# Three components of glucose dynamics – value, variability, and autocorrelation – are independently associated with coronary plaque vulnerability

**Authors**

Hikaru Sugimoto,[1] Ken-ichi Hironaka,[2] Tomoko Yamada,[3] Natsu Otowa-Suematsu,[3] Yushi Hirota,[3] Hiromasa Otake,[4] Ken-Ichi Hirata,[3] Kazuhiko Sakaguchi,[3] Wataru Ogawa,[3]* and Shinya Kuroda[1,2,5]*

**Affiliations**

[1]Department of Biochemistry and Molecular Biology, Graduate School of Medicine, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan

[2]Department of Biological Sciences, Graduate School of Science, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan

[3]Division of Diabetes and Endocrinology, Department of Internal Medicine, Kobe University Graduate School of Medicine, Kusunoki-cho 7-5-1, Chuo-ku, Kobe, Hyogo 650-0017, Japan

[4]Division of Cardiovascular Medicine, Department of Internal Medicine, Kobe University Graduate School of Medicine, Kusunoki-cho 7-5-1, Chuo-ku, Kobe, Hyogo 650-0017, Japan

[5]Lead contact

*Corresponding authors: Wataru Ogawa, ogawa@med.kobe-u.ac.jp; Shinya Kuroda, skuroda@bs.s.u-tokyo.ac.jp

## ABSTRACT

Impaired glucose homeostasis leads to numerous complications, with coronary artery disease (CAD) being a major contributor to healthcare costs worldwide. Given the limited efficacy of current CAD screening methods, we investigated the association between glucose dynamics and a predictor of coronary events measured by virtual histology-intravascular ultrasound (%NC), with the aim of predicting CAD using easy-to-measure indices. We found that continuous glucose monitoring (CGM)-derived indices, particularly average daily risk ratio (ADRR) and AC_Var, exhibited stronger predictive capabilities for %NC compared to commonly used indices such as fasting blood glucose (FBG), hemoglobin A1C (HbA1c), and plasma glucose level at 120 min during oral glucose tolerance tests (PG120). Factor analysis identified three distinct components underlying glucose dynamics – value, variability, and autocorrelation – each independently associated with %NC. ADRR was influenced by the first two components and AC_Var by the third. FBG, HbA1c, and PG120 were influenced only by the value component, making them insufficient for %NC prediction. Our results were validated using data sets from Japan (n=64), America (n=53), and China (n=100).

CGM-derived indices reflecting the three components of glucose dynamics can serve as more effective screening tools for CAD risk assessment, complementing or possibly replacing traditional diabetes diagnostic methods.

## INTRODUCTION

Diabetes mellitus (DM) is a metabolic disorder affecting more than 400 million people worldwide. Among its complications, coronary artery disease (CAD) accounts for a significant proportion of morbidity, mortality and healthcare costs in patients with type 2 DM (T2DM) (Bax et al., 2007). Various prognostic models (Fiarni et al., 2019; Ravaut et al., 2021) and diagnostic markers (Bax et al., 2007) have been developed to predict CAD; however, screening of CAD can be ineffective, costly, or laborious (Bax et al., 2007; Young et al., 2009). More effective approaches for identifying individuals at high risk for complications using readily available clinical variables are warranted.

Blood glucose levels are among the readily obtained predictors of the complications (Psoma et al., 2022). The disrupted conditions of glucose dynamics seen in impaired glucose tolerance (IGT) and T2DM are partly characterized by high concentrations of blood glucose levels (Monnier et al., 2008). High concentrations of blood glucose levels have been defined as having high hemoglobin A1c (HbA1c) levels, fasting blood glucose (FBG) levels, and plasma glucose concentration at 120 min during the oral glucose tolerance test (OGTT) (PG120) (Monnier et al., 2008). These indices, especially HbA1c, are associated with complications of T2DM (Selvin et al., 2010).

Recent studies have shown that glucose variability, in addition to absolute glucose concentration, significantly contributes to the prognosis of complications (Gerbaud et al., 2019; Gorst et al., 2015; Monnier et al., 2008; Psoma et al., 2022; Su et al., 2011; Zhou et al., 2018) and all-cause mortality (Cai et al., 2022). Continuous glucose monitoring (CGM) can estimate short-term glycemic variability (Service, 2013), and has been reported to predict T2DM complications (Tang et al., 2016). Standard deviation (Std) of glucose levels (CGM_Std), mean amplitude of glycemic excursion (MAGE), mean of daily difference (MODD), and continuous overlapping net glycemic action (CONGA) are established indices of glycemic variability, of which CGM_Std and MAGE are more highly correlated with

coronary plaque properties (Otowa-Suematsu et al., 2018). Among glucose level-related indices, including HbA1c and FBG, MAGE is an independent determinant of coronary plaque instability (Okada et al., 2015).

Other CGM-derived indices such as average daily risk ratio (ADRR), lability index (LI), J-index, mean absolute glucose (MAG), and glycemic risk assessment in diabetes equation (GRADE) have also been developed (Hill et al., 2011). We recently showed that AC_Mean and AC_Var, which are calculated from the autocorrelation function of glucose levels measured by CGM, can detect decreased abilities in glucose regulation that cannot be captured by FBG, HbA1c, or the other conventional CGM-derived indices (Sugimoto et al., 2023). The characteristics of glucose dynamics can also be estimated from insulin concentrations. The disposition index (DI), which is the product of insulin sensitivity and insulin secretion, reflects and predicts glycemic disability beyond FBG (Utzschneider et al., 2009). Several other glucose-related indices and the relationship between the indices have also been reported (Fabris et al., 2015, 2014; Keshet et al., 2023). Despite these advances, a comprehensive understanding of how these various indices can be optimally combined to predict T2DM complications, particularly CAD, remains elusive. Furthermore, the underlying factors that these indices represent and their individual associations with CAD remain to be fully elucidated.

