USE OF MACHINE LEARNING TO PREDICT INDIVIDUAL POSTPRANDIAL GLYCEMIC RESPONSES TO FOOD AMONG INDIVIDUALS WITH TYPE 2 DIABETES IN INDIA

Niteesh K. Choudhry, MD, PhD\textsuperscript{1,2}; Shweta Priyadarshini, PhD\textsuperscript{1}; Jaganath Swamy, MBA\textsuperscript{1}; Mridul Mehta, PhD\textsuperscript{1}

Author affiliations:

1. Decipher Health, Inc, Boston, MA;
2. Harvard Medical School, Boston, MA.

Corresponding Author: Niteesh K. Choudhry, MD, PhD; Email: nkc@post.harvard.edu

Keywords: Diabetes, India, Nutrition, Algorithms, Blood Sugar, Postprandial

Trial registration number: CTRI/2022/02/040619

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.
ABSTRACT

Background: Type 2 diabetes (DM2) is a leading cause of premature morbidity and mortality globally and affects more than 100 million people in the world’s most populous country, India. Nutrition is a critical and evidence-based component of effective blood sugar control and most dietary advice emphasizes carbohydrate and calorie reduction. Emerging global evidence demonstrates marked inter-individual differences in post-prandial glucose response (PPGR) although no such data exists in India and prior studies have primarily evaluated PPGR variation in individuals without diabetes.

Methods: This prospective cohort study seeks to characterize the PPGR variability in Indians with diabetes and to identify factors associated with these differences. Adults with type 2 diabetes and a hemoglobin A1c ≥7 are being enrolled from 14 sites around India. Subjects wear a continuous glucose monitor, eat a series of standardized meals, and record all free-leaving foods, activities, and medication use for a 14-day period. The study’s primary outcome is PPGR, calculated as the incremental area under the curve 2 hours after each logged meal.

Discussion: This study will provide the first large scale examination variability in blood sugar responses to food in India and will be among the first to estimate PPGR variability for individuals with DM2 in any jurisdiction. Results from our study will generate data to facilitate the creation of machine learning models to predict individual PPGR responses and to facilitate the prescription of personalized diets for individuals with DM2.
Type 2 diabetes mellitus (DM2) is the leading cause of chronic kidney disease, end-stage renal disease, blindness and non-traumatic amputation; it also substantially increases the risk of myocardial infarction, stroke and heart failure. Its prevalence is particularly high in India, which is now the most populous country in the world. As of 2023, more than 100 million Indians have diabetes, representing more than 11% of the population. An additional 136 million Indians have prediabetes. These numbers are anticipated to continue to grow rapidly. The lifetime risk of DM2 among obese 20-year-old Indians is estimated to be more than 86%. The rising prevalence of this condition in India is believed to be the result of changing diets, increasingly sedentary occupations, lower levels of physical activity in the context of urbanization, and rapidly increasing rates of obesity.

These trends are particularly concerning because of important differences between the presentation and consequences in DM2 in Indians compared with other racial and ethnic groups. These differences, which often referred to as the “Asian Indian Phenotype”, are characterized by the onset of DM2 at a younger age and substantially lower body-mass index (BMI) than people of other races and ethnicities. Indians have higher levels of insulin resistance (and for longer periods of time) and premature beta-cell failure. They are more likely to develop the fatal complications of DM2, most notably heart disease. These features are thought to result from a mix of lifestyle, epigenetics, and fetal programming factors.

