Active Care for Diabetes Patients

BGSE Data Science Center

May, 2019

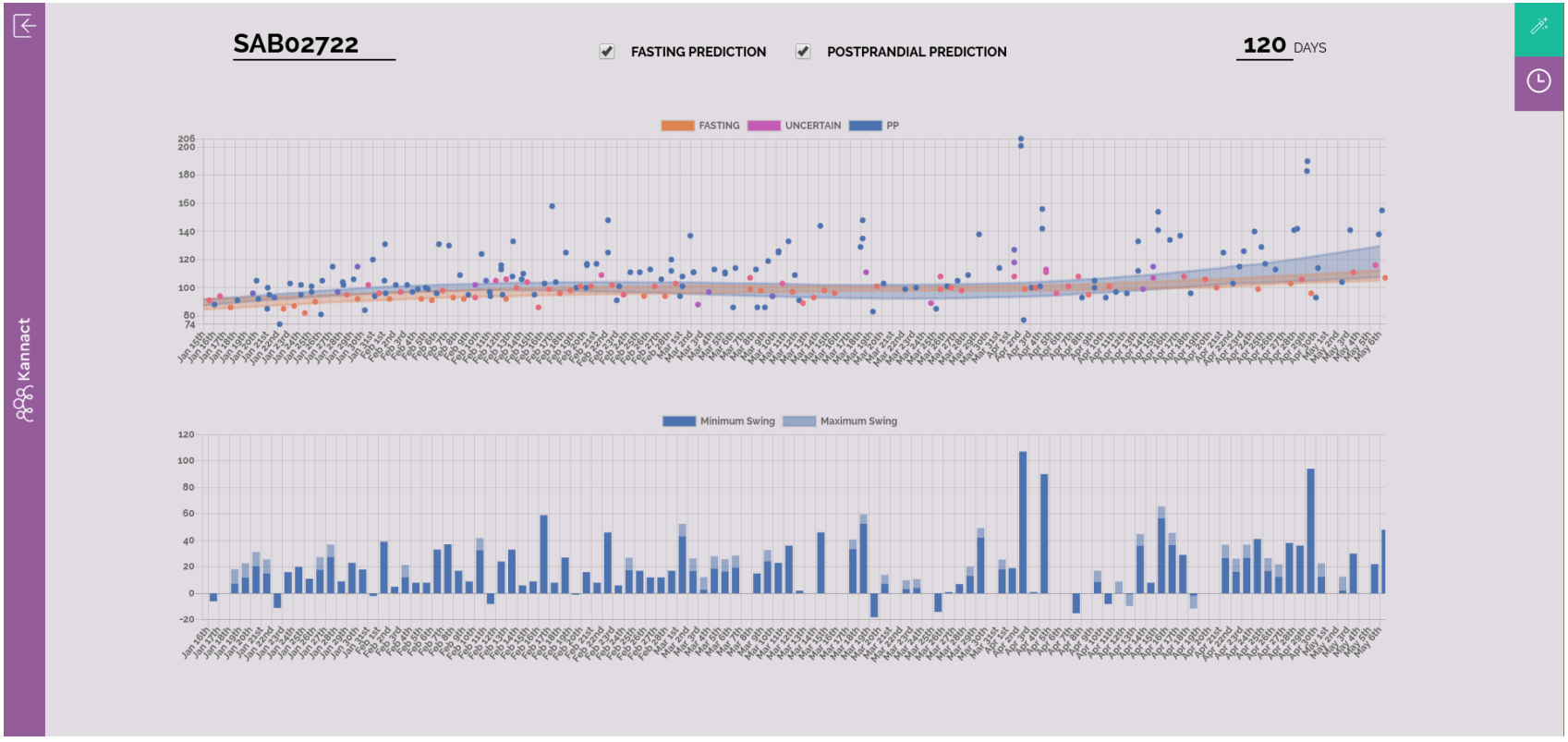
Barcelona GSE Data Science Center

Project Goal

Create a tool for active management of type II diabetes patients in large populations.

- Scaling to a large population requires scalable solutions for gathering data (**patient-generated data**).
- Active management implies recognizing problems faster than the traditional process (**mid-term trends**).