This study aims to address these knowledge gaps through three objectives: (i) to determine which clinical parameters are effective predictors of coronary plaque vulnerability; (ii) to identify the factors underlying these indices; and (iii) to elucidate how these factors are associated with coronary plaque vulnerability. We investigated the characteristics of 14 CGM-derived indices: 12 relatively well-known CGM-derived indices (Hill et al., 2011) and 2 indices (AC_Mean and AC_Var) as well as OGTT-derived indices, and investigated the relationship between these parameters and coronary plaque vulnerability assessed by virtual histology-intravascular ultrasound (VH-IVAS). We showed that three factors, namely, value,

variability, and autocorrelation, underly blood glucose level-related indices, and that the three are independently associated with coronary plaque vulnerability.

## RESULTS

### Mean, standard deviation, and autocorrelation function of glucose levels independently contribute to the prediction of coronary plaque vulnerability

We previously reported that AC_Var, calculated from the autocorrelation function of glucose levels, can capture decreased glucose handling capacities that cannot be captured by the mean (CGM_Mean) and standard deviation (CGM_Std) of glucose levels measured by CGM (Sugimoto et al., 2023). Based on the study, we hypothesized that AC_Var could identify individuals with high %NC, a widely used parameter of plaque vulnerability, independently from CGM_Mean and CGM_Std. To test this hypothesis, we conducted multiple regression analysis with CGM_Mean, CGM_Std, and AC_Var as input variables and %NC as the objective variable (Fig. 1A). For comparison, we also performed multiple regression analysis with established diagnostic markers of diabetes (FBG, HbA1c, and PG120) as input variables (Fig. 1B). We performed this analysis using a previously described cohort consisting of 8 individuals with NGT, 16 with IGT, and 29 with T2DM (Otowa-Suematsu et al., 2018).

The variance inflation factor (VIF) for CGM_Mean, CGM_Std, and AC_Var were 1.1, 1.1, and 1.0, respectively, indicating low multicollinearity among these variables. The $R^2$ of the model that predicted %NC from the three indices was 0.36. CGM_Mean, CGM_Std, and AC_Var had statistically significant independent positive correlations with %NC (Fig. 1A), suggesting that CGM_Mean, CGM_Std, and AC_Var are independently associated with %NC. In contrast, the $R^2$ of the model that predicted %NC from FBG, HbA1c, and PG120 was only 0.05 (Fig. 1B). Univariate and multivariate analyses including other indices further confirmed that several CGM-derived indices were significantly correlated with %NC, even after adjustment for multiple testing (Supplementary text).

**CGM-derived indices, particularly ADRR, AC_Var, MAGE, and LI, contribute to the prediction of coronary plaque vulnerability**

To address the challenge of unstable results when dealing with numerous variables and to investigate which variables are particularly useful in estimating %NC, we used two statistical techniques: Least Absolute Shrinkage and Selection Operator (LASSO) regression and Partial Least Squares (PLS) regression (Fig. 2) (Tibshirani, 1996; Wold et al., 2001). These regression models have been used for studies where the number of input variables is large relative to the sample size (Pei et al., 2023; Wang et al., 2005).

LASSO uses L1 regularization to produce models with fewer parameters and has been widely used for feature selection in predictive modeling (Wei et al., 2022). We included BMI, FBG, HbA1c, OGTT-derived indices, and CGM-derived indices as the input variables. The leave-one-out cross-validation identified the optimal regularization coefficient, lambda, as 0.849 (Fig. 2A). At the lambda, the coefficients of ADRR, AC_Var, MAGE, and LI were estimated to be non-zero coefficients (Fig. 2B, C), suggesting that CGM-derived indices, particularly ADRR, AC_Var, MAGE, and LI, contribute to the prediction of %NC. Even with the inclusion of SBP, DBP, TG, LDL-C, and HDL-C as additional input variables, the results remained consistent, with the coefficients of ADRR, AC_Var, MAGE, and LI still estimated as non-zero coefficients (Fig. S4).

To further validate the LASSO results and address potential instability, we performed PLS regression and examined the Variable Importance in Projection (VIP) scores (Fig. 2D). PLS regression is particularly useful when dealing with many input variables that may be highly collinear (Wold et al., 2001). The VIP scores of ADRR, AC_Var, MAGE, and LI, which were estimated to be non-zero coefficients by LASSO, were higher than 1, indicating that these four variables especially contribute to the prediction of %NC.

**Three components of glucose dynamics –value, variability, and autocorrelation – are associated with coronary plaque vulnerability**

To elucidate the underlying factors of clinical parameters and their association with %NC, we performed an exploratory factor analysis. Factor analysis reduces interrelated indices into a smaller set of underlying common factors, and has been employed to examine the interdependencies among various clinical parameters (Augstein et al., 2015; Cappelleri et al., 2000; Oh et al., 2004) and DM complications (Guo et al., 2020).

The optimal number of underlying factors was determined using Bayesian information criterion (BIC) and minimum average partial (MAP) methods, which indicated that five and six factors were appropriate, respectively. We first set the number of underlying factors as five. Figure 3A shows that FBG, HbA1c, PG120, I.I., oral DI, CGM_Mean, CONGA, HBGI, MVALUE, GRADE, JINDEX, and ADRR were included in the first factor. Given that most of these indices are related to the value of blood glucose concentration, factor 1 was labeled “value.” CGM_Std, MAGE, LI, MAG, MODD, JINDEX, and ADRR were included in the second factor. Given that these indices are related to glucose variability, factor 2 was labeled “variability.” Given that the definition of JINDEX is based on the sum of CGM_Mean and CGM_Std, and that of ADRR is based on both high and low values of glucose, the result that JINDEX and ADRR clustered in both factors 1 and 2 is plausible. Given that autocorrelation-derived indices, AC_Mean and AC_Var, were included in the third factor, factor 3 was labeled “autocorrelation.” BMI, PG120, composite index, and oral DI were included in the fourth factor. Factor 4 did not include any CGM-derived indices. Given that this combination of indices indicates a decrease in oral DI and associated increase in blood glucose due to decreased insulin sensitivity, factor 4 was labeled “sensitivity (without CGM)”. PG120, I.I., oral DI, and MAG were included in the fifth factor. Factor 5 did not have positive loadings of any CGM-derived indices. Given that this combination of the indices indicates a decrease in oral DI and associated increase in blood glucose due to

decreased insulin secretion (I.I.), factor 5 was labeled "secretion (without CGM)." The cumulative percentages of the total variance of the factors were 39%, 60%, 70%, 75%, and 80%, respectively.

