

Diabetes Case Study. Title to be decided for at the end.

Jeremy Chu

09/04/2021

Abstract

As the paper is in progress, not much can be written about the abstract. Note that this abstract was written after the first introduction was written and after the final blub. It reflects my second introduction instead. Originally I was planning on predicting blood sugar levels based on carb intake and insulin intake but with the way the data is structured I was getting completely opposite correlations. I am currently reworking my data to focus more on time and overnight sleep blood sugar levels when insulin intake and carb intake does not play a factor. What you see in the dicussion/statistical modelling section were my attempts at figuring out just what's going on and how I arrived at my conclusion. Any and all feedback is welcome!

Contents

1	Introduction	2
2	Intro - Revised	3
3	Data Source	4
4	Diabetes Terminology	4
5	About the Individual	6
6	Literature Review	6
7	Looking at the individual's performance	6
7.1	Eating Habits. Meal Times, Frequencies.	8
8	Model	10
8.1	Predicting bolus volume based on carb intake, time of day, basal amount	11
8.2	Causality	15

9 Discussion	15
9.1 Final Takeaways	15
9.2 Moving Forward	16
References	16

1 Introduction

Diabetes research is centered around two fields: Studying the factors leading to diabetes, and finding critical factors that impact the lives of diabetics. When approaching the latter topic, available research has been sparse and scattered, with some examining the relationships between meal frequency and meal timings with blood sugar levels (missing citation), and others comparing ECGs with CGMs to discover some previously unknown correlations. Data on the lives of diabetics is sparse, and even when institutions generously make their data public, the time period of data collection is often quite short (cite the open diabetics data github). Borrowing the foundational research questions laid upon by previous diabetes researchers, this project looks at a single case for an extended period of time. Rather than comparing the diabetic experience of multiple individuals for a few weeks, I instead opted to examine 1 individual for 9 months.

The individual in question is my fiancée, which makes the topic of consent much easier than otherwise. The aim of the project is two-fold. First and foremost, this is a project designed to derive value for my fiancée, hence the questions will be seeking answers practical in nature. Secondly, I hope to be able to contribute to other available research, and see how a more extended observation period compares to a shorter but wider scope. The project aims to address three key questions:

1. Does my fiancée's blood sugar level follow a pattern of time of day
2. What is the relationship between bolus intake versus carb intake
3. Do meal frequencies affect daily bolus intake volumes

By examining the data and testing for the correlations between the available parameters, I attempt to better illustrate how my fiancée's blood sugar level fluctuates on a day-to-day basis. Aiding the answering of these questions, regression models will be ran to predict:

1. Predict bolus volume based on carb intake, time of day, basal amount
2. Predict carb intake based on hour of the day
3. Predict blood sugar levels based on carb intake, time of day, basal amount
4. Predict blood sugar levels when sleeping (no carb intake) and look at the morning value

The goal here is to functionally help my fiancée better control her A1C levels by predicting whether she needs to correct for more insulin intake depending on the time of day and carbs consumed. In addition, predicting blood sugar levels based on carb intake, while a very simple regression that cannot draw conclusive evidence due to many other variables in a meal's nutritional value, can serve as a simple baseline for insulin intake. Moreover, an understanding on blood sugar levels based on meal timing, and blood sugar levels overnight, connects my research to a wider discourse

and ties back into *Insert Study Name Here*'s study especially. Ultimately, the goal is to personally derive value from this study, while at the same time offering the research community a unique set of findings alongside a long and extensive dataset to build upon.

2 Intro - Revised

Note that this introduction above is what I was aiming to do. With the results of my model now run it is looking quite impossible. Instead a better, revised introduction would be something like the blurb below. I am keeping the first introduction in case someone has feedback on how to approach my original idea, but I have currently pivoted my project to the direction outlined below

Diabetes research is centered around two fields: Studying the factors leading to diabetes, and finding critical factors that impact the lives of diabetics. When approaching the latter topic, available research has been sparse and scattered, with some examining the relationships between meal frequency and meal timings with blood sugar levels (missing citation), and others comparing ECGs with CGMs to discover some previously unknown correlations. Data on the lives of diabetics is sparse, and even when institutions generously make their data public, the time period of data collection is often quite short (cite the open diabetics data github). Borrowing the foundational research questions laid upon by previous diabetes researchers, this project looks at a single case for an extended period of time. Rather than comparing the diabetic experience of multiple individuals for a few weeks, I instead opted to examine 1 individual for 9 months.

The individual in question is my fiancée, which makes the topic of consent much easier than otherwise. The aim of the project is two-fold. First and foremost, this is a project designed to derive value for my fiancée, hence the questions will be seeking answers practical in nature. Secondly, I hope to be able to contribute to other available research, and see how a more extended observation period compares to a shorter but wider scope. The project aims to address three key questions:

1. Does my fiancée's blood sugar level follow a pattern of time of day
2. Does eating before sleep influence whether she wakes up with a high/in range/low blood sugar
3. Does taking insulin right before bed influence whether she wakes up with a high/in range/low blood sugar

The questions are structured around time rather than insulin and carb intake throughout the day due to the limitations of the data. As will be further detailed in the data source section, the data used for analysis is aggregated by hour, where insulin intake numbers increase when blood sugar level rises. This results in unfortunate regression lines where it gives off the illusion that the more insulin the subject takes, the higher her blood sugar. The project will therefore first go over and illustrate why the data is unfeasible to perform predictions on daytime blood sugar levels before analyzing the overnight changes of blood sugar in the subject.