Dashboard Screenshot



(Description in **APPENDIX 1**)

The Data

Type II diabetes is when cells develop insulin resistence and absorb glucose from the blood slowly. It is diagnosed by the following tests:

1. Glucose level in the blood after fasting for 8-10 hours (>125 mg/dl).
2. Glucose level in the blood 2 hours after drinking a special sugar-water concoction (> 199mg/dl).
3. Hemoglobin A1C test (> 6.5%). This measures your average glucose levels for the past 2-3 months, or in other words, a mixture of #1 and #2.

Type II Diabetes

Prediabetes and Early-stage Type II Diabetes patients are primarily prescribed “diet and exercise” to reduce glucose levels to a healthy range.

Mid-stage Type II Diabetes patients take oral medication (such as Metformin) to increase the cells’ ability to absorb glucose, as diet and exercise alone did not prove effective.

Late-stage Type II Diabetes patients inject insulin after meals to help absorb glucose into their cells, as their system no longer responds well enough to oral medications.

Type II Diabetes - Treatment

Treatment is deamed effective if:

1. Fasting glucose levels return to a “healthy” level.
2. Postprandial glucose levels return to a “healthy” level.
3. Hemoglobin A1C levels return to a “healthy” level.

(Or, if they don’t get better, they at least stop getting worse)

Type II Diabetes - Treatment

Doctors check the glucose levels and run A1C tests every 3-6 months, to see if the treatment is working for the patient.

Type II Diabetes - Treatment

3-6 months is a long time to wait just to see if your new, painful diet and exercise routine are having any impact on your disease.

And if they haven't, you don't have many 3-6 month iterations before you get put on medication (or a higher dose of medication or insulin).

Type II Diabetes - Data

Blood glucose levels at any point in time are caused by two factors:

1. The amount of glucose dumped into the blood when the patient eats.
2. The amount of glucose absorbed from the blood by the cells.

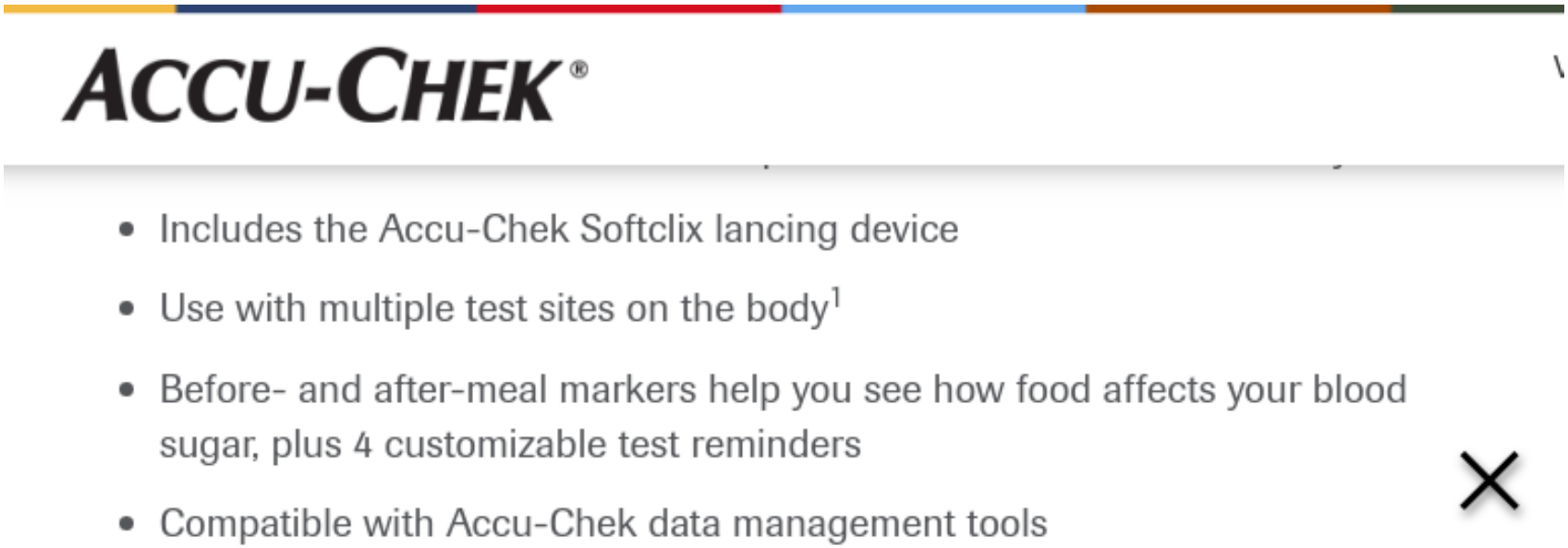
In the tests to determine how the disease is developing, #1 is minimized or controlled. It's noise.

