**Analyzing the Personalized Factors for Better Blood Glucose Control for a Type 1 Diabetic**

**Scott Wardle - University of Utah, Department of Biomedical Informatics**

**March 26, 2025**

**Abstract**

Managing Type 1 Diabetes requires consistent monitoring and many adjustments to insulin, diet, and lifestyle. The timing of insulin doses, meals, physical activity, and other daily routines plays a critical role in staying within a healthy blood glucose (BG) target range. A structured daily schedule and healthy routines - whether at home, at school or at work - helps diabetics maintain that control by ensuring consistency and oversight in accurate insulin delivery, carbohydrate intake, and exercise, which minimizes fluctuations in BG levels.

This paper explores the factors that explain and predict glycemic control in a healthy target range. It evaluates the relationships between the types of daily schedules and activity routines to gain an understanding about how those schedules can influence glycemic health to act on these factors on a personalized, prescriptive basis.

For this analysis a few months of data was gathered from continuous glucose monitors (CGMs) and insulin pumps of two diabetic patients. Two sample t-tests on the mean BG between different daily activity schedule groups were performed and found to be significant for times at school or work versus at home. The data was also to create personalized predictive models. These Logistic Regression Models predict with between 93-95% accuracy whether the targeted healthy BG range will be achieved a few hours into the future.

This knowledge can lead to better decisions now for improved control, fewer complications, and better overall health outcomes in the future for individuals with Type 1 Diabetes.

**Introduction and Background**

Managing Type 1 Diabetes involves constant learning and daily challenges, especially fo adolescents dealing with body changes and desire to fit in. As a parent of multiple children with diabetes, I deeply value reliable information and innovative products from companies like Tandem Diabetes and Dexcom that help families manage the disease. Guidance from endocrinologists have been invaluable, as have insightful articles from trusted sources on strategies for better BG control.

Patients with diabetes are encouraged to meet with their endocrinologist quarterly to assess their health and make needed corrections as they grow and as the disease progresses. Common adjustments needed include the insulin ‘basal’ rate (the amount the body needs, even without food), the food ‘bolus’ amount (insulin to carbohydrate ratio), and the correction ‘bolus’ rate (insulin required to lower high BG levels). These adjustments must also account for time-of-day variations, as insulin needs vary during the day. Additionally, diabetics must consider the effects of exercise (or lack thereof), stress, hormonal changes, and illness on their BG levels.

For diabetics and their caregivers, daily decisions about factors affecting BG levels are made with the goal of targeting a healthy BG range in the coming hours. However, several challenges can derail BG control—many of the same obstacles faced by those trying to stick to a weight loss goal. A lack structure in daily routines, loss of motivation and constant access to unhealthy food can make it difficult to maintain a healthy BG level. Additionally, for children and adolescents, the absence of qualified adult supervision while away from home can further complicate diabetes management.

Fortunately, numerous reliable resources offer valuable guidance to patients and parents of children with diabetes on managing the disease in healthy ways. One such resource is a Mayo Clinic article on diabetes management, which highlights the impact of lifestyle and daily routines on blood sugar levels. It emphasizes the importance of planning balanced meals and carbohydrate counting, along with the benefits of regular exercise and the need to monitor BG levels prior to exercise. Additionally, the article underscores the importance of proper medication management and being prepared for illness.

Another excellent resource which recommends the importance of setting daily routines is PEAQ Medical2.It provides a step-by-step guide for managing diabetes, highlighting how a well-structured daily routine can improve BG control. Key aspects include meal planning, regular exercise and consistent monitoring, all of which to maintaining healthy BG levels.

Many studies have explored data-driven modeling and BG prediction in diabetes management, including a paper published in the Elsevier Journal3 (linked from PubMed). This paper highlights advancements in Continuous Glucose Monitoring (CGM) technologies and potential for machine learning algorithms to use patient history data for more accurate BG predictions. However, developing a universal model that consistently produces accurate predictions for all patients and circumstances remains a challenge.

Adolescents with diabetes and their caregivers should consult reliable sources to better understand the medical devices they rely on for BG management. A great example is the products page on the Tandem Diabetes website4, which provides detailed guidance on using their devices effectively and getting needed support. Being well-versed in these features is essential for maintaining good control.

***What is already being done***

Most CGMs effectively alert users to high and low BG levels. They analyze monitor data to generate real-time graphs and detect daily patterns, helping diabetics and their caregivers anticipate and correct BG fluctuations before they lead to hyperglycemic or hypoglycemic events.