One of the primary worries for diabetics is waking up in the middle of the night with a low blood sugar, or waking up and realizing that they spent the night with a high blood sugar and start off the day groggy and tired. Rather than focusing my efforts on predicting carb to insulin ratios, where the now present CGM and insulin pump combinations already do with their in-built algorithms, this project aims to dissect what happens when the subject sleeps, and hopes to offer better insight into what to do to wake up in range. An understanding on the nightly shifts of blood sugar levels

Table 1: Original Insulin Data

Field Name	Description
Time	Time data was logged
Basal Amount	Amount of basal taken at the time of data log
Bolus Type	Whether bolus was fast-acting or normal
Bolus Volume	Amount of bolus taken at the time of data log
Immediate Volume	A setting on the pump where insulin can be quickly administered with a button press
Extended Volume	Type of bolus taken where users can take some insulin up front and have the rest administered later
Duration	No data
Carbs	Carbs taken at the time of data log
Total Daily Dose	Summation of total daily bolus taken
Total Daily Basal	Summation of total daily basal
Serial Number	Serial number of device

^a 1 of 2 original datasets

would, connect my research to a wider discourse and tie back into *Insert Study Name Here*'s study especially. Ultimately, the goal is to personally derive value from this study, while at the same time offering the research community a unique set of findings alongside a long and extensive dataset to build upon. All files associated with this project can be found on GitHub.¹

3 Data Source

The data for this project comes from my fiancée's (name omitted for privacy's sake) Dexcom Continuous Glucose Monitor (CGM). A CGM is a device that tracks glucose levels 24/7, providing constant updates for the individual's glucose levels, allowing for better diabetes management (Dexcom 2021). It also has an additional function of saving the data and allowing it to be extracted for further analysis. By connecting the CGM to a computer, individuals are able to download data on their glucose level readings, when they took insulin, when they had carbs etc. For the purposes of this project, I simply took that data, scrubbed it to ensure privacy, and cleaned it for use. The original datasets contained the fields shown in tables 1 and 2, and the cleaned version can be seen in table 3.

Table 2: Original CGM Data

Field Name	Description
Time	Time data was logged
mmol/L	Blood sugar level at the time of data log
...3	No data
Serial Number	Serial number of device

^a 2 of 2 original datasets

4 Diabetes Terminology

For readers unfamiliar with diabetes, as the project revolves around medical data, some clarification of the fields and terminology used is necessary before proceeding.

¹GitHub repository for study: <https://github.com/JeremyJChu/diabetes>

Table 3: Cleaned CGM Data

Field Name	Description
Year	Year data was logged
Month	Month data was logged
Day	Day of the month data was logged (1-31)
Hour	Hour of day data was logged (0-23)
Blood Sugar	Average hourly blood sugar level at the hour of data log
Basal Amount	Total hourly basal taken at the hour of data log
Bolus Volume	Total hourly bolus amount taken at the hour of data log
Carbs	Total hourly carbs taken at the hour of data log
Time of Day	Morning, afternoon, or evening based on the hour
Time of Day Coded	Time of day coded into 1, 2, 3
Range	What the range of blood sugar is for the hour (low, in range, high)
Range Coded	Range coded into 1,2,3
Insulin Food	Whether bolus taken was with food or not (1: No carbs, Yes insulin, 2: Yes carbs, Yes insulin, 3: Yes carbs, No insulin)

^a 1 of 1 cleaned and coded datasets

Type 1 and Type 2 Diabetes

The two types of diabetes. While the risk factors of type 2 diabetes is mainly attributed to lifestyle and age, those behind type 1 are less clear. They are usually associated with genetics and family history rather than more controllable factors. Causes for type 1 and type 2 also differ. Type 1 is an autoimmune disease where the pancreas is unable to make enough insulin whereas type 2 is the body being resistant to insulin (Diabetes Care Community 2021).

A1c

Throughout this paper the term A1c will be mentioned sporadically. A1c, also referred to as hemoglobinA1c or glycated hemoglobin, is simply just a test that measures the 2-3 month average of glucose in blood (Labtests Online 2021). In other words, it is a measure of how managed individual diabetics are. If they are usually in range, their A1c levels will reflect that, if they are uncontrolled, then likewise the A1c will return higher averages.

Blood Sugar/Glucose Levels

Blood sugar or glucose levels is simply the amount of sugar in a person's blood. For the purposes of this project, these values will be represented in mmol/L (millimoles per litre), the UK standard. In the US and continental Europe, the values would instead be represented in mg/dL (milligrams per decilitre). While the calculations differ slightly between the two, they perform the same function in measuring blood glucose concentration (Diabetes UK 2019). Once blood glucose levels fall or rise above a certain range, the individual can be labelled as currently having a low/high blood sugar. Low blood sugars can result in a loss of consciousness with a risk of death whereas high blood sugars for prolonged periods of time can result in a host of health complications.

The general target range usually classifies less than 4 mmol/L as a low blood sugar, and higher than 11 as a high blood sugar. The dataset used in this project will reflect this range.

Continuous Glucose Monitor (CGM) and Insulin Pump

The data comes from the Dexcom G6 Continuous Glucose Monitor System... To be filled in.

Basal

To be filled in

Bolus

To be filled in

5 About the Individual

A brief explanation on the circumstances surrounding the individual used for this case study. This does not impact the research done in this paper, but is rather used to provide further context for additional research down the line.