#1, however, contains information about the behavior of the patient and their diet.

Type II Diabetes - Data

It’s important to understand that take-home glucometers have historically been seen as useful for #1, determining the effect of food.

Marketing Screenshot (Glucometer)



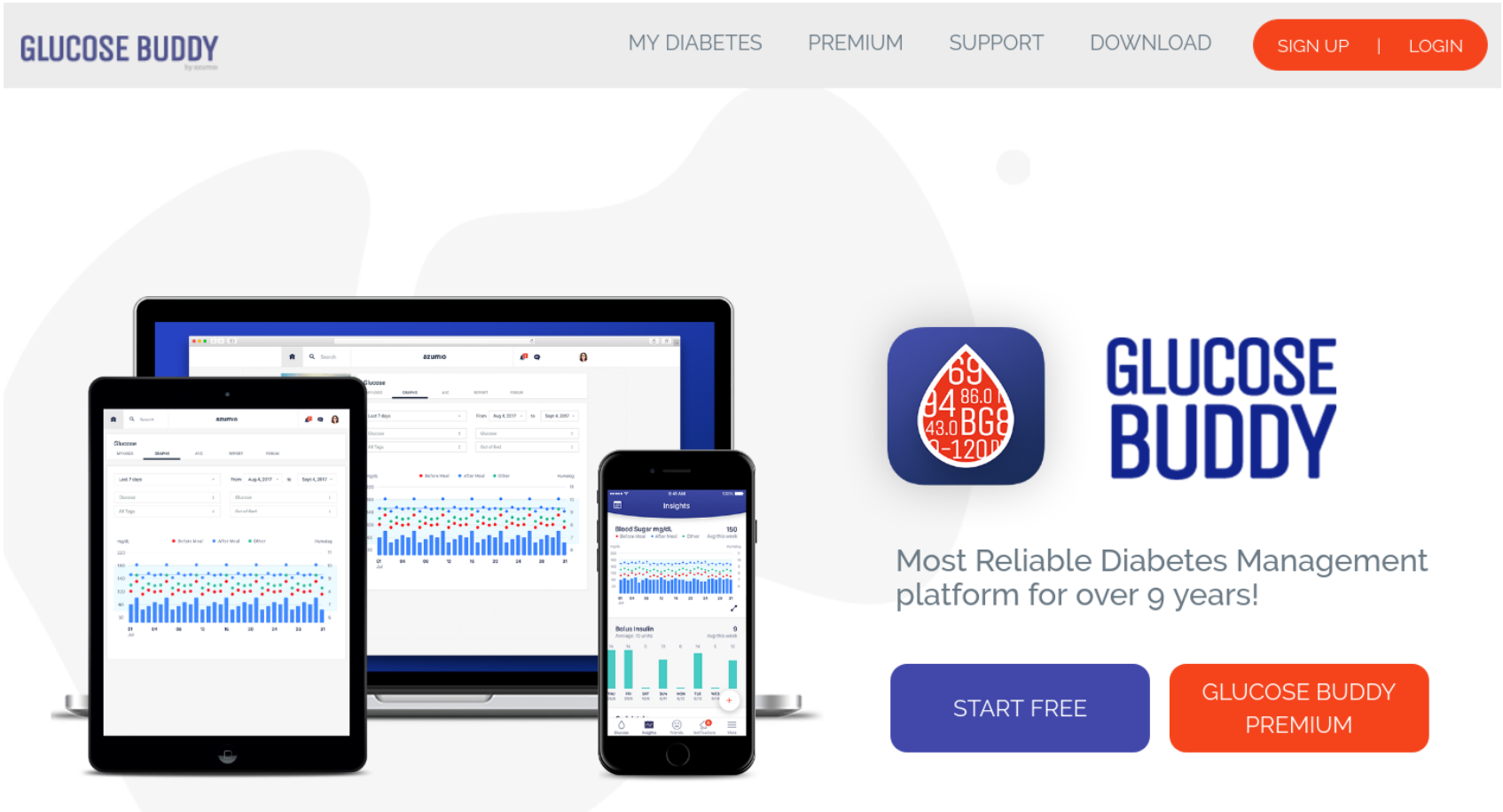
Marketing - Glucose Management

There is a bonanza of new apps to help people manage their diabetes.

Similarly, they focus on short-term changes in glucose, the effects of individual food items on glucose changes, and, at best, averages, without truly trying to measure medium-term trends in the levels that matter.

These apps do not seem to differentiate between different types of type II patients and their differing needs.

Marketing Screenshot (App)