The validity of the factor analysis was assessed according to previous studies (Cappelleri et al., 2000; Guo et al., 2020). To evaluate the applicability of the factor analysis, the Kaiser-Meyer-Olkin (KMO) and Bartlett's spherical test were performed. The KMO test indicated that the value of the measure of sampling adequacy for this data was 0.64, and Bartlett's spherical test indicated that the variables were statistically significantly intercorrelated ($P < 0.01$), suggesting that this dataset was applicable for the factor analysis. To evaluate internal consistency, Cronbach's α (Fig. 3B) and item–total correlations were calculated for each factor. Cronbach's α was 0.97 for factor 1, 0.93 for factor 2, 0.90 for factor 3, 0.72 for factor 4, and 0.66 for factor 5; these values were larger than 0.65 (Fig. 3B), suggesting that the internal consistency was satisfactory. While Cronbach's α of factor 5 was relatively low, exclusion of MAG increased the Cronbach's α to 0.84, indicating that the association between factor 5 and decrease in oral DI and associated increase in blood glucose due to decreased insulin secretion could be considered reliable. Item–total correlations ranged from 0.63 to 0.97 for factor 1, 0.72 to 0.94 for factor 2, 0.82 for factor 3, 0.54 to 0.76 for factor 4, and 0.37 to 0.86 for factor 5. With the exception of MAG, item–total correlations ranged from 0.84 to 0.91 for factor 5. The correlation coefficient of MAG was 0.37, which can be considered a modest correlation (Cappelleri et al., 2000), and the item–total correlations were generally reasonably strong in demonstrating reliability.

We also investigated a 6-factor solution (Fig. S5A). Factors 1, 2, and 3 could be interpreted as value, variability, and autocorrelation, respectively, similar to the 5-factor solution. Given that factor 6 had no factor loadings ≥ 0.3, we applied the 5-factor solution in the subsequent analysis. Furthermore, the inclusion of SBP, DBP, TG, LDL-C, and HDL-C into the input variables did not change the presence of the three components (value,

variability, and autocorrelation) in glucose dynamics (Fig. S5B). Since we only included only individuals with well-controlled serum cholesterol and BP levels in this study, we applied the 5-factor solution without these indices (Fig. 3A) to the following analysis.

To further examine the stability of the results of the factor analysis, we also conducted hierarchical clustering analysis (Fig. S6). The optimal number of clusters was determined based on silhouette analysis. A large positive silhouette coefficient indicates that each cluster is compact and distinct from the others. The analysis indicated that the four clusters were appropriate (Fig. S6A). Clusters 1, 2, and 3 can be interpreted as value, variability, and autocorrelation, respectively (Fig. S6B).

To investigate the association between these underlying factors and %NC, we investigated the correlation between the factor scores and %NC (Fig. 3C). The factor value and variability showed significant positive correlations with %NC, whereas autocorrelation showed a significant negative correlation. Factors 4 and 5, which were less related to the CGM-derived indices, showed weaker correlations with %NC. Collectively, we conclude that glucose dynamics has three components – value, variability, and autocorrelation – and that these three components are associated with %NC.

To assess the robustness and generalizability of our factor analysis results, we performed similar analyses using previously published datasets from diverse populations (Figs. S7, 8). Factors that could be interpreted as representing the value, variability, and autocorrelation of glucose dynamics were consistently observed across diverse populations, including Japanese (Sugimoto et al., 2023) (Fig. S7A), American (Hall et al., 2018) (Fig. S7B), and Chinese (Zhao et al., 2023) (Fig. S8A) cohorts. In the Chinese dataset, all three factors were significantly different between individuals with and without diabetic macrovascular complications (Fig. S8B). Taken together, these results support the reproducibility and cross-cultural validity of our three-component model of glucose dynamics.

224

225 **Overview of the three components of glucose dynamics –value, variability, and**
226 **autocorrelation –**

227 We have shown that three components of glucose dynamics – value, variability, and
228 autocorrelation – are associated with %NC. To overview the characteristics of glucose
229 dynamics with different values of the components, we simulated glucose fluctuations using a
230 previously reported mathematical model (De Gaetano and Arino, 2000) (Figs. 4, S9).

231 We could generate glucose fluctuations with almost the same standard deviation
232 (Std) and AC_Var but with a different mean (Fig. 4A). Similarly, we could also simulate
233 glucose fluctuations with almost the same mean and AC_Var but different Std, and with
234 almost the same mean and Std but different AC_Var. These three components were
235 independently adjustable by changing parameters within the range of values for NGT
236 individuals (Fig. S9B). Individuals with higher AC_Var tended to have higher %NC (Fig. 1);
237 however, comparing the glucose dynamics with higher and lower AC_Var, the maximum
238 value of blood glucose was lower in individuals with higher AC_Var (Fig. 4A).

239 To facilitate a more comprehensive exploration of the three components of glucose
240 dynamics, we developed a web-based application (https://simulator-glucose.streamlit.app/)
241 using a more detailed mathematical model (Dalla Man et al., 2007). This model incorporates
242 additional physiological parameters and allows for a more comprehensive simulation of
243 glucose-insulin interactions. The web application allows users to manipulate a wide range of
244 parameters, including insulin sensitivity, beta cell function, and insulin clearance. By
245 systematically varying these parameters, users can examine how the Mean, Std, and AC_Var
246 of glucose concentrations change under different conditions.