The fundamental goal of diabetes management is to maintain near-normal glucose levels. A variety of self-management behaviors, in particular adherence to diet and regular exercise, are central to this goal. An extensive body of evidence demonstrates that aiding DM2 patients with self-management behaviors is associated with improvements in a wide range of outcomes including knowledge, self-care behaviors, weight, quality of life, hemoglobin A1c (A1c), all-cause mortality and health care costs. Guidelines recommend that nutritional guidance be personalized based on nutritional status, lifestyle, and metabolic goals. Despite this, most dietary advice for individuals with DM2
remains generic emphasizing reductions in calories and minimization of carbohydrates.\textsuperscript{16} However, there are marked inter-individual responses in post-prandial glucose response (PPGR).\textsuperscript{17} A study conducted in Israel found substantial PPGR variability to standardized meals for individuals without diabetes.\textsuperscript{18} Similar data has been generated in the UK, US and China.\textsuperscript{19-21} There have been no studies characterizing food responsiveness among Indians and virtually no published data, from any judgment, in the variability in PPGR for individuals with DM2.\textsuperscript{22} Given the unique Indian diabetes phenotype and differences between Indian and western diets, there are very likely to be differences in blood sugar responses to food and exercise in India than observed elsewhere, just as there have been in Indians’ responses to diabetes medications.\textsuperscript{25} Accordingly, the goal of this study was to characterize and identify factors associated the variability in PPGR among individuals with DM2 in India.

METHODS

This prospective cohort study seeks to evaluate the relationship between PPGR and self-management activities including diet, exercise, and other daily routines, for individuals with DM2 in India. This study was approved by the ethics committees at all institutions enrolling patients and is registered with Clinical Trials Registry-India (CTRI/2022/02/040619). The authors are responsible for the design, conduct and analysis of this study and all met International Committee of Medical Journal Editors (ICJME) criteria.

Study setting

This trial is being conducted at 14 outpatient clinics in geographically distinct regions across India. Sites were identified and managed by a multinational contract research organization and were included if they specialized in the care of individuals with diabetes, had
an established research infrastructure for the conduct of diabetes-related studies including a Site Principal Investigator who is a diabetologist (with clinical training in Endocrinology or Internal Medicine) and a local ethics committee to provide study oversight, and a sufficient volume of potentially-eligible patients. Study enrollment began in May 2022.

**Eligible subjects and enrollment**

The study population consists of adults with diabetes and suboptimal disease control, classified A1c ≥7%. Complete inclusion and exclusion criteria are summarized in Table 1.

Potentially eligible patients are identified from clinic records and are invited to attend an in-person screening visit at which time eligibility is confirmed and written informed consent obtained. Consenting patients are asked to provide sociodemographic and medical information (specifically, age, sex, predominant diet, health conditions, family history, and current medications) and to complete baseline surveys including the World Health Organisation STEPwise Approach to NCD Risk Factor Surveillance (STEPS) survey, World Health Organisation-Five Well-Being Index (WHO-5), Diabetes Distress Scale, Wilson Adherence Scale and the Pittsburgh Sleep Quality Index.

Biometric data including blood pressure, heart rate, weight, height, and body measurements at the upper arm, thigh, calf, waist, and hips, are collected by study coordinators at each site. Finally, enrolled subjects provide blood samples including a complete blood count, hemoglobin A1c, blood electrolytes, creatinine, and cholesterol, as well as a urinalysis.

After completing baseline assessments, subjects are fitted with an Abbott Freestyle Libre continuous glucose monitors (CGM) sensor on their upper, non-dominant arm and are provided with a Xiaomi Mi Band Smart Wristband (heart rate monitor) and a Roche Accu-Chek glucometer with testing supplies, and dietary supplements to be consumed with their standardized meals (see **Follow-up procedures** below). Study-specific applications are
downloaded on to the subjects’ smartphones to allow them to log dietary intake and synchronize their continuous glucose and heart rate monitors. As a back-up, subjects are given a paper dietary logbook and a Freestyle Libre CGM reader with which to collect protocol-specified data.

Follow-up procedures

Subjects are followed for 14 days. They are instructed to wear the CGM and heart rate monitor. The heart rate monitor is to be always worn, including during sleep, and only removed for re-charging. Subjects are also asked to check their capillary glucose on days 2 through 6 before breakfast and dinner.

Subjects log their full dietary intake using the study app or logbook over the 14-day study period, including all standardized test meals and free-living foods (including snacks), beverages (including water) and medications. Participants also log all exercise.