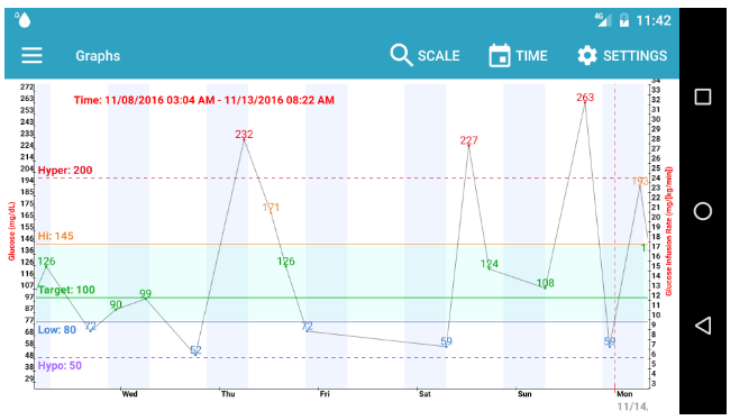


Marketing Screenshot (App)

Diabetes : M

HOME FEATURES BLOG CONTACTS USER'S GUIDE DIABETES:M LOGIN

Detailed Graph



(placeholder screen)

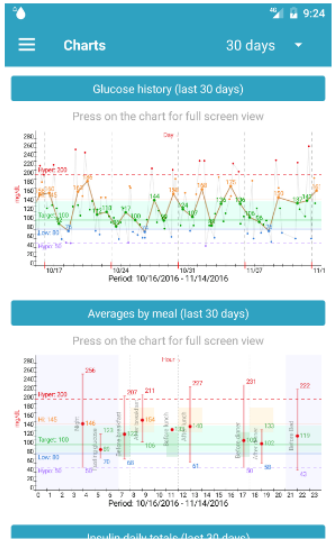
Diabetes:M can give you all the blood sugar entries as a timeline graph.

Analytical Charts

Diabetes:M gives you a set of charts and diagrams to present the data in various ways. These are displays that are organized to give you the big picture of the data at a glance.

The available charts list:

- Distribution by days
- Distribution by category
- Daily glucose overlay
- Glucose history
- Metabolic control
- Averages by meal
- Insulin daily totals

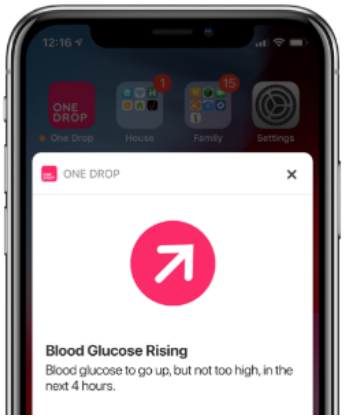


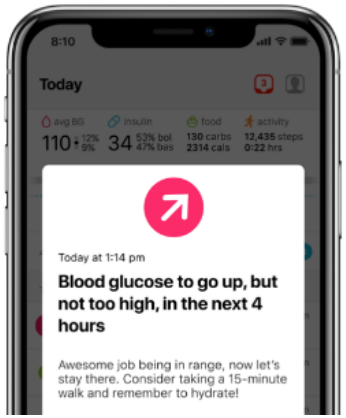
Marketing Screenshot (App)