247 We also investigated the relationship between the three components and shapes of
248 the glucose response curve after OGTT. Patterns of the glucose response curve after OGTT
249 were heterogeneous, and four distinct patterns, denoted class 1–4, were previously identified

by analyzing the glucose dynamics of 5861 individuals in Denmark (Fig. 4B) (Hulman et al., 2018). The classes with a high mean did not necessarily have high Std or AC_Var (Fig. 4C). The classes with high Std did not necessarily have high mean and AC_Var (Fig. 4C), consistent with the result that mean, Std, and AC_Var had low multicollinearity with each other. Compared to class 2, only class 1 was lower in mean and Std, and was higher in AC_Var. Compared to class 2, only class 3 was higher in mean, Std, and AC_Var. Compared to class 2, only class 4 was higher in mean and Std, and lower in AC_Var. Collectively, the three components could characterize the previously reported four distinct patterns during OGTT.

Class 3 was characterized by normal FBG and PG120 values, but is reportedly associated with increased risk of diabetes and higher all-cause mortality rate, suggesting that subgroups at high risk may not be identified by investigating only FBG and PG120 (Hulman et al., 2018). Std and AC_Var were high in class 3 (Fig. 4C), suggesting that high Std and high AC_Var indicate glycemic disability independent of FBG and PG120.

## DISCUSSION

Here, we identified three distinct components of glucose dynamics: value, variability, and autocorrelation, each independently associated with coronary plaque vulnerability. We previously reported that AC_Var, an index reflecting autocorrelation, can detect decreased abilities in glucose regulation independently of other CGM-derived indices including CGM_Mean and CGM_Std, which reflect value and variability components, respectively (Sugimoto et al., 2023). Diabetes diagnosis has been based on elevated FBG, PG120, and HbA1c levels. However, these indices primarily reflect only the value component of glucose dynamics, and consequently the predictive performance of the prediction model for %NC using FBG, PG120, and HbA1c was relatively modest compared to that of the model using all three components of glucose dynamics. This result is partially consistent with a previous

notion that glucose dynamics include two components: amplitude and timing (Cobelli and Facchinetti, 2018). Collectively, CGM-derived indices reflecting the three components of glucose dynamics – value, variability, and autocorrelation – can outperform indices used to diagnose diabetes reflecting only the value component in predicting glycemic control capacity and coronary plaque vulnerability.

We also showed that CGM-derived indices, especially ADRR and AC_Var, contribute to the prediction of %NC by using LASSO and PLS (Fig. 2). Given that the definition equation for ADRR is affected by both high and low concentrations of blood glucose (Hill et al., 2011), it is likely affected by both glucose concentration values and glycemic variability. Factor analysis (Fig. 3) also showed that ADRR was included in both factor 1 (value) and factor 2 (variability). Since three factors, value, variability, and autocorrelation, contribute independently to the prediction of the complication, it would be useful to examine ADRR, which is influenced by both value and variability, and AC_Var, which is influenced by autocorrelation, in predicting %NC with a minimal number of variables. Therefore, the result of the LASSO showing that ADRR and AC_Var are particularly effective in predicting %NC is consistent with the results of the factor analysis that the three components contribute to the prediction.

This study also provided evidence that autocorrelation can vary independently from the value and variability components by using simulated data. As shown in Figure 4, these three components could be varied independently by simply changing the parameters related to glucose regulation within the range of NGT individuals. In addition, simulated glucose dynamics indicated that even subjects with high AC_Var did not necessarily have high maximum and minimum blood glucose levels. This study also indicated that these three components qualitatively corresponded to the four distinct glucose patterns observed after glucose administration, which were identified in a previous study (Hulman et al., 2018). Glycemic variability is involved in T2DM complications by oxidative stress and endothelial

dysfunction (Ceriello et al., 2008; Monnier et al., 2008); however, the reasons why the three components, especially autocorrelation, independently contribute to the prediction remain unknown. The underlying biological mechanisms and effects of the three components on living systems need to be investigated in future studies.

The current study had several limitations. LASSO and factor analysis indicated that CGM-related features were particularly important in predicting %NC. However, these results do not mean that other clinical parameters do not associate with T2DM complications, because we only included subjects with well-controlled serum cholesterol and blood pressure levels in this study. A previous study identified components of interday variability and hypoglycemia in CGM-derived indices (Augstein et al., 2015) that were not observed in our analysis. This discrepancy may be due to the relatively small number of T2DM subjects in our study. We acknowledge that factor analyses of data from longer measurement periods, including more patients with T1DM and T2DM, could potentially yield different results. However, it is noteworthy that our analysis of longer-term CGM data sets from Japanese and American populations confirmed the existence of the same three factors - value, variability, and autocorrelation - in glucose dynamics. Moreover, even with the short measurement period, CGM-derived indices reflecting these three factors demonstrated superior predictive accuracy for %NC compared to traditional indices such as FBG, HbA1c, and PG120, underscoring the potential utility of CGM. Although time in range (TIR) was not included in the main analyses due to the small number of T2DM patients and the large number of patients with TIR greater than 70%, the CGM-derived indices still outperformed FBG, HbA1c, and PG120 in predicting %NC. The multiple regression analysis between factor scores and TIR showed that only factors 1 (value) and 2 (variability) were significantly associated with TIR (Fig. S10), further supporting the existence of three components in glucose dynamics and the potential value of examining AC_Var in addition to other CGM-derived indices.