The individual is a female aged between 20-30, ethnically Indian, and a type 1 diabetic. What differentiates her from typical type 1 diabetics is that she is a type 1 with insulin resistance. While previously uncommon, these cases are becoming more and more prevalent and as such so will the usefulness of this case study. Additionally, the individual was misdiagnosed with type 2 diabetes at age 19 before getting a correct diagnosis as a type 1 diabetic with insulin resistance at age 22.

6 Literature Review

To be filled in

7 Looking at the individual's performance

Before I get any deeper into running models to analyze and predict my fiancée's blood sugar levels and insulin intake, I would like to preface that during the year 2020 (the year the data was collected), she would be what you would call an uncontrolled diabetic.

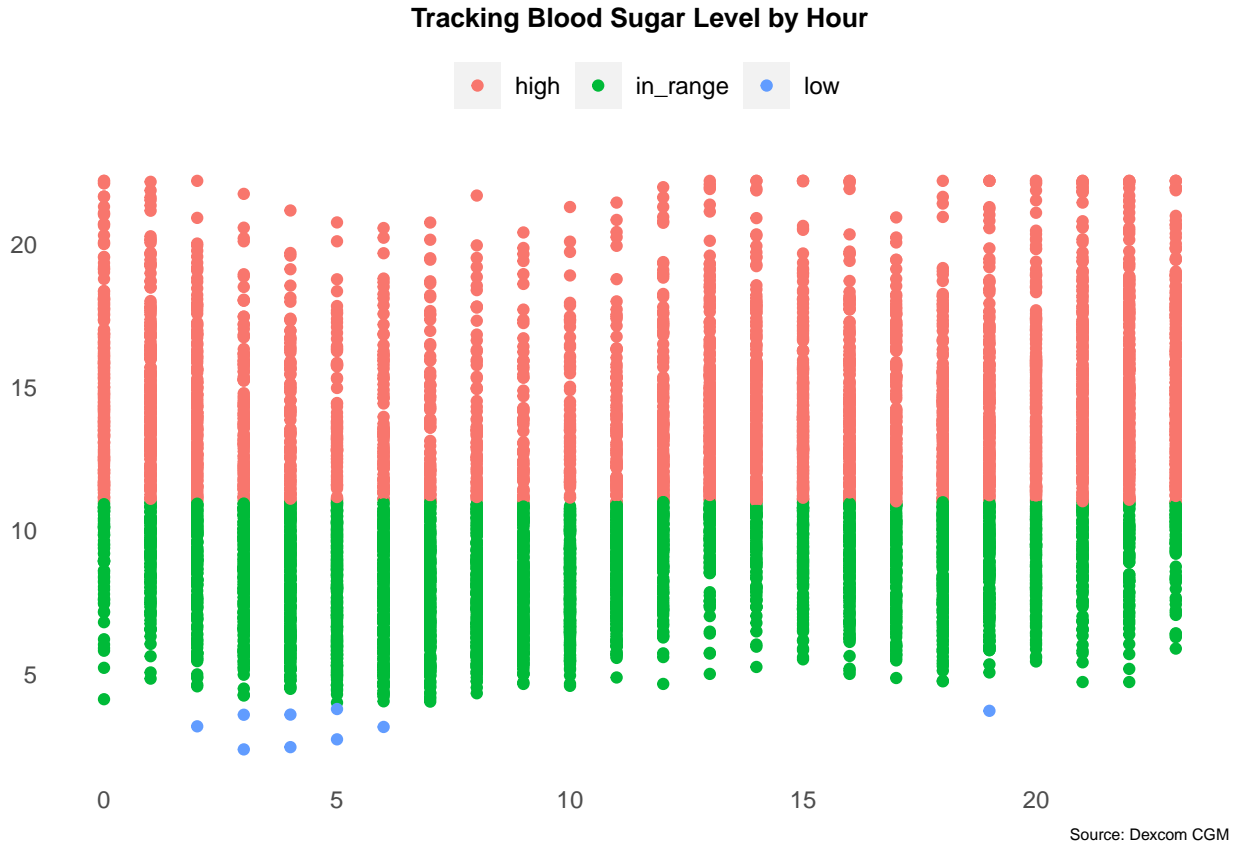


Figure 1: The individual spends significantly more time in a state of high blood sugar

Almost consistently across all hours of the day, for 9 months, my fiancée has been in a state of high blood sugar than in range (See Figure 1). This suggests that the data reflects a situation in which the insulin she takes is insufficient to counteract both her natural blood sugar generation and the amount of carbs she eats. As the data itself is skewed, note that any prediction done will be reflective of insulin intake for an uncontrolled diabetic. It is not representative of diabetics in control of their blood sugar levels and the insulin intake presented in this project will undoubtedly be significantly higher if compared to in control individuals. There will be more datapoints in which insulin is taken without carb consumption (See Figure 2) because of either insufficient insulin taken in the previous meal, or incorrect carb-insulin settings in the pump that failed to account for a need for increased insulin intake.

