**PROBAST**

Study:

Proactive Identification of Patients with Diabetes at Risk of Uncontrolled Outcomes during a Diabetes Management Program: Conceptualization and Development Study Using Machine Learning

Step 2: Type of prediction study

**Is the study a diagnostic or a prognostic study?**

Diagnostic

**Is the study a development only, development and validation or validation only study?**

Development only

**What is the model of interest?**

LightGBM

**What is the outcome of interest?**

Diabetes risk assessment

Step 3: Assess risk of bias

**Domain 1: Participants**

**Describe the sources of data and criteria for participant selection**

Participants enrolled in the Livongo for Diabetes RDMP between January 1, 2019, and January 1, 2022, with an activated BG meter who met the criteria to be categorized as observable at month 12 in the program were included as study participants(N>200,000).

**1.1 Were appropriate data sources used, e.g. cohort, RCT or nested case-control study data?**

Y

**1.2 Were all inclusions and exclusions of participants appropriate?**

Y

**Risk of bias introduced by selection of participants:**

Low

**Rationale of bias rating**

General population of T1D and T2D patients without any further eligibility criteria

**Domain 2: Predictors**

**List and describe predictors included in the final model, e.g. definition and timing of assessment**

Participant attributes, including survey data, BG data, medication fill data, and health signals, were used for feature construction within the ML models.

Survey data gathered from the participants at enrollment included demographic information, such as age, gender, race, ethnicity, height, weight, BMI, language, and diabetes type.

BG data were measured through RDMP-provided BG meters with blood from a finger prick applied to a test strip, and participants were asked to select “feel tags” and “meal tags” from a set of options. Features were constructed from an accumulation of SMBG readings by 30-day aggregates, BG readings broken down by meal and feel tags, and A1c. The30-day BG check aggregates generated data of the total number of readings and the number of hypoglycemic and hyperglycemic BG levels. Since the ML models were designed in monthly checkpoints, 30-day aggregates were important indicators of a participant’s diabetes pattern change.

Medication fill data

Health signal data

**2.1 Were predictors defined and assessed in a similar way for all participants?**

Y

**2.2 Were predictor assessments made without knowledge of outcome data?**

Y

**2.3 Are all predictors available at the time the model intended to be used?**

Y

**Risk of bias introduced by predictors or their assessment**

Low

**Rationale of bias rating**

All predictors were assessed the same way and can be applied when intended

**Domain 3: Outcome**

**Describe the outcome, how it was defined and determined, and the time interval between predictor assessment and outcome determination:**

HbA1c cannot be directly measured with a BG meter, which only measures the current BG level in the blood. To calculate HbA1c from BG meter readings, an algorithm that considers the average BG levels and the frequency of measurements was used: A1c = [average glucose (mg/dL) –46.7]/28.7 [13]. In this algorithm, the more frequent and consistent the BG readings, the more accurate the estimate ofA1c (eA1c). A participant was considered “observable” if enough BG readings over 90 preceding days provided statistical confidence to estimate clinically meaningful A1c, otherwise the participant was considered “unobservable.”

Ground truth labeling as “cases” and “controls” for diabetes management conditions was performed using 12-month eA1c values as follows: (1) participants with month 12 eA1c≥7.5% were labeled as cases and defined as participants at risk of uncontrolled diabetes management outcomes and (2) participants with month 12 eA1c<7.5% were labeled as controls and defined as participants not at risk of uncontrolled diabetes management outcomes. The ratio of cases to controls in the study population was 23.5%-76.5%.

**3.1 Was the outcome determined appropriately?**

Y

**3.2 Was a pre-specified or standard outcome definition used?**

Y

**3.3 Were predictors excluded from the outcome definition?**

N

**3.4 Was the outcome defined and determined in a similar way for all participants?**

Y

**3.5 Was the outcome determined without knowledge of predictor information?**

N

**3.6 Was the time interval between predictor assessment and outcome determination appropriate?**

Y

**Risk of bias introduced by the outcome or its determination**

High

**Rationale of bias rating**

BG values were taken as predictors but they were also taken to calculate the eA1c value which served as outcome determination. Thus there is dependence of outcome to predictor.

**Domain 4: Analysis**

**Describe number of participants, number of candidate predictors, outcome events and events per candidate predictor**

Over 200,000 patients included. The ratio of cases to controls in the study population was 23.5%-76.5%.

**Describe how the model was developed, predictor selection and risk group definition**

As previously mentioned, from the first day of Livongo BGdevice activation, at each monthly step, a prediction ML model was trained to predict the binary outcome of the diabetes management status of each participant at 12 months in the program. At each month of the program journey, there were 2 separate models to develop based on observability.

The following considerations were taken to reduce bias in the ML model: class imbalance handling, feature selection and inclusion, observability consideration, missing imputation, cross-validation, and temporal changes.

**Describe whether and how the model was validated, either internally (cross validation, random split sample) or externally (e.g. temporal validation, geographical validation, different setting, different type of participants)**

In modeling, for each model, data were split into 2 subsets. The first subset, which contained a larger portion of the data, was used as training and validation data to train the model and tune the hyperparameters, as well as find the optimal parameters of the model to achieve the highest accuracy. The second subset was used as testing data, which was held separate from the training data and used to assess how well the model could work  
with new data. Each model randomly split the data into 85% training and 15% testing subsets.

**Describe the performance measures of the model, e.g. calibration, discrimination, classification, net benefit, and whether they were adjusted for optimism**

AUC, ACC, F1, REC, PRE, SPE

**Describe any participants who were excluded from the analysis**

**Describe missing data on predictors and outcomes as well as methods used for missing data**

For unobservable participants with missing eA1c values, a robust imputation approach was used, which involved a mixture of historical eA1c data interpolation, leveraging their past records, and incorporation of similarity features from other participants, considering factors such as age and diabetes medications. This approach aimed at reducing bias in the imputation process, ensuring a more accurate estimation of missing eA1c values.

**4.1 Were there a reasonable number of participants with the outcome?**

Y

**4.2 Were continuous and categorical predictors handled appropriately?**

Y

**4.3 Were all enrolled participants included in the analysis?**

U

**4.4 Were participants with missing data handled appropriately?**

Y

**4.5 Was selection of predictors based on univariable analysis avoided?**

Y

**4.6 Were complexities in the data (e.g. censoring, competing risks, sampling of controls)**

**accounted for appropriately?**

Y

**4.7 Were relevant model performance measures evaluated appropriately?**

Y

**4.8 Were model overfitting and optimism in model performance accounted for?**

N

**4.9 Do predictors and their assigned weights in the final model correspond to the results**

**from multivariable analysis?**

Y

**Risk of bias introduced by the analysis**

High

**Rationale of bias rating**

Not mentioned how many patients were excluded. Only a simple validation split with 85/15 was taken for validation.

**Overall Risk of bias**

High