We acknowledge the potential concern of multiple testing in our study. However, even after adjusting for multiple comparisons, the CGM-derived indices retained significant correlations with %NC (Fig. S1). The consistency of our findings across different analytical approaches (Lasso, PLS, and factor analysis) and different data sets further supports the robustness of our conclusions regarding the characteristics of glucose dynamics. While we used methods that assume linearity, such as LASSO, we also examined nonlinear relationships using Spearman's correlation for index relationships and factor loadings with %NC, and found significant associations. We acknowledge that some significant correlations appear to be relatively small. However, these findings, combined with our predictive models showing improved accuracy using CGM compared to traditional diabetes diagnostic indices, and the theoretical framework showing that conventional markers only consider the "value" component of glucose dynamics, can demonstrate the clinical significance. Finally, although we analyzed three different datasets with a total of 270 subjects, the sample size may still be considered relatively small to comprehensively examine the relationships between the variables in this study. Larger, prospective studies are needed to provide a more accurate assessment of these variables in predicting abnormalities and to further validate our findings.

In conclusion, glucose dynamics has three components: value, variability, and autocorrelation. These three components are associated with coronary plaque vulnerability. CGM-derived indices reflecting these three components can be valuable predictive tools for T2DM complications, compared to conventional diabetes diagnostic markers reflecting only the value component. This new predictive model has the potential to improve the diagnosis and management of diabetes worldwide. To facilitate this CGM-derived prediction, we created a web application that performs a multiple regression model with these three components as input variables (https://cgm-basedregression.streamlit.app/).

## METHODS

| Key Resources Table | | | | |
|---|---|---|---|---|
| **Reagent type** | **Designation** | **Source or reference** | **Identifiers** | **Additional information** |
| software, algorithm | SciPy v1.10.1 | Virtanen et al., 2020 | | |
| software, algorithm | scikit-learn v1.0.2 | https://scikit-learn.org/stable/ | | |

### Subjects and measurements

This was a retrospective observational study approved by the ethics committee of Kobe University Graduate School of Medicine (UMIN000018326; Kobe, Japan), as described previously (Otowa-Suematsu et al., 2018). The study included 53 participants who underwent a 75-g oral glucose tolerance test (OGTT), continuous glucose monitoring (CGM) with the use of an iPro2 CGM system (Medtronic, Northridge, CA, USA), and percutaneous coronary intervention (PCI). During PCI, VH-IVUS was carried out to assess the plaque components. Among the 53 participants, eight, 16, and 29 individuals were categorized as having normal glucose tolerance (NGT), impaired glucose tolerance (IGT) and T2DM, respectively. Of note, with a type I error of 0.05, a power of 0.8, and an expected correlation coefficient of 0.4, a sample size of 47 was required to detect a significant difference from zero in the correlation coefficient.

Detailed participant characteristics have been reported in the previous study (Otowa-Suematsu et al., 2018). Briefly, participants aged 20–80 years with LDL-C levels < 120 mg/dL under statin administration or < 100 mg/dL under other treatments for dyslipidemia, including lifestyle intervention, were included in this study. Participants with acute coronary syndrome, unsuitable anatomy for virtual VH-IVUS, poor imaging by

VH-IVUS, hemodialysis, inflammatory disease, shock, low cardiac output, or concurrent malignant disease were excluded from this study.

To validate the findings on glucose dynamics, the present study analyzed a previously reported data set from Japan (Sugimoto et al., 2023). The study was conducted in accordance with the Declaration of Helsinki and approved by the Kobe University Hospital Ethics Committee (Approval No. 1834). Briefly, individuals aged ≥ 20 years with no previous diagnosis of diabetes were recruited from Kobe University Hospital between January 2016 and March 2018. Exclusion criteria included use of medications that affect glucose metabolism (e.g., steroids, β-blockers), psychiatric disorders, pregnancy or lactation, and ineligibility as determined by treating physicians. Participants wore a CGM device (iPro; Medtronic, USA). The study included 52 individuals with NGT, nine with IGT, and three with T2DM.

Further validation of the glucose dynamics findings was performed using a previously reported CGM dataset (Dexcom G4 CGM System; Dexcom, Fort Lauderdale, FL, USA) obtained in the United States (Hall et al., 2018). The study included 53 individuals with no previous diagnosis of diabetes.

In addition, the present study analyzed a previously reported CGM (FreeStyle Libre H, Abbott Diabetes Care, Witney, UK) dataset from China (Zhao et al., 2023). Participants were recruited from the DiaDRIL registry at Shanghai East Hospital (September 2019 to March 2021) and Shanghai Fourth People's Hospital (June 2021 to November 2021). Inclusion criteria for this dataset were: diagnosis of diabetes according to the 1999 World Health Organization (WHO) criteria, age 18 years or older, willingness to provide informed consent, and CGM recording for at least three days. Exclusion criteria included reported alcohol or drug abuse, inability to comply with study protocols, or inability to participate as determined by the investigators. We extracted and analyzed glucose profile characteristics from the first three days of CGM data for each participant. 100 individuals with T2DM were

398 analyzed in this study. Our primary objective was to explore the relationship between the
399 CGM-derived indices and the presence of diabetic macrovascular complications.
400
401 **Calculation of clinical indices**
402 CGM-derived indices:
403 The CGM data collected in this study were obtained at different sampling frequencies in the
404 different data sets. For the two Japanese cohorts and the American cohort, glucose
405 measurements were collected at 5-minute intervals. In contrast, the Chinese cohort collected
406 glucose measurements at 15-minute intervals.
407 Fourteen CGM-derived indices were evaluated in this study: twelve well-established
408 CGM-derived indices (Hill et al., 2011) and two indices (AC_Mean and AC_Var) that have
409 been reported to capture glucose handling capacity independently of the aforementioned
410 twelve indices (Sugimoto et al., 2023). For the datasets with 5-minute sampling intervals (the
411 two Japanese datasets and the American dataset), AC_Mean and AC_Var were calculated as
412 the mean and variance of the autocorrelation functions at lags 1-30 of the glucose levels,
413 respectively. For the data set with 15-minute sampling intervals (the Chinese data set),
414 AC_Mean and AC_Var were calculated as the mean and variance of the autocorrelation
415 functions at lags 1-10 of the glucose levels, respectively. These calculations correspond to a
416 time window of 150 minutes. CGM_Mean and CGM_Std indicate the mean value and
417 standard deviation of glucose levels measured by CGM, respectively. CONGA, LI, JINDEX,
418 HBGI, GRADE, MODD, MAGE, ADRR, MVALUE, and MAG were calculated using
419 EasyGV software (Hill et al., 2011). The calculating formulae of these indices are shown in
420 Table S1.
421 In the two Japanese datasets and the American dataset, there was a relatively small
422 proportion of T2DM patients and a substantial number of subjects with time in range (TIR)
423 (Battelino et al., 2023; Larkin et al., 2019) values (glucose levels between 70 and 180 mg/dL)