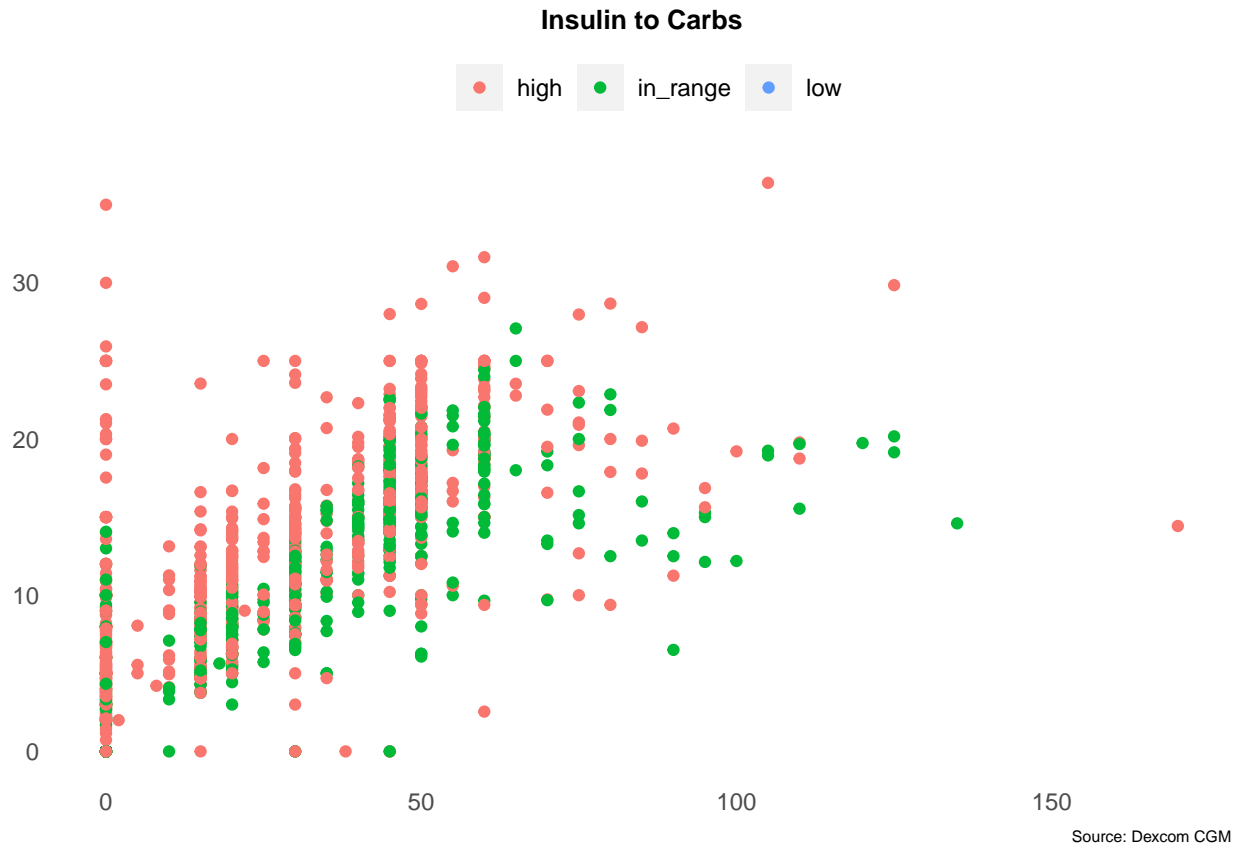


Figure 2: As expected, the individual takes more insulin when eating more carbs, but there is a significant amount of insulin taken when no carbs are eaten

7.1 Eating Habits. Meal Times, Frequencies.

In terms of insulin taking habits (See Figure 3), the individual typically starts the morning off in range. She then noticeably has the least amount of times she takes insulin in the afternoon. When the evening comes around, her blood sugar level is more often high than in range.

What time of day does the subject take insulin, and what is the blood sugar level when doing so

