time, determining how much that correlation affects the study results became an important consideration. Of the three models specified, glmer, glmmPQL, and glmer with bootstap, the glmmPQL model had the least bias and best confidence interval coverage. Despite bootstrapping the glmer model, the assumption of no serial correlation proved too strong of a misspecfication to recover good confidence interval coverage or reduce model bias. This could be due to the fact that the data were bootstrapped at the subject level, so even though the variability of the bootstrap samples changed, the correlation between glucose measurements never did, and every glmer model fit assumed it was 0. That could also explain why the bias for the variance of the intercept was so much lower in the bootstrapped glmer model than for the non-bootstrapped model. From one bootstrap sample to the next, the variance for the intercept was able to change, and the bootstrap could therefore approximate it.

To further investigate the properties of the glmer and glmmPQL models, both serial correlation and the variance of the random intercepts were changed, and the models were fit again (excluding the glmer bootstrap). This sought to do two things. First, by changing the serial correlation, it allowed the team to assess how misspecified the glmer and glmmPQL

models are in terms of bias and confidence interval coverage. Second, it allowed the study team to disentangle whether model misspecification was due more to serial correlation or population heterogeneity.

The results show that for even low levels of serial correlation, the bias in the variance of the glmer model is already large (over 11% off for all parameters) and the type 1 error rate is inflated. As the serial correlation increases, the bias only becomes worse and the type 1 error grows. Both cases make intuitive sense - the glmer model does not specify serial correlation, and therefore it fails to account for a non-zero source of variation. By assuming there's less variation than there actually is, the standard errors are too narrow and the model underestimates the true parameters. Conversely, the glmmPQL model fit does well with all levels of serial correlation, but as the population heterogeneity (RIntSD) increases, the glmmPQL model begins to underestimate the standard errors and the type 1 error rate is too high.

These simulation study results reveal the challenges associated with analyzing these kind of complex data in R. As mentioned earlier, this study population was designed to be homogeneous, so while the glmmPQL model fit has inflated type 1 error rates, the problem was minimized by a strong study design, and the study team had specified the best model based on the simulation results.

Chapter 7

Conclusion

This study estimated and profiled the hypoglycemic risk associated with Tresiba insulin as compared to continuous subcutaneous insulin infusion (CSII). By using Dexcom continuous glucose monitors (CGMs) to measure blood glucose on the study subjects, the study team was not only able to quantify the effects of Tresiba on average over 24 hours, but also able to look at any time interval (down to 5 minutes) throughout the day. Importantly, the study team found that while no statistically significant differences were detected between Tresiba and CSII across the full 24 hours, there were in fact differences in the risk of hypoglycemia in the early morning hours (midnight to 5:00am), midday hours (1:00pm to 8:30pm) and the later evening hours (8:30pm to midnight). By offering a nuanced understanding of the risk profile of Tresiba, the study team hopes to allow physicians and patients to make more informed decisions about their diabetes treatment regimen.

The study team also built context for understanding how the concentration of risk in a population changes as between-person variability (population heterogeneity) increases. Recognizing that the hypoglycemic risk profile for a population is tied to variation beyond what a model can control for is especially important in diabetes studies due to the plethora of factors that affect an individual's blood glucose. Specifically, this finding plays an important role in study design. By allowing a study population to be more heterogeneous, the observed hypoglycemic risk will be more extreme than for a homogeneous population. Therefore, the statistical analysis should be appropriately specified to ensure important differences and variations are not averaged out, especially if time spent in hypoglycemia is a study endpoint.

Lastly, the study team sought to better understand how characteristics of the data, specifically serial correlation and between-person variability, affected different model fits. The simulation study showed that the type 1 error was inflated, even when specifying a model that accounts for serial correlation. Increasing serial correlation and population heterogeneity exacerbated bias, especially in the generalized linear mixed effects model that could not specify auto-correlation, which completely missed the dependence structure in the data. Though the study team did not find an immediate solution to this problem in R, they hoped to draw attention to the issue for study analysts who work with these kind of complex data, and they encourage careful examination of model results before drawing inferences or making conclusions.

Bibliography

Battelino, Tadej et al. (Aug. 2019). "Clinical Targets for Continuous Glucose Monitoring Data Interpretation: Recommendations From the International Consensus on Time in Range". In: *Diabetes Care* 42.8, pp. 1593–1603. ISSN: 0149-5992, 1935-5548. DOI: 10.2337/dci19-0028. URL: http://care.diabetesjournals.org/lookup/doi/10.2337/dci19-0028 (visited on 09/07/2019).

Breslow, Norman (2004). "Whither PQL?" In: *Proceedings of the Second Seattle Symposium in Biostatistics*. Springer, pp. 1–22.