424 greater than 70% (Hall et al., 2018; Otowa-Suematsu et al., 2018; Sugimoto et al., 2023).
425 Given this distribution, we decided not to include TIR in our primary analyses. However,
426 given the potential interest of this metric, we provided correlations between TIR and other
427 indices as supplementary information.
428
429 OGTT-derived indices:
430 Three OGTT-derived indices were calculated as previously described (Matsuda and
431 DeFronzo, 1999; Otowa-Suematsu et al., 2018). The insulinogenic index (I.I.) indicates
432 insulin secretion and is calculated from the ratio of the increment of immunoreactive insulin
433 (IRI) to that of plasma glucose at 30 min after onset of the OGTT. The composite index
434 indicates insulin sensitivity, which can be calculated from fasting plasma glucose, fasting IRI,
435 mean blood glucose levels, and mean serum IRI concentrations during the OGTT. The oral
436 disposition index (Oral DI) was calculated from the product of the composite index and the
437 ratio of the area under the insulin concentration curve from 0 to 120 minutes to that for
438 plasma glucose from 0 to 120 minutes, without using the data measured at 90 min, in the
439 OGTT.
440
441 VH-IVUS-derived index:
442 VH-IVUS was carried out using the Eagle Eye Platinum 3.5-Fr 20-MHz catheter (Volcano,
443 Rancho Cordova, CA, USA), as previously described (Otowa-Suematsu et al., 2018). The
444 intraclass correlation coefficients for interobserver and intraobserver reliability of external
445 elastic membrane volume were 0.95 and 0.97, respectively (Otowa-Suematsu et al., 2018),
446 indicating high reproducibility. The VH-IVUS categorized plaque into four components:
447 fibrous, fibrofatty, necrotic core, and dense calcium. Following the previous study, our
448 investigation focused specifically on the ratio of necrotic core to total plaque volume (%NC),

449 a widely used parameter of plaque vulnerability. For patients with multiple plaques, the
450 mean %NC was calculated.
451
452 **Prediction models and statistical analyses**
453 In this study, we conducted multiple linear regression, LASSO regression, and PLS
454 regression. The input variables in these models included the following 26 variables: BMI,
455 SBP, DBP, TGs, LDL-C, HDL-C, FBG, HbA1c, PG120, I.I., composite index, oral DI,
456 CGM_Mean, CGM_Std, CONGA, LI, JINDEX, HBGI, GRADE, MODD, MAGE, ADRR,
457 MVALUE, MAG, AC_Mean, and AC_Var. In conducting these models, z-score
458 normalization on each input variable was performed.
459 The predictive performance of multiple linear regression was evaluated by the
460 coefficient of determination ($R^2$), the adjusted coefficient of determination (Adj $R^2$), or AIC.
461 The multicollinearity of the input variables was estimated by VIF. LASSO regression is a
462 kind of linear regression with L1 regularization (Tibshirani, 1996). The optimal
463 regularization coefficient, lambda, was based on leave-one-out cross-validation and
464 mean-squared error (MSE). The importance of the input variables in predicting %NC was
465 evaluated by the VIP scores (Wold et al., 2001) that were generated from PLS regression.
466 These models were conducted using scikit-learn v1.0.2, a Python-based toolkit
467 (https://scikit-learn.org/stable/).
468 Relationships among indices were also evaluated using Spearman's correlation
469 coefficients ($r$), and the correlation coefficients were reported with 95% CIs through
470 bootstrap resampling. The number of resamples performed to form the distribution was set at
471 10000. Benjamini–Hochberg's multiple comparison test was also performed with a
472 significance threshold of $Q < 0.05$.
473
474 **Factor analysis and hierarchical clustering analysis**

The intercorrelations of the clinical parameters and their associations with %NC were assessed using exploratory factor analyses and hierarchical clustering analyses. We followed the previously reported approach (Cappelleri et al., 2000; Guo et al., 2020; Lakka et al., 2002; Oh et al., 2004) with some modifications in conducting our exploratory factor analyses. BIC and MAP methods were used to determine the number of underlying factors. Variables with factor loadings of ≥ 0.30 were used in interpretation. To improve the interpretation, orthogonal (varimax) rotation was used. To evaluate the applicability of the factor analysis, KMO and Bartlett's spherical test were performed. To evaluate internal consistency of each factor, Cronbach's α and item–total correlations were calculated. The association of the factor scores with %NC was assessed using Spearman's correlation.

Hierarchical clustering analysis was conducted using a method that combines a Euclidean distance measure and Ward linkage. I.I., composite index, oral DI, and AC_Mean were inverted negatively so that the value of indices increased in subjects with abnormalities. The quality of the hierarchical clustering analysis was evaluated based on silhouette analysis (Rousseeuw, 1987). These analyses were performed after Z score normalization using scikit-learn v1.0.2, a Python-based toolkit (https://scikit-learn.org/stable/).