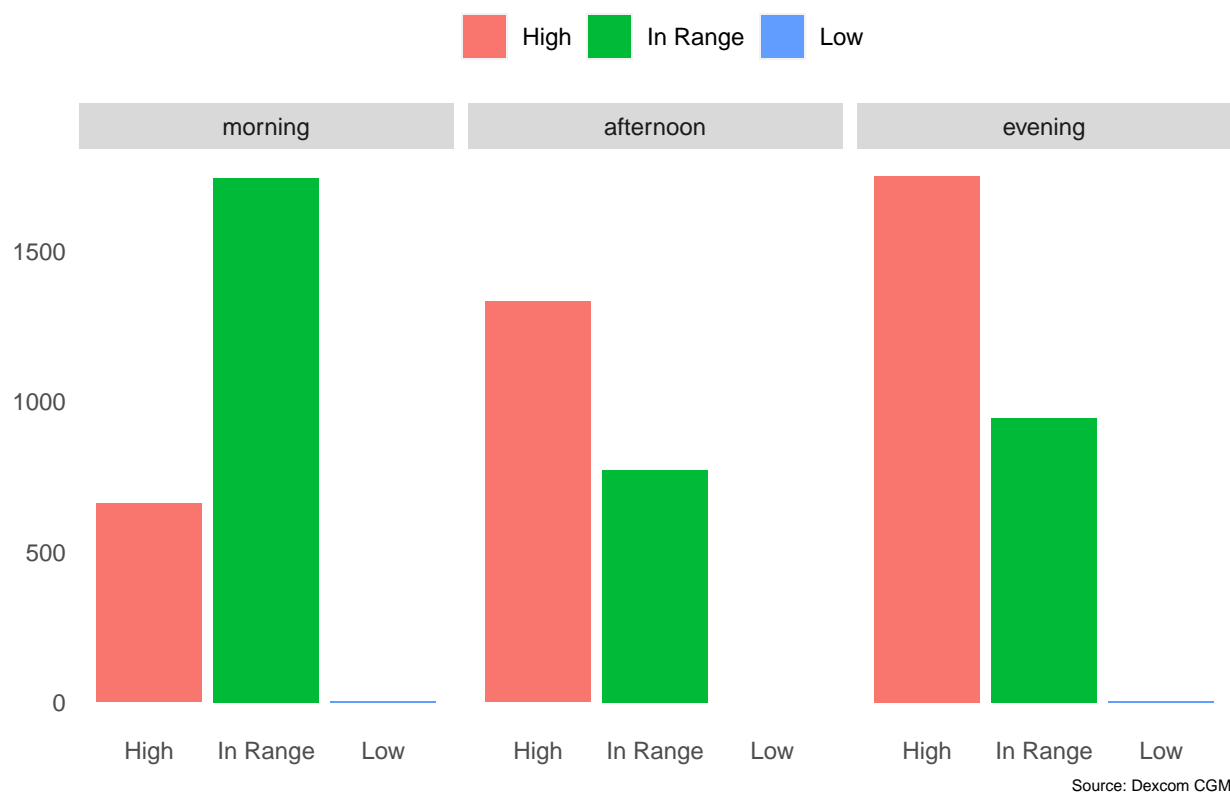


Figure 3: The subject starts the day off in typically in range, then gets more lax before ending up with typically high blood sugar levels in the evening

For meal frequencies, the subject on average eats around 3-4 meals a day (See Figure 4). Meals in this situation are defined as the consumption of carbs, so snacks are counted as 1 meal. A brief glance at WebMD suggests that the Recommended Daily Allowance is 130 grams of carbs per day (cite WebMD). A more in-depth look from the Centers for Disease Control and Prevention (CDC) reveals that ultimately there is no standard, but rather diabetics should aim to get half their calories from carbs (cite CDC). Since I know the subject personally, the doctor recommendation provided was 30 grams of carbs for breakfast, 45 grams of carbs each for lunch and dinner, and 30 grams of carbs for snacks throughout the day. This brings up the subject's daily allowance to up to 150 per day. With the exception of December 2019, the subject has mostly fallen within this range (See Table 4). While the type of carbs consumed elicits varied reactions on blood sugar levels, the project will be progressing with the assumption that the subject is not overeating carbs and therefore leading to high spikes in blood sugar levels, but rather the reasoning for high blood sugar levels stem elsewhere.

Table 4: Carb Intake Breakdown

Year	Month	Total Carbs	Average Carbs/Day
2019	December	5498	183
2020	January	4055	135
2020	May	4460	149
2020	June	3195	106
2020	July	4745	158
2020	August	4640	155
2020	September	4475	149
2020	October	3387	113
2020	November	3013	100
2020	December	3820	127