Some CGMs can also integrate with insulin pumps to enhance glucose management. When BG levels are high, they can automatically deliver a correction bolus, and when levels are low or dropping rapidly, they can pause basal insulin to prevent further decline.

As mentioned in the Elsevier article, new and emerging technologies have given rise to a ‘closed-loop’ Artificial Pancreas. This system integrates a CGM, which measures BG every 5 minutes, with an insulin pump and an algorithm that determines proper insulin delivery in real time. This is an exciting development, as it allows for automatic insulin dosing to lower BG, and some systems may even administer glucagon to raise BG when needed. Over the past few years, these technologies have undergone extensive regulatory approvals to ensure their safety and effectiveness.

One of the most challenging aspects of accurate insulin delivery remains is accounting for variations in carbohydrate intake. While there are excellent apps and tools to help diabetics track nutritional information—especially carbohydrate content—accurate insulin dosing still depends on correctly estimating the amount of carbohydrates consumed in advance. Maintaining consistent portion sizes, such as using individual serving sizes for common snacks, can help improve accuracy in insulin calculations and overall glucose management.

**Research Questions**

In this study, we investigate the personalized factors that are already available in the technology in use by most diabetics today which can be used to improve their BG outcomes and overall health. We hypothesize that a diabetics daily schedule impacts BG levels. Specifically, we will test whether the mean BG level is the same during school or work hours versus at home, against the alternative hypothesis that there is a difference between those times.

Furthermore, we propose that by incorporating additional customized settings and user inputs into CGMs and insulin pumps, diabetics can enhance their devices ability to optimize insulin delivery and other features. By entering key details—such as their schedule, typical eating and exercise patterns and weight—these devices could generate a personalized prediction model that is both safe and accurate. This model could then be automatically fine-tuned to improve BG levels in real time.

To demonstrate this, we create two logistic regression models, one for each patient using their individual data and known personal details about their daily schedule. Our analysis evaluates whether their BG levels remain within the target range of 70–150 mg/dL one to three hours later.

**Methodology**

**Study Objective and Design**

The objective of this study is twofold:

* **Comparative Analysis:** Evaluate the differences in blood glucose (BG) levels between different activity locations (At School vs. Not at School) using hypothesis testing.
* **Predictive Modeling:** Develop a logistic regression model using data from a CGM and insulin pump to predict whether blood glucose will be in the target range within the next few hours.

This study leverages CGM data, insulin delivery records and carbohydrate intake logs to assess real-world variations in BG control and explore predictive capabilities. The design is observational, as it relies on naturally occurring data collected from insulin pumps rather than a controlled intervention. Data is retrospectively analyzed to understand patterns and relationships in BG fluctuations.

**Data Sources and Variable Description**

The datasets consist of BG readings taken every 5 minutes from a CGM over a two-month period for two patients. It also contains insulin pump event data over the same period and includes a few important raw independent variables including:

* **BG (Blood Glucose):** A primary variable measured in mg/dL.
* **Insulin Delivery:** The amount of insulin administered at each time point.
* **Carbohydrate Intake:** The amount of carbohydrates consumed (in grams).

Additionally, the following feature-engineered independent variables were created to facilitate some other useful predictive variables. These include the following:

* **Recent Blood Glucose Trends:** Rolling averages over different time windows.
* **Recent Insulin Delivery & Carbohydrate Intake:** Amounts of recent insulin/carb intake.
* **Activity Locations:** Categorized as "At School" or "Not at School" (home or other locations).
* **Missed Food Bolus:** Categorical variable -- if no food bolus was entered in over 6 hours during daytime.
* **Time Features:** Whether the reading occurred on a weekday, weekend, or holiday.
* **Lagged Features:** Rolling averages of insulin, and carbohydrate intake over previous time windows.

These datasets allow for both comparative analysis (hypothesis testing) and predictive modeling (logistic regression).