**Mathematical models used for simulating the characteristics of glucose dynamics**

In simulating the characteristics of glucose dynamics, we used a simple and stable model (De Gaetano and Arino, 2000), which can be written as follows:

$$\frac{dG}{dt} = -k_{\mathrm{glu}}G - k_{\mathrm{sen}}IG + k_{\mathrm{pro}} + f$$

$$\frac{dI}{dt} = \frac{k_{\mathrm{sec}}}{k_{\mathrm{tim}}}\int_{t-k_{\mathrm{tim}}}^{t} G\,ds - k_{\mathrm{cle}}I$$

where the variables $G$ and $I$ denote blood glucose and insulin concentrations, respectively. We simulated 240-minutes profiles of $G$, and calculated the mean, Std, and AC_Var of $G$. The parameters were changed within the range participants could take (De Gaetano and

Arino, 2000). Five mg/dL/min glucose was applied for 10 minutes at 30 minutes as the external input of glucose $f$.

For more detailed simulations of the postprandial response, we used a complex, fine-grained model (Dalla Man et al., 2007). Briefly, the model consists of two interconnected subsystems: the glucose subsystem and the insulin subsystem. The glucose subsystem is represented by a two-compartment model. The first compartment ($G_{\mathrm{p}}$) represents glucose in plasma and rapidly equilibrating tissues, while the second compartment ($G_{\mathrm{t}}$) represents glucose in slowly equilibrating tissues. The dynamics of this subsystem is described by the differential equations, as follows:

$$\frac{dG_p}{dt} = -k_1 G_{\mathrm{p}} + k_2 G_{\mathrm{t}} + \mathrm{EGP}(t) + \mathrm{Ra}(t) - U_{\mathrm{ii}}(t) - E(t)$$

$$\frac{dG_t}{dt} = k_1 G_{\mathrm{p}} - k_2 G_{\mathrm{t}} - U_{\mathrm{id}}(t)$$

$$\frac{dG}{dt} = \frac{G_{\mathrm{p}}}{V_{\mathrm{G}}}.$$

In these equations, $k_1$ and $k_2$ are rate constants, EGP is endogenous glucose production, Ra is the rate of glucose appearance from meal absorption, $U_{\mathrm{ii}}$ is insulin-independent glucose utilization, $E$ is renal excretion, $U_{\mathrm{id}}$ is insulin-dependent glucose utilization, $G$ is plasma glucose concentration, and $V_{\mathrm{G}}$ is the distribution volume of glucose.

The insulin subsystem is also modeled with two compartments, one for plasma insulin ($I_{\mathrm{p}}$) and another for liver insulin ($I_{\mathrm{l}}$). The dynamics of this subsystem is described by the differential equations, as follows:

$$\frac{dI_l}{dt} = -(m_1 + m_3) I_{\mathrm{l}} + m_2 I_{\mathrm{p}} + \mathrm{S}(t)$$

$$\frac{dI_l}{dt} = -(m_2 + m_4) I_{\mathrm{p}} + m_1 I_{\mathrm{l}},$$

where, S is the insulin secretion rate, $m_1$, $m_2$, and $m_4$ are rate parameters, $m_4$ is peripheral insulin clearance. For simplicity, we assumed that insulin sensitivity and beta cell responsivity to glucose are constant. We simulated 48-hour profiles of $G$, and calculated the

Mean, Std, and AC_Var of $G$. The code that simulates the glucose dynamics is available from the repository (https://github.com/HikaruSugimoto/Simulator) and the web application (https://simulator-glucose.streamlit.app/). The simulations were conducted using SciPy v1.10.1 (Virtanen et al., 2020).

## Characterization of glucose patterns during the OGTT

We investigated the characteristics of previously reported glucose patterns during the OGTT (Hulman et al., 2018). In the study, 5861 subjects without diabetes in Denmark underwent the OGTT with measurements of glucose levels at three time points (0, 30, and 120 min), and four distinct glucose patterns associated with long-term outcomes including diabetes onset, cardiovascular disease, and all-cause mortality rate were identified. For the calculation of mean, Std, and AC_Var of glucose levels, each time point was linearly imputed. Here, AC_Var was calculated from the autocorrelation function at lags 1–20, as we had glucose data available for only 2 h after the OGTT.

## Acknowledgments

**Author Contributions.** H.S. analyzed the data. H.S., K.H., T.Y., N.O.S., Y.H., H.O., K.H., K.S., W.O., and S.K. wrote the manuscript. W.O. and S.K. supervised the study.

**Conflict of Interest.** The authors have no conflicts of interest to declare.

**Funding and Assistance.** This study was supported by the Japan Society for the Promotion of Science (JSPS) KAKENHI (JP21H04759), CREST, the Japan Science and Technology Agency (JST) (JPMJCR2123), The Uehara Memorial Foundation, and The Takeda Science Foundation.

## Data and code availability

We used only previously published data sets. The code that calculates AC_Mean and AC_Var and that performs regression analysis with CGM-derived indices as input variables

543 are available from the repository (https://github.com/HikaruSugimoto/CGM_regression_app)

544 and the web application (https://cgm-basedregression.streamlit.app/).

545

## Figure Legends

**Figure 1. Multiple regression analyses for predicting %NC.**

Multiple regression analysis between %NC and CGM_Mean, CGM_Std, and AC_Var (A). That between %NC and FBG, HbA1c, and PG120 (B). Scatter plots for predicted %NC versus measured %NC (the left). Each point corresponds to the values for a single subject. Bars represent the 95% CIs of the coefficients of the regression models (the right).

**Figure 2. LASSO and PLS regression analyses for predicting %NC.**

(A) Relationship between regularization coefficients (lambda) and the MSE based on the leave-one-out cross-validation in predicting %NC. Dotted vertical line indicates the optimal lambda, which provides the least MSE. The optimal lambda was 0.849.

(B) LASSO regularization paths along the lambda in predicting %NC. Cyan, magenta, and gray lines indicate the estimated coefficients of AC_Mean, AC_Var, and the other input variables, respectively. Dotted vertical line indicates the optimal lambda.