^a Average carbs rounded to nearest interger

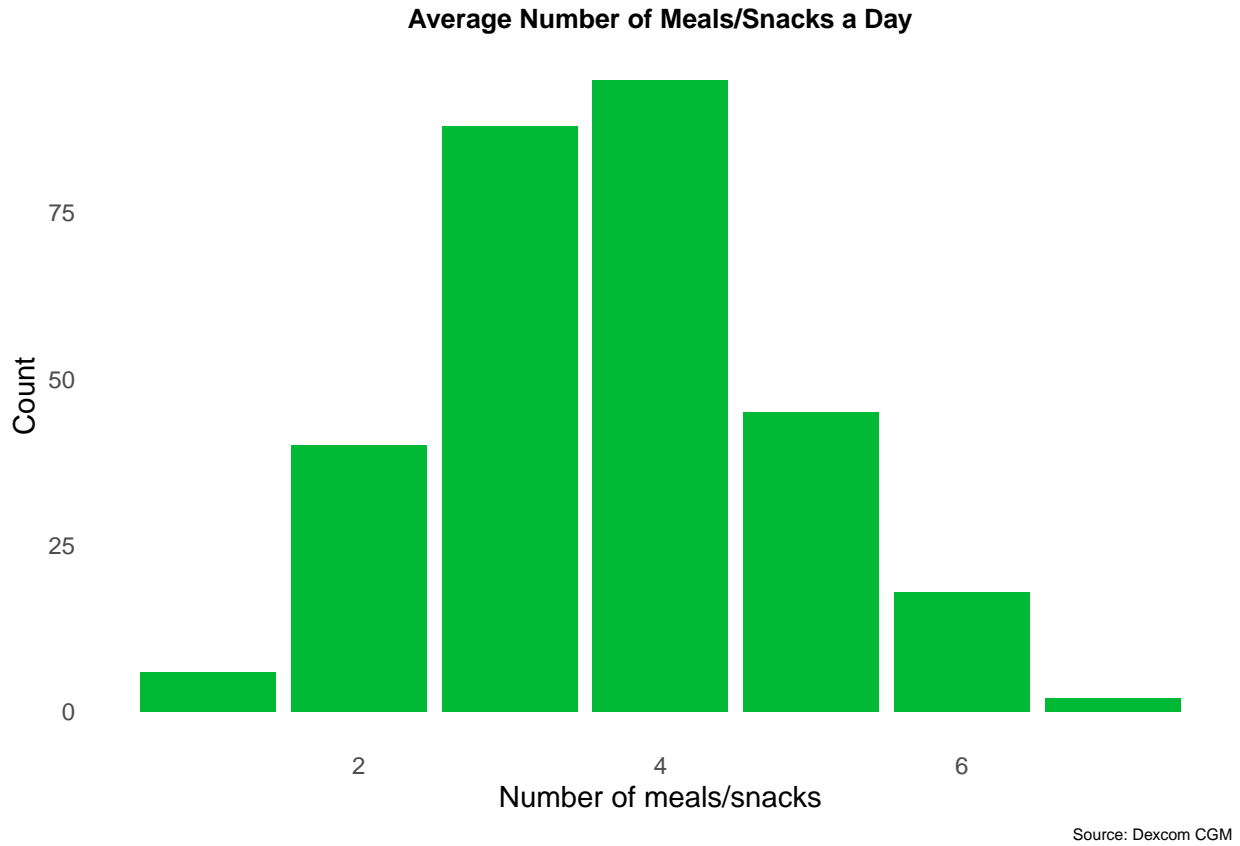


Figure 4: The subject typically eats 3-4 times a day

8 Model

With all the factors outlined above, I decided to run a few statistical models on varying dependent variables. Mainly I wanted to evaluate and predict volume of insulin intake and blood sugar levels. Predicting insulin intake would allow for better adjustment of the tandem pump insulin settings

and carb ratios, whereas predicting blood sugar levels could reveal possible patterns in blood sugar levels in relation to meal frequencies or time of day.

8.1 Predicting bolus volume based on carb intake, time of day, basal amount

To start, I looked at whether carb intake, time of day, and basal amount had a correlation with how much insulin the subject took. Based on previous observations, it would be a natural assumption that insulin intake, that is bolus volume increases with the number of carbs consumed. The question is, would time of day also matter? As we were not shown otherwise during data exploration, my null hypothesis is no, time of day is not a significant factor correlated to how much insulin the subject takes. *Insert relationship between basal and bolus here. Insufficient research right now* Results in table 5.

Table 5

	(1)
(Intercept)	3.506 *** (0.306)
carbs	0.251 *** (0.004)
time_of_day_coded	0.040 (0.112)
basal_amount	-1.002 *** (0.082)
N	2376
R2	0.686
logLik	-6796.572
AIC	13603.144

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$.

The results were as expected. Carb intake shows a positive, significant correlation with insulin intake (See figure 5). Insulin intake should increase by 0.25 units per 1 extra gram of carbs consumed. *Insert blub about basal negative correlation here after research*

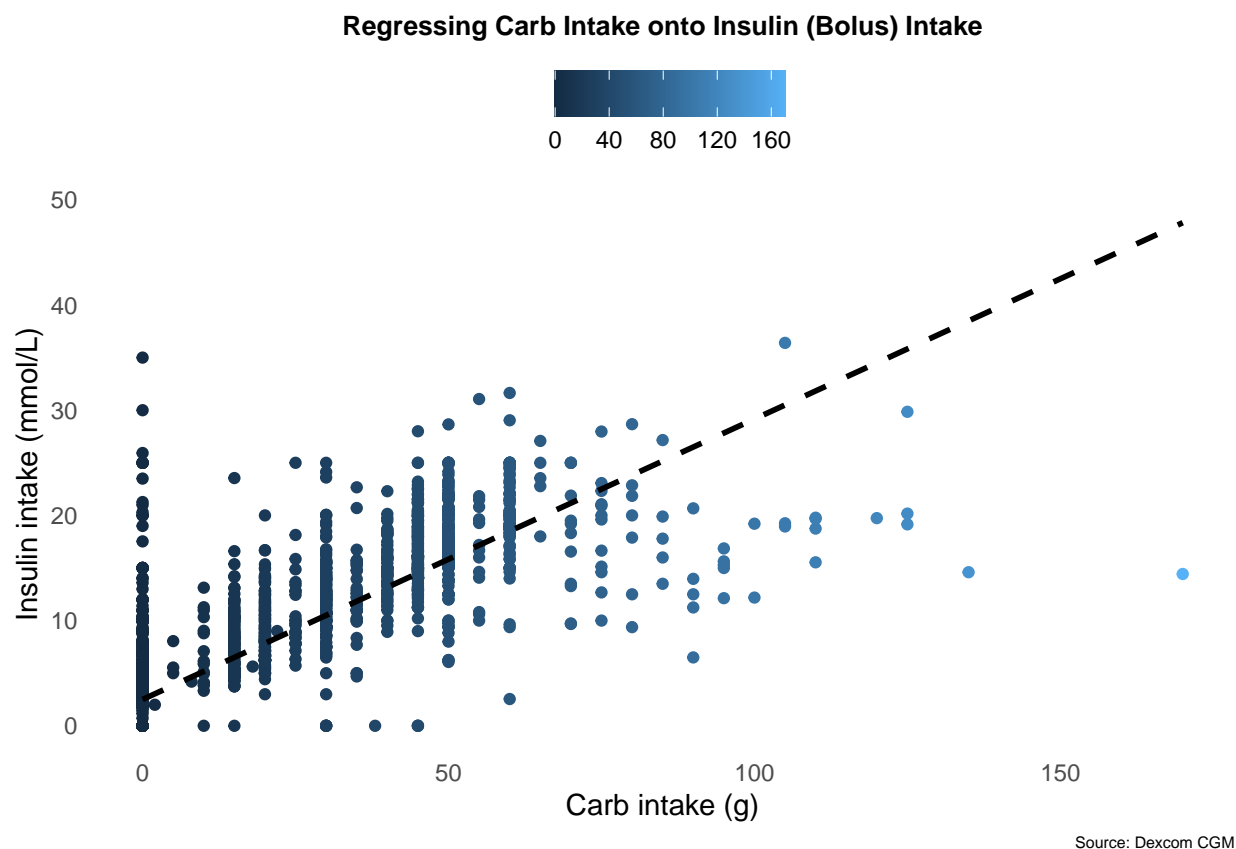


Figure 5: Positive correlation between carb intake and insulin intake

Rather than insulin intake, which is a calculated measure that the insulin pump also calculates, a more valuable question is figuring out factors correlating with blood sugar levels. Table 6 shows the results of a regression predicting blood sugar levels based on carb intake, time of day, bolus amount, and basal amount.