**Software and Packages Used**

This analysis was conducted using Python in a Jupyter Notebook. This allowed for an interactive way to do data exploration and modeling. The following libraries and tools were used:

* **Data Manipulation and processing**
  + **Pandas**: Used for loading, cleaning and manipulating the datasets. This includes the creation of several engineered features that were needed.
  + **NumPy**: Used for handling numerical operations including calculations related to rolling averages.
* **Statistical Analysis**
  + **SciPy**: For conducting hypothesis testing, specifically the two-sample t-test to compare blood glucose levels between school/work and non-school/work conditions.
  + **Statsmodels:** Used for the estimation of statistical logistic regression modeling, for providing statistical summaries and diagnostics.
  + **Scikit-learn:** Used for data pre-processing including scaling, performance evaluation with accuracy, precision, recall and ROC analysis
* **Data Visualizations**
  + **Matplotlib and Seaborn:** Used to create visual diagrams of the data including histograms, boxplots, time-series plots and ROC curves for model evaluation.

**Statistical Analysis**

To examine differences in BG levels based on activity location, we define the following hypotheses:

* **Null Hypothesis (H0):** The average blood glucose levels are equal for observations when the patient is (At School) vs. (Not at School).
* **Alternative Hypothesis (H1):** The average blood glucose levels differ between the two conditions.

A **paired** **two-sample t-test** was conducted on a sample of the data to assess whether there is a significant difference in BG levels between these two groups.

A **logistic regression model** was built to predict whether the patient's blood glucose will be in the target range in the next few hours based on current and historical data. The target variable is defined as:

* Y = 1, if BG is in Target Range for more than 50% of the observations in window 1 to 3 hours in future.
* Y = 0, otherwise

A training/testing split is performed:

* The **first month** of data is used for training.
* The **second month** is used for testing and evaluation.

Model evaluation performance is assessed using:

* **Accuracy** (train vs. test performance)
* **Precision, Recall, and F1-score**
* **AUC-ROC Curve** (to measure classification performance)

Whenever overfitting is detected, we have used regularization and/or other selection adjustments.

**Ethical Considerations**

This study relies on patient health data, including BG levels and insulin administration. Ensuring privacy is a top priority. To protect sensitive information:

All personally identifiable information was removed before performing analysis. Data storage and processing were conducted in a secure environment. While logistic regression can identify patterns and predict future BG levels, they should not replace clinical judgment. Instead, these models:

* Serve only as decision support tools to help patients/parents and healthcare providers and make informed personalized choices.
* Can identify potential high-risk scenarios, prompting further medical evaluation.
* May reduce the burden of manual monitoring, allowing for more proactive adjustments in diabetes management.

**Results**

Patient 1 is a male adolescent high school student that has had Type 1 Diabetes for 8 years. For this patient, there were 16,349 BG measurements taken from his Dexcom CGM every 5 minutes over a two-month period from November 2024 to January 2025. Patient 2 is a male young working adult that has had Type 1 Diabetes for 22 years. He had 16,499 BG measurements taken from his Dexcom CGM also over a two-month period from January to March 2025.

**Descriptive Statistics**

For patient 1

* The mean BG was 199.02 with SD 63.26 during the 2-month observation period.
* The mean BG recorded during school hours was 204.17 with SD 48.22
* The mean BG recorded outside of school hours was 198.23 with SD 65.23

For patient 2

* The mean BG was 148.55 with SD 51.66 during the 2-month period
* The mean BG recorded during work hours was 159.11 with SD of 53.51
* The mean BG recorded outside of work hours was 145.04 with SD of 50.55

The distribution of blood glucose levels in both groups for patient 1 is shown in Figure 1