Subjects are required to consume protocol-specified meals and to perform light activity, as described in Table 2. The standardized meals consist of vegetarian breakfast foods which subjects are to prepare in their homes. The meals vary in their proportion of carbohydrate, fiber, protein, and fat (see Table 3). Subjects are instructed to: (a) fast for a minimum of 8 hours prior to and 3 hours after consuming the standardized breakfast meal; (b) during these fasting periods, limit exercise and drink only still (not sparkling) water, tea or coffee in moderation; (c) eat the meal, in its entirety, within 20 minutes. After completing the post-meal fasting period on standardized test meal days, subjects may consume other foods as they normally would unless there are other meal modifications specified by the protocol later the same day.

On other days, subjects are asked to consume normal foods with protocol-specified constraints. For example, on different days, subjects vary the types of mixed protein (e.g., different types of lentils with or without added protein), ordering with which foods are consumed (e.g., protein before carbohydrate v. protein with carbohydrate), consume water before their meal, go for a walk after eating or eat what they perceive to be a healthy meal. Where
applicable, subjects are given several options as to which of their usual foods are acceptable for each protocol specified food modification.

During the follow-up period subjects are contacted by phone and text messages to ensure protocol compliance. If the outreach identifies a problem with a CGM (i.e., it was damaged, fell off or malfunctioned), study staff provide a replacement within 24 hours during which time subjects are asked to pause their meal protocol and restart once their CGM has been re-applied and recalibrated. If a test meal was not consumed as intended, participants are provided with the option to repeat the meal.

On Day 15, subjects are asked to remove their CGM and will complete endline surveys. If need be, study staff help with recording into the study app or logbook of missing or inaccurate food, activity, and medication data.

**Statistical analysis plan**

The study’s primary outcome is PPGR. Following the Wolever and Jenkins method,28 as adapted by Zeevi et al,18 Mendes-Soares et al20 and Berry et al,29 logged meal times and continuous glucose measurements will be used to calculate the incremental area under the curve (iAUC). Prior to conducting analyses, meals logged less than 30 minutes apart will be merged and meals logged within 90 min of other meals will be removed. Very small (<15 g and <70 calories) meals and meals with very large (>1 kg) components, meals with incomplete logging and meals consumed at the first and last 12 hour of the CGM connection will also be removed. To reduce noise, the median of all glucose values from the 30-minute period prior to the meal will be taken as the initial glucose level, above which the incremental area will be calculated. Meals that had incomplete glucose measurements in the time window of 30 minutes before and 2 hours after the logged mealtime will be filtered out.
Descriptive statistics will be used to plot the range of PPGRs responses to standardized test meals as well as the correlation between PPGR and the nutritional composition of the logged meals (i.e., carbohydrates, fat, and protein).

A machine learning predictor will be developed based on stochastic gradient boosting regression (XGBoost, version 0.6)\textsuperscript{30} using the XGBRegressor class. Postprandial glycemic responses will be predicted as the sum of predictions from thousands of decision trees. Trees will be inferred sequentially, with each trained on the residual of all previous trees and making a small contribution to the overall prediction. The features incorporated in each tree are selected by an inference procedure from a pool of features representing meal content, demographics, health habits, baseline laboratory values, as well as CGM, heart rate and activity data.

Performance will be assessed by 5-fold cross-validation in which participants are divided into 10 groups, the model will be trained on 9 parts, and the performance will be measured by the ability to accurately predict meals reported by the left-out participants (out-of-bag predictions). Prediction results on subjects from all left folds will be aggregated, and Pearson product moment correlation with the measured post-prandial glucose responses will be reported. The standard error for the calculated performance will be assessed using 50 iterations of bootstrapping. Random data sets of the same size as the original will be sampled with replacement from the original data set, and the entire training and validation process was repeated.