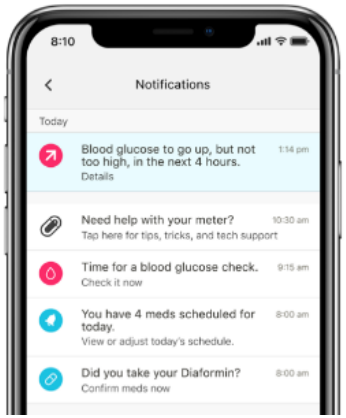
SUPPORT BLOG ONE DROP

How it Works

A user enters a single blood glucose moment into One Drop | Mobile. **One Drop | Automated Decision Support** analyzes this moment, referencing 1.3 billion existing food, activity, medication, and glucose data points from over 1 million users . The user automatically receives an in-app notification including a relatively precise future blood glucose report and suggestions to counteract a possible blood glucose spike or drop.







Medium-term Trends

Daily, or inter-daily, measurements can only reveal a very short-term trend.

At a daily level, the signal that matters for understanding the evolution of early- and mid-stage type II diabetes (the change in the cells' resistence to insulin) is lost.

3-6 months is a long-term trend.

At 3-6 months, the effectiveness of a regime becomes clear, but it might be too late to change or adjust.

Nowcasting

This is why we have proposed to nowcast the trends observed at a 3-6 month level, as they evolve on a daily/weekly basis.

Even if after one has the measurements of interest, this would require the reduction of the measurements to a sufficient statistic and a model that can capture the evolution of that statistic over time.

Barcelona GSE Data Science Center

Our Data

Data and inspiration for this problem comes from Kannact.

Kannact is a healthcare technology startup based in the U.S. focused on managing populations of patients with diabetes and cardiovascular disease.

Glucose readings are collected by 4G-connected glucometers that are sent to all their patients.

These glucometers have buttons for to label readings as “pre-meal” and “post-meal”.

Those user-defined labels, along with the timestamp, are the only metadata provided with the readings.

Our Data

Some patients label all of their readings as “pre-meal” or “post-meal”.

Most patients forget some of the time.

Many patients forget all of the time.

Process

In order to recover the labels *of interest* (fasting/postprandial) we need to consider two dimensions:

1. The reading value
2. The time of day in which the reading took place.

Where the time of day can help us infer if the reading is taken while fasting (immediately after waking up).

Where possible, we can also take advantage of:

3. The user-provided label (pre-meal/post-meal)

Model

Given the noisy data, we treat this problem as one of probabalistic inference. As such, we would like our model to learn all the following elements simultaneously via the maximum likelihood principle:

1. The label (fasting/postprandial) of each glucose reading.
2. A parameterized distribution for each label type (fasting/postprandial).
3. The smooth, potentially non-linear, evolution of (one of) the parameters from the distributions over time.

Model

For each patient, we learn the following model:

Let y_{ij} be the i^{th} glucose reading on date j and time $t_{ij} \in [0, 24]$.

Let $z_{ij} \in 1, 2$ be the latent fasting/postprandial label. We formulated the following mixture model in two dimensions (glucose and time):

$$(y_{ij}, t_{ij} | z_{ij} = 1) \sim Laplace(\mu_j^1, \sigma^1),\ Beta(\alpha_i, \beta_i)$$

$$(y_{ij}, t_{ij} | z_{ij} = 2) \sim Laplace(\mu_j^2, \sigma^2, \delta^2),\ Unif(0, 24)$$

Where we assume $y_{ij} \perp t_{ij} \mid z_{ij}$.

Model

In order to allow changes over time, we incorporate a change to the location parameters μ given the date.

This is done via a regression on a non-linear basis transformation of the date value:

$$\mu_j^k = \mathbf{w}\Phi(j)$$

Where **w** is a vector of learned coefficients, ***j*** is the date, and **Φ** represents the feature transformation function, in this case, a Fourier basis.