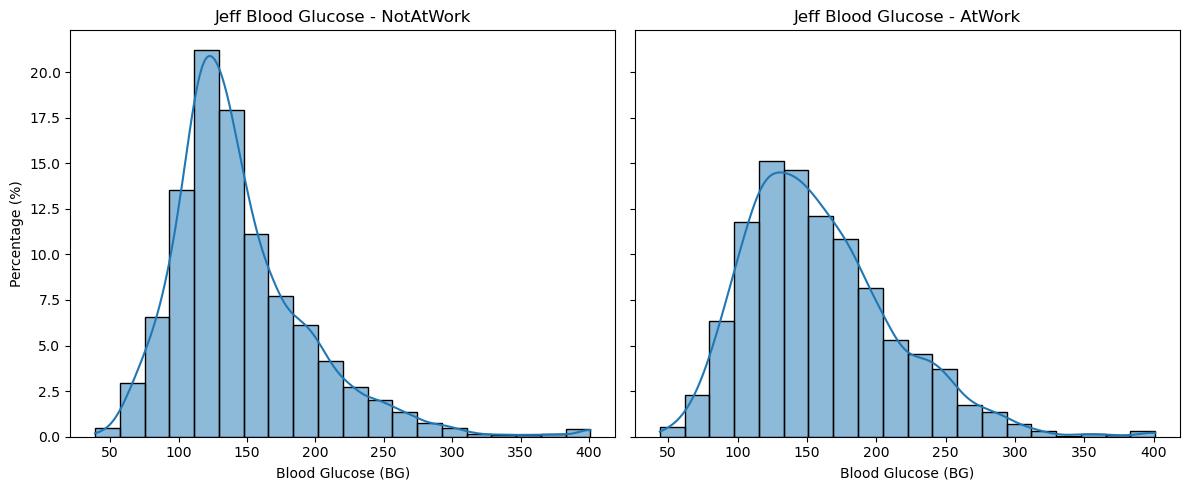
**Figure 1**

A graph of a graph of a graph

AI-generated content may be incorrect.

The distribution of blood glucose levels in both groups for patient 2 is shown in Figure 2

**Figure 2**



**Hypothesis Test Results**

A two-sample independent t-test with a random sample size of 500 per group was conducted to compare the means between the **At School** and **Not at School** groups for patient 1

* The Null Hypothesis is the mean BG from the **At School** measurement and the **Not At School** measurement are equal.
* The Alternative Hypothesis is that there is a difference in the mean BG between the locations.
* The test yielded a t-statistic of 3.27 with p-value of 0.0011
* Since the p-value is below 0.05, we therefore reject the Null Hypothesis and conclude that there is difference in mean BG for **At School** versus **Not at School**.

Another two-sample independent t-test with a random sample size of 500 per group was conducted to compare the means between the **At Work** and **Not at Work** groups for patient 2

* The Null Hypothesis is the mean BG from the **At Work** measurement and the **Not At Work** measurement are equal.
* The Alternative Hypothesis is that there is a difference in the mean BG between the locations.
* The test yielded a t-statistic of 4.02 with p-value of 0.0001
* Since the p-value is below 0.05, we therefore reject the Null Hypothesis and conclude that there is difference in mean BG for **At Work** versus **Not at Work**.

**Model Performance Metrics**

A logistic regression model was trained to predict whether patient 1’s BG would be in the target range (70-150 mg/dL) for more than half of the measurements one to three hours later. This time range is chosen due to the time effect of both insulin and glucose (from food) in the bloodstream normally peaks during those hours.

* The model was able to achieve:
  + A Psuedo-Rsquare of 0.7244
  + A Training accuracy of 0.939
  + A Testing accuracy of 0.953
  + A precision of 0.95
  + A recall of 0.91
  + An F1 value of 0.93
  + A ROC (AUC) Score of 0.988

Another logistic regression model was trained to predict whether patient 2’s BG would be in the target range for more than half of the measurements one to three hours later.

* The model was able to achieve:
  + A Psuedo-Rsquare of 0.7565
  + A Training accuracy of 0.940
  + A Testing accuracy of 0.936
  + A precision of 0.95
  + A recall of 0.96
  + A ROC (AUC) Score of 0.972
  + An F1 value of 0.96

The model ROC Curve for Patient 1 is shown below in Figure 3

Figure 3

A graph of a curve

AI-generated content may be incorrect.

**Main Analysis**

To prepare the data for the modeling exercise, all the features were created and analyzed for their relationships before being included in the model. Figure 4 below displays one of the heatmaps generated for Patient 1 after several variables were removed from the model.

Figure 4

A screenshot of a graph

AI-generated content may be incorrect.

Additionally, Variance Inflation Factor (VIF) analysis was conducted on the predictor variables to identify and address multicollinearity issues. By removing variables with high VIF values, remaining coefficients in the model become more stable and reliable, while reducing standard errors. A VIF greater than 10 indicates a problematic level of multicollinearity. Table 1 below presents the VIF analysis for Patient 1 prior to removing some factors.

Table 1

A screenshot of a computer program

AI-generated content may be incorrect.

Since the predictor variables in this dataset have different scales (e.g., raw BG values vs. categorical indicators), standardization was performed on the variables using a transformation to a z-scale. This ensures that each variable has a mean of 0 and standard deviation of 1, preventing those variables that have higher numeric variability from dominating the model.

Several iterations of the model were performed, each time looking at those variables that did not contribute to the model by examining and removing those with p-values greater than 0.05. The remaining variables in the model are shown in Table 2 below.