**Sample size considerations**

1050 individuals will be targeted for recruitment. Assuming a 5% loss-to-follow-up, this corresponds to 1000 evaluable individuals at the end of the study. The study has been designed to predict postprandial glucose responses based on individual characteristics and 1000 subjects followed for 14-days will result more than 4 million glucose readings (assuming glucose
readings from the CGMs every 5 minutes) and 42,000 meals (assuming 3 meals per participant per day). This volume of data will also provide more than 80% power to detect correlations of a magnitude of $r=0.13$ ($R^2=0.017$) with an $p<0.005$. We will also be sufficiently powered to detect effects of $r=0.165$ ($R^2=0.027$, i.e., explaining 2.7% of interindividual variation) with $p<0.00001$, i.e., accounting for 5000 independent hypothesis tests.

**DISCUSSION**

The prevalence of DM2 has grown increased rapidly in India such that 1 in 6 people in the word with this condition live in this country.\textsuperscript{31} Along with medications and physical activity, diet is a key tenant of effective blood sugar control.\textsuperscript{15} Guidelines call for individualization of meal planning, which is sometimes referred to as “Medical Nutritional Therapy”. Despite this, personalization of dietary plans are generally based upon broad constructs like age, activity level, health status and preferences, and, for all patients, tend to emphasize overall calorie reductions and minimization of carbohydrate,\textsuperscript{16} especially added sugars and refined grains, in favor of the consumption of non-starchy vegetables and foods that are high in protein.\textsuperscript{32}

However, emerging data demonstrates that there are marked inter-individual responses to food,\textsuperscript{17} attributed to differences in physical activity,\textsuperscript{33} gut microbiome,\textsuperscript{18,19,34} and genetics\textsuperscript{35} including in variations in skeletal glucose transporters related to insulin resistance.\textsuperscript{36} For example, a study conducted in the US among non-diabetic individuals with a mean BMI of 27 found PPGR to a standardized meal of bagel and cream cheese ranged from 6 to 94 mg/dL.\textsuperscript{20} A similar study conducted in Israel enrolled non-diabetic individuals of whom three-quarters had a BMI $\geq$ 25 and found mean PPGR to bread and butter of 44 mg/dL*hour but the bottom decile had responses of $\leq$ 15 mg/dL*hour and the top decile has responses $\geq$ 79 mg/dL*hour.\textsuperscript{18} Similar data has been generated for individual without diabetes in the UK and US\textsuperscript{19,20} and for individuals with type 1 diabetes in Israel.\textsuperscript{37}
There is exceptionally limited data for variability in PPGR for individuals with DM2 in the peer-reviewed although it is highly likely that such variability exists. The primary goal of our study is to fill this void and to generate an India-specific machine-learning models on the basis of which PPGR can be predicted with high accuracy for DM2. Similar models have been built in other jurisdictions. For example, a machine learning algorithm trained on CGM data, dietary, activity, anthropometrics and gut microbiota for non-diabetic individuals in Israel was much more accurate at predicting PPGR than generic models based on the carbohydrate content or the amount of calories in a meal. A separate US based study had similar findings.

Among individuals with diabetes, a study in the Netherlands that included a small number of individuals with DM2 along with individuals with pre-diabetes and normal glucose metabolism, a machine learning model based on CGM data was highly accurate and predicting future glucose values but this study did not specifically evaluate the ability to predict PPGR. A US study of 1,000 patients of whom one-quarter had DM2 found that a machine learning model trained on CGM, HRM data and food logs was highly accurate at predicting PPGR but this study has, to our knowledge, only been published in abstract form. These studies have all relied on CGM data to make their predictions. While these devices are increasingly used, practice guidelines do not recommend their long-term use for most individuals with DM2. Accordingly, a key goal of our study will be to explore the ability to predict PPGR response without reliance on CGM data or with very limited blood sugar data from patients.