Breslow, Norman E and David G Clayton (1993). "Approximate inference in generalized linear mixed models". In: *Journal of the American statistical Association* 88.421, pp. 9–25.

Cryer, Philip E. (Apr. 2, 2007). "Hypoglycemia, functional brain failure, and brain death". In: *Journal of Clinical Investigation* 117.4, pp. 868–870. ISSN: 0021-9738. DOI: 10.1172/JCI31669. URL: http://www.jci.org/cgi/doi/10.1172/JCI31669 (visited on 12/23/2019).

Everitt, Brian and David C. Howell, eds. (2005). *Encyclopedia of statistics in behavioral science*. Hoboken, N.J: John Wiley & Sons. 4 pp. ISBN: 978-0-470-86080-9.

Fowler, Michael J (2008). "Microvascular and Macrovascular Complications of Diabetes". In: *Clinical Diabetes* 26.2, p. 6.

Harrell Jr, Frank E (2015). *Regression modeling strategies: with applications to linear models, logistic and ordinal regression, and survival analysis*. Springer.

Kalra, Sanjay and Yashdeep Gupta (Mar. 2015). "Clinical use of Insulin Degludec: Practical Experience and Pragmatic Suggestions". In: *North American Journal of Medical Sciences* 7.3, pp. 81–85. ISSN: 2250-1541. DOI: 10.4103/1947-2714.153918. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4382770/ (visited on 12/23/2019).

Pickup, J. (Mar. 23, 2002). "Glycaemic control with continuous subcutaneous insulin infusion compared with intensive insulin injections in patients with type 1 diabetes: meta-analysis of randomised controlled trials". In: *BMJ* 324.7339, pp. 705–705. ISSN: 09598138, 14685833. DOI: 10.1136/bmj.324.7339.705. URL: http://www.bmj.com/cgi/doi/10.1136/bmj.324.7339.705 (visited on 11/16/2019).

Pinheiro, José C and Edward C Chao (Mar. 2006). "Efficient Laplacian and Adaptive Gaussian Quadrature Algorithms for Multilevel Generalized Linear Mixed Models". In: *Journal of Computational and Graphical Statistics* 15.1, pp. 58–81. ISSN: 1061-8600, 1537-2715. DOI: 10.1198/106186006X96962. URL: https://www.tandfonline.com/doi/full/10.1198/106186006X96962 (visited on 12/07/2019).

Rodbard, David (June 1, 2017). “Continuous Glucose Monitoring: A Review of Recent Studies Demonstrating Improved Glycemic Outcomes”. In: *Diabetes Technology & Therapeutics* 19 (S3), S–25. ISSN: 1520-9156. DOI: 10.1089/dia.2017.0035. URL: https://www-liebertpub-com.proxy.library.vanderbilt.edu/doi/full/10.1089/dia.2017.0035 (visited on 09/07/2019).

Schaepelynck, Pauline et al. (June 2011). “A Recent Survey Confirms the Efficacy and the Safety of Implanted Insulin Pumps During Long-Term Use in Poorly Controlled Type 1 Diabetes Patients”. In: *Diabetes Technology & Therapeutics* 13.6, pp. 657–660. ISSN: 1520-9156, 1557-8593. DOI: 10.1089/dia.2010.0209. URL: http://www.liebertpub.com/doi/10.1089/dia.2010.0209 (visited on 11/16/2019).

Tambascia, Marcos Antonio and Freddy Goldberg Eliaschewitz (Dec. 2015). “Degludec: the new ultra-long insulin analogue”. In: *Diabetology & Metabolic Syndrome* 7.1, p. 57. ISSN: 1758-5996. DOI: 10.1186/s13098-015-0037-0. URL: https://dmsjournal.biomedcentral.com/articles/10.1186/s13098-015-0037-0 (visited on 12/23/2019).

Thomas, Danne, Nimri Revital, Battelino Tadej, et al. (2017). “International Consensus on Use of Continuous Glucose Monitoring”. In: 40, p. 10.

Tibshirani, Ryan (2014). *Advanced Methods for Data Analysis: Spring 2014*. URL: https://www.stat.cmu.edu/~ryantibs/advmethods/notes/smoothspline.pdf (visited on 12/15/2019).

Weinstock, Ruth S. MD PhD (Nov. 2019). “Management of blood glucose in adults with type 1 diabetes mellitus”. In: ed. by Ted. W. Post. Waltham, MA": UpToDate.

Wood, Simon N (2017). *Generalized additive models: an introduction with R*. Chapman and Hall/CRC.