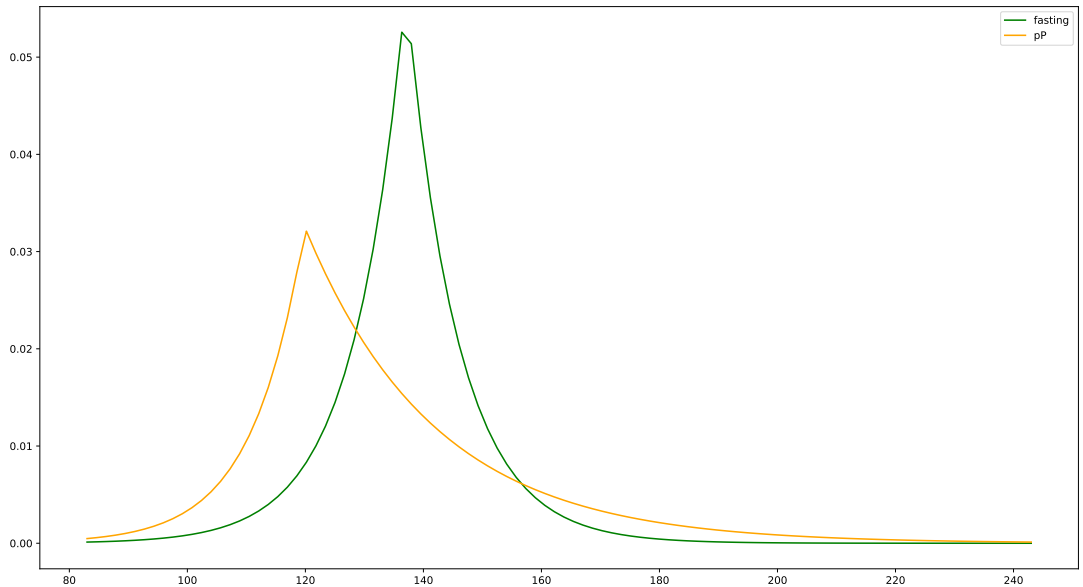
Model

The number of Fourier bases used to represent the time evolution of any given patient is done by comparing AIC scores between the full models.

The evolution of the location parameters is reported with 95% confidence intervals created via bootstrap resampling of the evolution regression given the best basis function.

Model

Fitted distributions, on a single day, for a mid-stage patient on oral medications:



Sufficient Statistics

It is important to emphasize the importance of having sufficient statistics (and thus reasonably parameterized distributions) if we want to track changes in an underlying “level” of a process that produces recordable data with some noise.

Success

Does our model accurately summarize the blood glucose and it's evolution by fitting these two distributions?

Predictive Ability

To validate our modelling of fasting and postprandial scores, we use a recorded A1C test score which we have for a small subset (~44) of our patients.

A1C can be thought of as a measure of the average level of glucose in a body for 2-3 months.

Predictive Ability - Correlation

Medical studies [1] comparing A1C scores with with average fasting and postprandial readings can give us a test to see whether the sufficient statistics of our parameterized distributions have captured the glucose levels sufficiently.

Results in these well-controlled studies suggest that the correlations between readings and A1c scores should be:

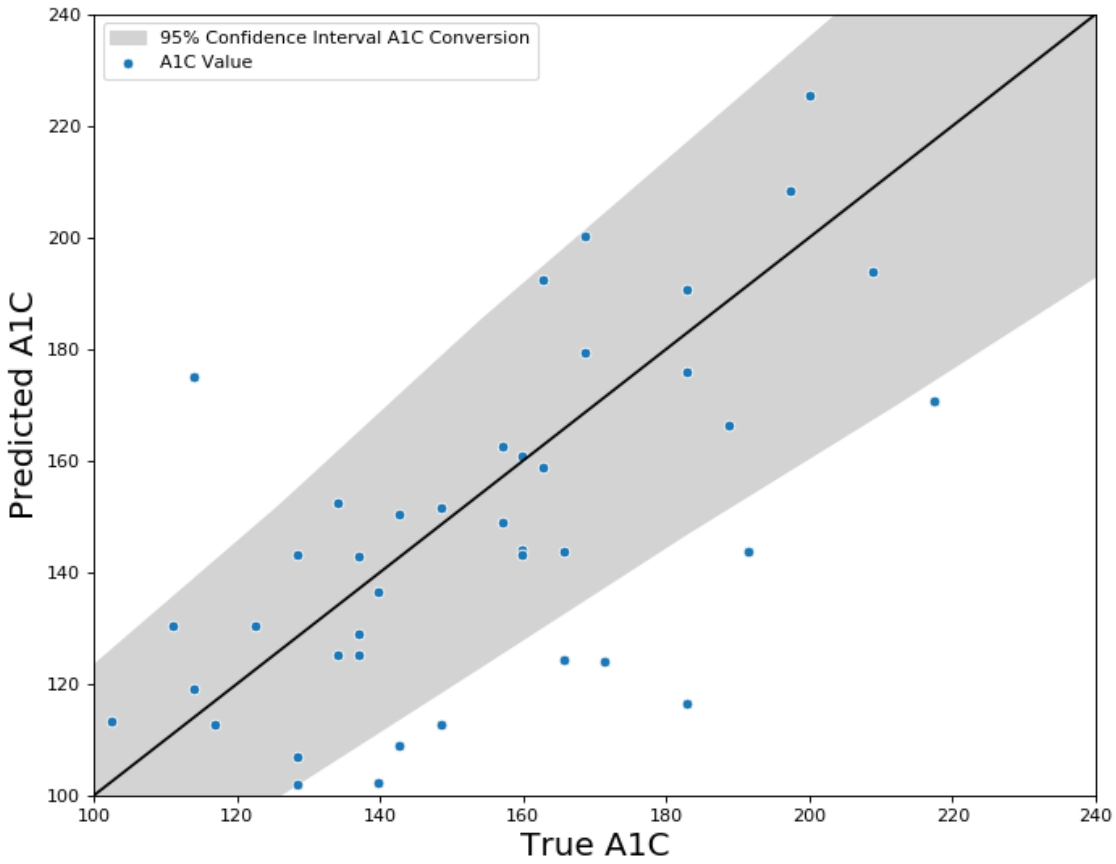
1. For fasting levels, between .46 and .71, the upper end being for populations known to have diabetes. We recovered .66 with our model and patients.
2. For postprandial levels, between .33 and .79. We recovered .54 with our model and patients.

[1] Van't Riet et al 2010, Monnier et al. 2006

Predictive Ability - Regression

Similarly, a simple linear combination of the two means calculated from the parameters of our fitted distributions recovers a good approximation of the A1C scores up to the confidence intervals provided with the A1C test and conversion into mg/dL itself.

Predictive Ability - Regression



Conclusions

Fasting, and postprandial levels represent a combination of information about user behavior and information about the biological transformations happening in the system of a diabetes patient.

Recovering the mid-term trends in the latter are needed to allow patients to see feedback from their exercise and diet regimes.

A focus on recovering mid-term trends is sorely lacking in both medical literature and commercial offerings.

A probabilistic model is necessary to separate signal from noise and provide it in a usable way for health coaches.

Appendix 1 - Project Results

The dashboard enables:

- Active-management professionals (health coaches and doctors) to monitor the evolution of key variables for early- and mid-stage type II diabetes patients.
- Coaches and patients to see the effects of short-term diet and exercise behaviors on mid- and long-term evolution of the biological process underlying the disease.
- Coaches and patients to see the quality, regularity, and reliability of the blood-glucose readings taken by the patient and the effects of improving.

Appendix 1 - Project Results

The dashboard displays data on blood glucose levels.

- Blood glucose levels are the primary and most important data for monitoring diabetes.
- They are recorded from self-administered blood tests by patients via a glucometer.
- Blood glucose levels are a noisy signal of multiple biological processes happening simultaneously.

Appendix 1 - Project Results

From the noisy signal of raw glucose levels, the dashboard presents the inferred variables that align with the medical theory on managing type II diabetes:

1. Fasting glucose levels and their evolution over time.
2. Postprandial glucose levels and their evolution over time.
3. Daily “swings” in glucose levels that causes cardiovascular stress to the system.

Appendix 1 - Project Results

Technical results needed to build the dashboard and present these variables in a usable manner:

- Infer the “type” of blood-glucose reading from the partially-labelled patient data provided by the glucometer (fasting/postprandial).
- Infer a “location” parameter for each type of reading that is robust to the outliers and skewness present in patient-recorded glucose data.
- Model the evolution of that location parameter for each patient over time, including prediction and interpolation when missing readings.