There are several limitations to our approach. Our approach is purposely pragmatic and is intended to simulate real-world circumstances for individuals with DM2 living in India. Similar to studies conducted in other jurisdictions, we rely on self-reported dietary and activity information. And, while we are auditing patient logs on an ongoing basis, there may nevertheless be issues with protocol adherence that may undermine the accuracy of the data we collect. Subjects are being recruited from clinics predominantly caring for individuals with diabetes, are required to have functional English literacy and a cellphone capable of running study specific devices.
Thus, our results may not be fully generalizable to patients who do not fulfill these criteria. Finally, some of enrollment has overlapped with the COVID-19 pandemic which may have influenced access to health care, dietary practices and glucose control for individuals with DM2.

In conclusion, this study will provide the first large scale examination variability in blood sugar responses to food in India and will be among the first to estimate PPGR variability for individuals with DM2 in any jurisdiction. Results from our study will generate data to facilitate the creation of machine learning models to predict individual PPGR responses and to ultimately facilitate the prescription of truly personalized diets for individuals with DM2.
TABLES AND FIGURES

FIGURE 1: Overall study design

- Health surveys
- Biometric measurements
- Baseline laboratory

<table>
<thead>
<tr>
<th>Activities</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dietary and activity logging</td>
</tr>
<tr>
<td>Continuous heart rate tracking</td>
</tr>
<tr>
<td>Continuous glucose monitoring</td>
</tr>
<tr>
<td>Intermittent standardized meals</td>
</tr>
<tr>
<td>Intermittent low intensity exercise</td>
</tr>
</tbody>
</table>

IN CLINIC (DAY 1)  AT HOME (DAY 1-15)
**TABLE 1: Subject inclusion and exclusion criteria**

| Inclusion Criteria | § Age ≥ 18 and <75 years  
|§| Physician-diagnosed type 2 diabetes treated with ≥1 oral hypoglycemic agents  
|§| (concurrent treatment with an injectable non-insulin agent is permitted)  
|§| Hemoglobin A1c (HbA1c) ≥ 7.0% recorded within the past 30 days  
|§| Mobile phone capable of running the protocol-specified applications  
|§| Functional English literacy  

| Exclusion Criteria | § Unable or unwilling to provide informed consent or comply with the study-specified procedures  
|§| Current use of prandial insulin including a continuous insulin infusion pump  
|§| Currently pregnant or planning to become pregnant  
|§| Estimated life expectancy ≤ 12 months  
|§| Active cancer  
|§| Myocardial infarction or stroke in the last 6 months  
|§| Receiving or planned to initiate dialysis for end-stage renal disease  
|§| Receiving oral or intravenous steroids  
|§| Any contraindication to using a continuous glucose monitor (CGM)  