A.Appendix

Table A.1: Operating characteristics for glmer, glmmPQL, and bootstrapped glmer models

	Intercept = 0.50	**Treatment = −0.50**	**Time = 0.25**	**Treatment:Time = 0.25**
model	$\hat{\beta}$ $s^2_{\hat{\beta}}$ Bias_β Bias_{var} CI coverage	$\hat{\beta}$ $s^2_{\hat{\beta}}$ Bias_β Bias_{var} CI coverage	$\hat{\beta}$ $s^2_{\hat{\beta}}$ Bias_β Bias_{var} CI coverage	$\hat{\beta}$ $s^2_{\hat{\beta}}$ Bias_β Bias_{var} CI coverage
glmer	-0.68 0.009 33.4 -31.6 42.5	0.66 0.012 31.7 -82.2 24.8	0.33 0.008 30.4 -68.6 57.8	0.33 0.021 33.9 -74.6 58.5
PQL	-0.48 0.004 -4.07 -2.22 92.4	0.47 0.005 -4.65 -9.02 92.5	0.24 0.004 -3.30 -7.02 93.8	0.25 0.011 -1.83 -8.11 93.2
bootstrap	-0.69 0.010 37.6 -3.60 50.8	0.68 0.013 35.7 -83.06 22.8	0.34 0.010 34.6 -70.9 54.2	0.34 0.022 36.6 -75.5 56.4

Table A.2: Confidence interval coverage for glmer and glmmPQL models

model	gamma	RIntSD	Intercept	Treatment	Time	Treatment:Time
glmer	0	0.5	0.956	0.958	0.948	0.952
glmer	0	1.0	0.952	0.946	0.953	0.952
glmer	0	2.0	0.940	0.950	0.954	0.959
glmer	2	0.5	0.813	0.712	0.829	0.790
glmer	2	1.0	0.890	0.732	0.806	0.794
glmer	2	2.0	0.938	0.761	0.815	0.817
glmer	4	0.5	0.425	0.248	0.586	0.590
glmer	4	1.0	0.602	0.289	0.602	0.603
glmer	4	2.0	0.794	0.377	0.628	0.628
pql	0	0.5	0.924	0.947	0.946	0.952
pql	0	1.0	0.938	0.925	0.946	0.947
pql	0	2.0	0.926	0.915	0.938	0.944
pql	2	0.5	0.936	0.940	0.948	0.944
pql	2	1.0	0.905	0.898	0.935	0.932
pql	2	2.0	0.912	0.882	0.913	0.922
pql	4	0.5	0.930	0.929	0.933	0.944
pql	4	1.0	0.876	0.862	0.934	0.940
pql	4	2.0	0.896	0.838	0.887	0.892

Table A.3: Average Estimates of $\hat{\beta}$ for glmer and glmmPQL models

model	gamma	RIntSD	Intercept $\beta = -0.50$	Treatment $\beta = 0.50$	Time $\beta = 0.25$	Treatment:Time $\beta = 0.25$
glmer	0	0.5	-0.50	0.50	0.25	0.25
glmer	0	1.0	-0.50	0.50	0.25	0.25
glmer	0	2.0	-0.50	0.50	0.25	0.25
glmer	2	0.5	-0.54	0.54	0.27	0.27
glmer	2	1.0	-0.54	0.54	0.27	0.27
glmer	2	2.0	-0.53	0.53	0.27	0.27
glmer	4	0.5	-0.66	0.66	0.32	0.34
glmer	4	1.0	-0.66	0.65	0.32	0.33
glmer	4	2.0	-0.65	0.64	0.32	0.32
pql	0	0.5	-0.49	0.49	0.24	0.24
pql	0	1.0	-0.48	0.48	0.24	0.24
pql	0	2.0	-0.47	0.48	0.24	0.24
pql	2	0.5	-0.49	0.48	0.24	0.25
pql	2	1.0	-0.46	0.47	0.24	0.24
pql	2	2.0	-0.45	0.47	0.24	0.24
pql	4	0.5	-0.48	0.48	0.24	0.25
pql	4	1.0	-0.45	0.45	0.24	0.24
pql	4	2.0	-0.44	0.46	0.24	0.24

Table A.4: Bias in β Estimates for glmer and glmmPQL models

model	gamma	RIntSD	Intercept	Treatment	Time	Treatment:Time
glmer	0	0.5	-0.3	-0.3	0.1	0.0
glmer	0	1.0	0.3	-0.1	0.2	-0.5
glmer	0	2.0	-0.1	0.1	-0.3	0.2
glmer	2	0.5	8.9	8.6	7.2	9.6
glmer	2	1.0	7.2	7.1	6.7	8.4
glmer	2	2.0	6.6	6.3	6.5	7.6
glmer	4	0.5	33.0	31.5	29.7	34.7
glmer	4	1.0	31.5	30.2	29.7	31.4
glmer	4	2.0	30.2	28.1	28.8	29.3
pql	0	0.5	-2.8	-2.8	-2.3	-2.6
pql	0	1.0	-3.7	-3.6	-3.3	-4.2
pql	0	2.0	-5.9	-3.4	-3.6	-3.4
pql	2	0.5	-2.8	-3.1	-3.2	-1.8
pql	2	1.0	-7.2	-6.8	-5.2	-5.0
pql	2	2.0	-9.5	-6.3	-4.5	-4.8
pql	4	0.5	-4.3	-4.6	-3.9	-1.1
pql	4	1.0	-9.5	-9.4	-5.4	-4.4
pql	4	2.0	-11.3	-8.2	-2.7	-2.7