Table 2

**A screenshot of a computer screen

AI-generated content may be incorrect.**

This analysis shows that the key variables explaining the most variation in the model are Avg\_BG\_2hrs (Average BG during the past 2 hours), InBGTarget (Patient is currently in BG Target range), BG\_Delta\_120 (Change in BG level over last 2 hours), and Sum\_Insulin\_2hrs (Total Insulin bolus in last 2 hours).

Data for Patient 2’s modeling was prepared using the same process for Patient 1. While the top three variables remained consistent between the two patients, it is notable that different additional variables contributed to the model. Another interesting finding is that Patient 2’s model required fewer variables (5) compared to Patient 1’s model (9), yet the simpler model for Patient 2 achieved a higher Psuedo-RSquare. Table 3 below presents the final model for Patient 2.

Table 3A screenshot of a computer screen

AI-generated content may be incorrect.

**Conclusions**

**Summary of Key Results**

* The Logistic Regression model demonstrated a strong ability to predict future BG levels based solely on insulin pump and CGM data.
* There was a high accuracy on both training and testing data, indicating a good fit on future data.
* The hypothesis tests comparing blood glucose levels during school hours vs. non-school hours and work hours vs. non-work hours found statistically significant differences, suggesting environmental or behavioral factors impact glucose control.
* Correlation analysis highlighted several relationships between variables such as insulin delivery amounts, carbohydrate intake, and glucose levels, justifying their inclusion in the model.

**Interpretation**

* The findings confirm that real-time insulin pump and CGM data can be reliably used to predict future BG levels.
* The significance difference found between school or work and home may indicate varying stress or activity factors can impact BG outcomes.

**Comparison with Literature**

* Prior studies have demonstrated that machine learning models can assist in diabetes management. Although these models focus mostly on CGM data rather than logistic regression that incorporates insulin pump data as well.
* This study, unlike studies that incorporate other personal health data (i.e. heart rate, sleep patterns), relies on insulin pump and CGM metrics alone, which may limit its power.
* The results from this study align well with existing research on glucose level variability in different contexts (i.e. school vs. home). This demonstrates the importance of personalized diabetes management strategies.

**Limitations**

* The datasets are only from two patients, which limits generalizability. A larger and more diverse population would more fully demonstrate the proof of concept.
* As this is an observational study, causal relationships cannot be definitively established.

**Future Directions**

* Expand the analysis to include more patients across age groups and management strategies.
* Experiment with other machine learning techniques (i.e. decision trees, linear regression, etc.) to improve predictive accuracy.
* Incorporate behavioral and health factors (i.e. sleep, exercise, illness) to refine predictions.
* Develop real-time applications for insulin pump users based on these models to assist in proactive glucose management, giving feedback or suggestions for improved outcomes.

**Implications**

* This analysis supports the potential for using more predictive modeling in diabetes care. It can reinforce the need for data-driven personalized treatment.
* Findings could inform endocrinologists and other healthcare providers about key factors influencing glucose control for those with diabetes, especially among children and adolescents.
* If models like these are refined and validated, they could integrate into insulin pump software to provide early warnings and suggestions for avoiding hypo/hyperglycemia, reducing risks for patients.

**References**

1. Mayo Clinic. (January 6, 2024). *Diabetes management: How lifestyle, daily routine affect blood sugar*. Mayo Clinic. Retrieved [February 9, 2025], from <https://www.mayoclinic.org/diseases-conditions/diabetes/in-depth/diabetes-management/art-20047963>
2. Peaq Medical. (October 24, 2024). *Daily routines for managing diabetes: A step-by-step guide*. Peaq Medical. Retrieved [February 9, 2025], from <https://peaqmedical.com/daily-routines-for-managing-diabetes-a-step-by-step-guide>
3. Woldaregay, A. Z., Årsand, E., Walderhaug, S., Albers, D., Mamykina, L., Botsis, T., & Hartvigsen, G. (2019). Data-driven modeling and prediction of blood glucose dynamics: Machine learning applications in type 1 diabetes. *Artificial intelligence in medicine*, *98*, 109–134. <https://doi.org/10.1016/j.artmed.2019.07.007>
4. Tandem Diabetes. (n.d.). *Mobile apps*. Tandem Diabetes Care. <https://www.tandemdiabetes.com/products/software-apps/mobile-apps>