# TABLE 2: Meal schedule during the at-home study period

<table>
<thead>
<tr>
<th>Day</th>
<th>Timing</th>
<th>Meal type</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Breakfast</td>
<td>Fasting</td>
</tr>
<tr>
<td></td>
<td>Lunch</td>
<td>As desired</td>
</tr>
<tr>
<td></td>
<td>Dinner</td>
<td>As desired</td>
</tr>
<tr>
<td>2</td>
<td>Breakfast</td>
<td>Perceived healthy meal</td>
</tr>
<tr>
<td></td>
<td>Lunch</td>
<td>Post meal exercise</td>
</tr>
<tr>
<td></td>
<td>Dinner</td>
<td>Typical mixed protein with dinner</td>
</tr>
<tr>
<td>3</td>
<td>Breakfast</td>
<td>Fiber supplement with standardized test meal</td>
</tr>
<tr>
<td></td>
<td>Lunch</td>
<td>Pre-meal exercise</td>
</tr>
<tr>
<td></td>
<td>Dinner</td>
<td>Alternative mixed protein with dinner</td>
</tr>
<tr>
<td>4</td>
<td>Breakfast</td>
<td>Fiber supplement prior to standardized test meal</td>
</tr>
<tr>
<td></td>
<td>Lunch</td>
<td>As desired</td>
</tr>
<tr>
<td></td>
<td>Dinner</td>
<td>As desired</td>
</tr>
<tr>
<td>5</td>
<td>Breakfast</td>
<td>Protein supplement with standardized test meal</td>
</tr>
<tr>
<td></td>
<td>Lunch</td>
<td>As desired</td>
</tr>
<tr>
<td></td>
<td>Dinner</td>
<td>Alternative mixed protein with dinner</td>
</tr>
<tr>
<td>6</td>
<td>Breakfast</td>
<td>Protein supplement prior to standardized test meal</td>
</tr>
<tr>
<td></td>
<td>Lunch</td>
<td>Added fiber with lunch</td>
</tr>
<tr>
<td></td>
<td>Dinner</td>
<td>Alternative mixed protein with added protein</td>
</tr>
<tr>
<td>7</td>
<td>Breakfast</td>
<td>Regular breakfast</td>
</tr>
<tr>
<td></td>
<td>Lunch</td>
<td>Added fruit</td>
</tr>
<tr>
<td></td>
<td>Dinner</td>
<td>Cooled carbohydrate</td>
</tr>
<tr>
<td>8</td>
<td>Breakfast</td>
<td>Regular breakfast with added protein</td>
</tr>
<tr>
<td></td>
<td>Lunch</td>
<td>Added fruit with protein</td>
</tr>
<tr>
<td></td>
<td>Dinner</td>
<td>Regular carbohydrate</td>
</tr>
<tr>
<td></td>
<td>Breakfast</td>
<td>Lunch</td>
</tr>
<tr>
<td>---</td>
<td>--------------------------------</td>
<td>-----------------------------------</td>
</tr>
<tr>
<td>9</td>
<td>Protein followed by carbohydrate</td>
<td>Water 30 minutes before lunch</td>
</tr>
<tr>
<td>10</td>
<td>Standardized test meal</td>
<td>As desired including dessert</td>
</tr>
<tr>
<td>11</td>
<td>Standardized test meal with post-meal exercise</td>
<td>As desired including desert and cinnamon</td>
</tr>
<tr>
<td>12</td>
<td>Standardized test meal with post-meal exercise</td>
<td>As desired</td>
</tr>
<tr>
<td>13</td>
<td>Water before late breakfast + post-meal exercise</td>
<td>Low glycemic index lunch</td>
</tr>
<tr>
<td>14</td>
<td>Water before late breakfast + exercise after meal</td>
<td>Low glycemic index lunch + exercise after meal</td>
</tr>
</tbody>
</table>
### TABLE 3: Nutritional composition of standardized test meals

<table>
<thead>
<tr>
<th>Meal type</th>
<th>Example meal</th>
<th>Energy (kcal)</th>
<th>Carbohydrate (g)</th>
<th>Fat (g)</th>
<th>Protein (g)</th>
<th>Fiber (g)</th>
<th>Carbohydrate (%energy)</th>
<th>Fat (%energy)</th>
<th>Protein (%energy)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Carbohydrate</td>
<td>2 aloo paratha with curd</td>
<td>474.4</td>
<td>61.5</td>
<td>12.5</td>
<td>7.7</td>
<td>8.3</td>
<td>51.9</td>
<td>23.8</td>
<td>6.46</td>
</tr>
<tr>
<td>Added fiber</td>
<td>2 aloo paratha with curd + fiber supplement</td>
<td>506.4</td>
<td>64.1</td>
<td>12.5</td>
<td>7.8</td>
<td>21.9</td>
<td>50.6</td>
<td>22.3</td>
<td>6.05</td>
</tr>
<tr>
<td>Added protein</td>
<td>2 aloo paratha with curd + protein supplement</td>
<td>594.4</td>
<td>62.9</td>
<td>14.6</td>
<td>31.9</td>
<td>8.3</td>
<td>42.3</td>
<td>22.1</td>
<td>21.4</td>
</tr>
</tbody>
</table>
REFERENCES


30. Chen T, Guestrin C. XGBoost: A Scalable Tree Boosting System. presented at: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; 2016; San Francisco, California, USA. [https://doi.org/10.1145/2939672.2939785](https://doi.org/10.1145/2939672.2939785)