The results show significance across the board (Will be written into paragraph form later).

- Negative correlation with carbs. Every 1g carb eaten, blood sugar drops 0.08 mmol/L. Weird. Why?
- Positive correlation with time of day. Every time the time of day shifts from morning -> afternoon -> evening, blood sugar increases by 1.7 mmol/L. Sounds about right with the data
- Basal. TBD
- Positive correlation with insulin. Blood sugar rises 0.3 for every unit increase in insulin. Why?

Need to break this down further. Do I have too many variables?

Running more models

Still showing weird correlations. (Detailed writeup later)

Table 6

	(1)
(Intercept)	7.768 *** (0.257)
carbs	-0.079 *** (0.005)
time_of_day_coded	1.659 *** (0.091)
basal_amount	-0.678 *** (0.069)
bolus_volume	0.326 *** (0.017)
N	2339
R2	0.325
logLik	-6203.912
AIC	12419.824

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$.

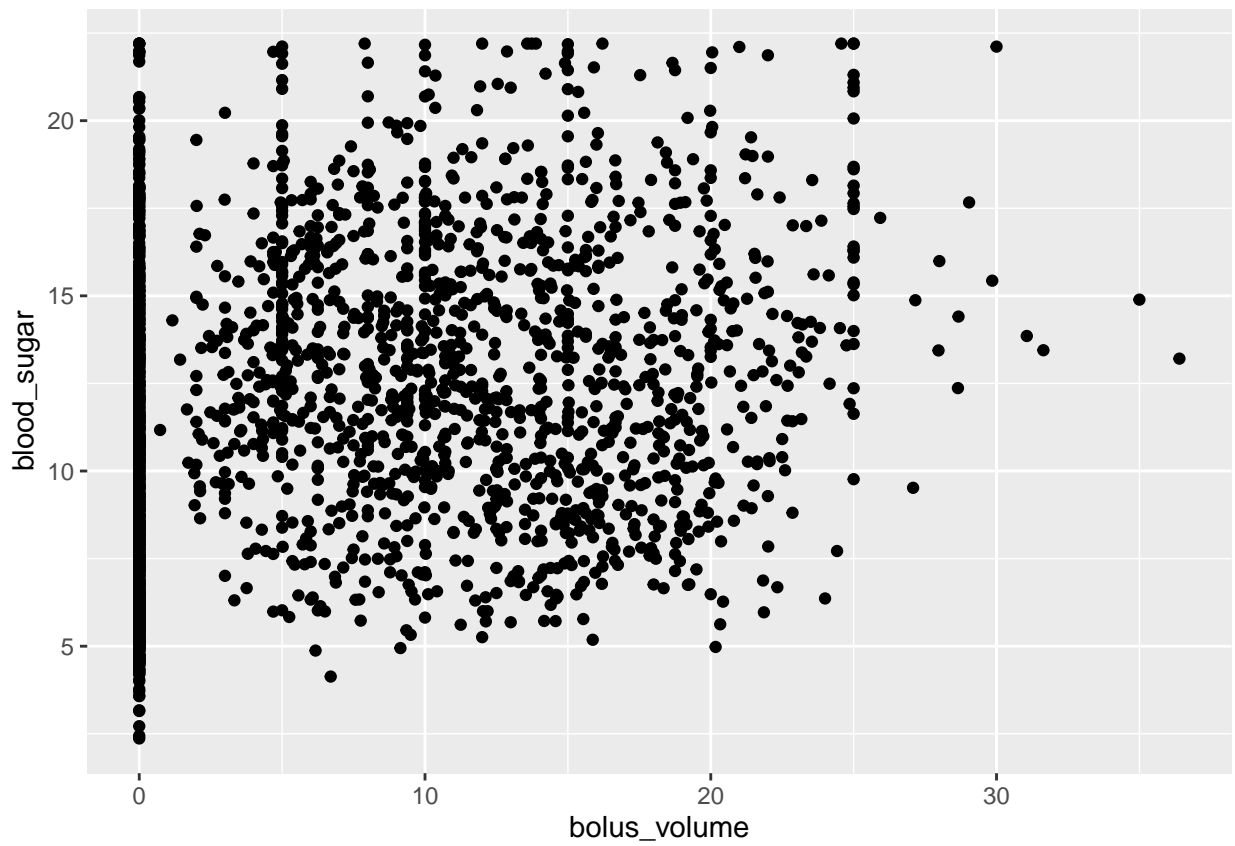
- Hour positive correlation
- Insulin Positive (Weird).
- Carbs Negative (Weird)

Or is it weird. In actuality, the limits of the data actually influence this. As the data is cleaned in a way so that it aggregates by hour, what the data actually says is that for hour 1, the blood sugar is high (15) and so let's say (20) units of insulin is taken. Therefore it makes sense that the results reflect insulin usage rising in relation to blood sugar levels. This means I need to rethink my introduction at the end. Figure 6 shows what a mess it is.