Table A.5: Estimated Variance (multiplied by 100) of $\hat{\beta}$ for glmer and glmmPQL models

model	gamma	RIntSD	Intercept	Treatment	Time	Treatment:Time
glmer	0	0.5	0.13	0.16	0.20	0.40
glmer	0	1.0	0.29	0.18	0.23	0.45
glmer	0	2.0	0.92	0.24	0.30	0.59
glmer	2	0.5	0.22	0.17	0.22	0.44
glmer	2	1.0	0.41	0.19	0.24	0.48
glmer	2	2.0	1.14	0.25	0.31	0.63
glmer	4	0.5	0.58	0.21	0.27	0.53
glmer	4	1.0	0.89	0.23	0.29	0.58
glmer	4	2.0	2.04	0.30	0.37	0.74
pql	0	0.5	0.13	0.15	0.19	0.38
pql	0	1.0	0.27	0.17	0.21	0.41
pql	0	2.0	0.80	0.20	0.25	0.50
pql	2	0.5	0.22	0.30	0.32	0.75
pql	2	1.0	0.34	0.30	0.32	0.75
pql	2	2.0	0.82	0.34	0.36	0.82
pql	4	0.5	0.37	0.44	0.36	0.98
pql	4	1.0	0.48	0.43	0.36	0.96
pql	4	2.0	0.94	0.46	0.40	1.07

Table A.6: True Variance (times 100) of $\hat{\beta}$ for glmer and glmmPQL models

model	gamma	RIntSD	Intercept	Treatment	Time	Treatment:Time
glmer	0	0.5	0.12	0.15	0.21	0.39
glmer	0	1.0	0.28	0.18	0.22	0.44
glmer	0	2.0	0.98	0.24	0.29	0.58
glmer	2	0.5	0.28	0.40	0.45	1.05
glmer	2	1.0	0.48	0.46	0.51	1.12
glmer	2	2.0	1.24	0.57	0.63	1.31
glmer	4	0.5	0.86	1.13	0.86	2.12
glmer	4	1.0	1.20	1.27	0.89	2.18
glmer	4	2.0	2.62	1.52	1.21	2.90
pql	0	0.5	0.12	0.14	0.20	0.37
pql	0	1.0	0.26	0.17	0.21	0.41
pql	0	2.0	0.83	0.23	0.27	0.54
pql	2	0.5	0.21	0.31	0.33	0.81
pql	2	1.0	0.35	0.33	0.36	0.84
pql	2	2.0	0.83	0.43	0.46	1.00
pql	4	0.5	0.39	0.46	0.41	1.07
pql	4	1.0	0.52	0.52	0.40	1.07
pql	4	2.0	1.06	0.70	0.60	1.52

Table A.7: Bias in variance of $\hat{\beta}$ for glmer and glmmPQL models

model	gamma	RIntSD	Intercept	Treatment	Time	Treatment:Time
glmer	0	0.5	6.0	6.5	-2.1	3.5
glmer	0	1.0	3.0	-1.3	2.1	2.1
glmer	0	2.0	-5.3	-1.4	2.8	1.5
glmer	2	0.5	-20.9	-56.8	-50.6	-58.5
glmer	2	1.0	-15.6	-58.2	-52.5	-56.8
glmer	2	2.0	-8.4	-55.7	-50.4	-52.4
glmer	4	0.5	-32.2	-81.3	-69.1	-75.3
glmer	4	1.0	-25.2	-81.5	-67.1	-73.5
glmer	4	2.0	-22.0	-80.3	-69.2	-74.6
pql	0	0.5	6.5	5.7	-2.6	3.1
pql	0	1.0	3.5	-3.2	0.1	0.3
pql	0	2.0	-3.7	-10.4	-6.4	-7.4
pql	2	0.5	1.0	-2.7	-2.5	-7.6
pql	2	1.0	-1.7	-7.3	-10.6	-10.2
pql	2	2.0	-1.6	-21.8	-21.5	-17.7
pql	4	0.5	-3.7	-4.2	-11.4	-8.3
pql	4	1.0	-7.1	-17.5	-11.5	-9.6
pql	4	2.0	-11.7	-34.0	-32.7	-29.7