Table 7

	(1)
(Intercept)	9.315 *** (0.169)
hour	0.108 *** (0.011)
bolus_volume	0.372 *** (0.018)
carbs	-0.092 *** (0.006)
N	2339
R2	0.204
logLik	-6396.291
AIC	12802.581

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$.



14
Figure 6: Mess 1

For carbs. I need to figure out why carbs has a negative correlation. It's saying that for each additional gram of carb consumed, blood sugar level drops. Plotting out figure (7) shows a similarly mess of a relationship.

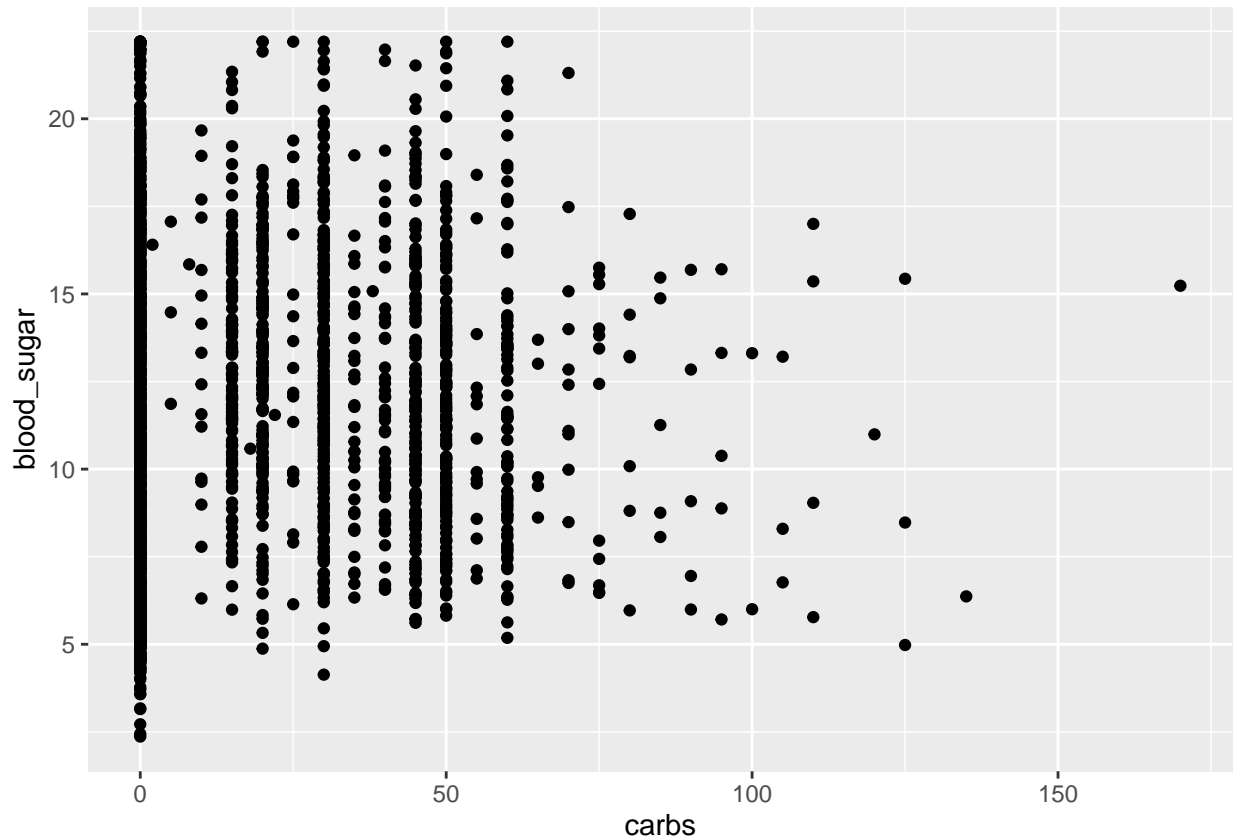


Figure 7: Mess 2

So my plan is to rework the analysis and modelling a bit so to show the limits of my data, and basically focus more on the time aspect of the data. Based on how the datasets are I don't know if it's possible to be able to predict what the blood sugar level will be in like say 4 hours down the line after taking insulin. I could maybe filter out the data for let's say sleep time only, where I know 100% there's no food consumption and then run the regression there. That way I will be able to restructure my analysis to instead show maybe like "Did the subject eat food X hours before sleep" and then "Did the subject wake up with her blood sugar in range".

8.2 Causality

To be filled in. Maybe a DAG? Not sure if I can even speak to causality but will address that if so.

9 Discussion

9.1 Final Takeaways

9.2 Moving Forward

References

Dexcom. 2021. “What Is Continuous Glucose Monitoring (Cgm)?” <https://www.dexcom.com/en-CA/what-cgm>.

Diabetes Care Community. 2021. “Differences Between Type 1 Diabetes and Type 2 Diabetes.” https://www.diabetescarecommunity.ca/diabetes-overview/differences-between-type-1-diabetes-and-type-2-diabetes/?gclid=Cj0KCQjwmIuDBhDXARIsAFITC_7-_kvWgylcWlAApAQ1Tk6L2Ck6QPHR1K0r1wcB.

Diabetes UK. 2019. “Blood Sugar Converter.” <https://www.diabetes.co.uk/blood-sugar-converter.html#:~:text=mmol%2FL%20gives%20the%20molarity,this%20case%20milligrams%20per%20decilitre>.

Labtests Online. 2021. “Hemoglobin A1c.” <https://labtestsonline.org/tests/hemoglobin-a1c>.