(C) Estimated coefficients with the optimal lambda. Only variables with non-zero coefficients are shown. Input variables include the following 21 variables: BMI, FBG, HbA1c, PG120, I.I., composite index, oral DI, CGM_Mean, CGM_Std, CONGA, LI, JINDEX, HBGI, GRADE, MODD, MAGE, ADRR, MVALUE, MAG, AC_Mean, and AC_Var.

(D) VIP generated from the PLS regression predicting %NC. Variables with a VIP $\geq$ 1 (the dotted line) were considered to significantly contribute to the prediction.

**Figure 3. Factor analysis of the clinical parameters.**

(A) Factor analysis after orthogonal rotation. The values and colors were based on the factor loadings. The columns represent each factor. The rows represent input indices.

(B) Cronbach's $\alpha$ for each factor. Bars represent the 95% CI.

(C) Scatter plots and fitted linear regression lines for factor scores versus %NC. Each point corresponds to the values for a single subject. *r* is Spearman's correlation coefficient, and the value in parentheses is the 95% CI.

**Figure 4. Overview of the three components of glucose dynamics.**

(A) 240 min simulated glucose concentration. The colors of the line are based on the mean value (Mean), Std, and AC_Var of the simulated blood glucose. Red and gray dotted horizontal lines indicate the minimum or maximum values of blood glucose, respectively.

(B) Previously reported patterns of blood glucose during the OGTT (Hulman et al., 2018). Green, class 1; light blue, class 2; dark blue, class 3; red, class 4.

(C) Mean, Std, and AC_Var of the glucose during the OGTT. Colors are based on the class shown in Figure 4B.

583 **References**

584 Augstein P, Heinke P, Vogt L, Vogt R, Rackow C, Kohnert K-D, Salzsieder E. 2015.
585 Q-Score: development of a new metric for continuous glucose monitoring that enables
586 stratification of antihyperglycaemic therapies. *BMC Endocr Disord* **15**:22.
587 Battelino T, Alexander CM, Amiel SA, Arreaza-Rubin G, Beck RW, Bergenstal RM,
588 Buckingham BA, Carroll J, Ceriello A, Chow E, Choudhary P, Close K, Danne T,
589 Dutta S, Gabbay R, Garg S, Heverly J, Hirsch IB, Kader T, Kenney J, Kovatchev B,
590 Laffel L, Maahs D, Mathieu C, Mauricio D, Nimri R, Nishimura R, Scharf M, Del
591 Prato S, Renard E, Rosenstock J, Saboo B, Ueki K, Umpierrez GE, Weinzimer SA,
592 Phillip M. 2023. Continuous glucose monitoring and metrics for clinical trials: an
593 international consensus statement. *Lancet Diabetes Endocrinol* **11**:42–57.
594 Bax JJ, Young LH, Frye RL, Bonow RO, Steinberg HO, Barrett EJ, ADA. 2007. Screening
595 for coronary artery disease in patients with diabetes. *Diabetes Care* **30**:2729–2736.
596 Cai J, Yang Q, Lu J, Shen Y, Wang C, Chen L, Zhang L, Lu W, Zhu W, Xia T, Zhou J. 2022.
597 Impact of the complexity of glucose time series on all-cause mortality in patients with
598 type 2 diabetes. *J Clin Endocrinol Metab*. doi:10.1210/clinem/dgac692
599 Cappelleri JC, Gerber RA, Kourides IA, Gelfand RA. 2000. Development and factor analysis
600 of a questionnaire to measure patient satisfaction with injected and inhaled insulin for
601 type 1 diabetes. *Diabetes Care* **23**:1799–1803.
602 Ceriello A, Esposito K, Piconi L, Ihnat MA, Thorpe JE, Testa R, Boemi M, Giugliano D.
603 2008. Oscillating glucose is more deleterious to endothelial function and oxidative
604 stress than mean glucose in normal and type 2 diabetic patients. *Diabetes*
605 **57**:1349–1354.
606 Cobelli C, Facchinetti A. 2018. Yet Another Glucose Variability Index: Time for a Paradigm
607 Change? *Diabetes Technol Ther* **20**:1–3.
608 Dalla Man C, Rizza RA, Cobelli C. 2007. Meal simulation model of the glucose-insulin
609 system. *IEEE Trans Biomed Eng* **54**:1740–1749.
610 De Gaetano A, Arino O. 2000. Mathematical modelling of the intravenous glucose tolerance
611 test. *J Math Biol* **40**:136–168.
612 Fabris C, Facchinetti A, Fico G, Sambo F, Arredondo MT, Cobelli C, MOSAIC EU Project
613 Consortium. 2015. Parsimonious Description of Glucose Variability in Type 2
614 Diabetes by Sparse Principal Component Analysis. *J Diabetes Sci Technol*
615 **10**:119–124.
616 Fabris C, Facchinetti A, Sparacino G, Zanon M, Guerra S, Maran A, Cobelli C. 2014.
617 Glucose variability indices in type 1 diabetes: parsimonious set of indices revealed by
618 sparse principal component analysis. *Diabetes Technol Ther* **16**:644–652.

Fiarni C, Sipayung EM, Maemunah S. 2019. Analysis and Prediction of Diabetes Complication Disease using Data Mining Algorithm. *Procedia Comput Sci* **161**:449–457.

Gerbaud E, Darier R, Montaudon M, Beauvieux M-C, Coffin-Boutreux C, Coste P, Douard H, Ouattara A, Catargi B. 2019. Glycemic Variability Is a Powerful Independent Predictive Factor of Midterm Major Adverse Cardiac Events in Patients With Diabetes With Acute Coronary Syndrome. *Diabetes Care* **42**:674–681.

Gorst C, Kwok CS, Aslam S, Buchan I, Kontopantelis E, Myint PK, Heatlie G, Loke Y, Rutter MK, Mamas MA. 2015. Long-term Glycemic Variability and Risk of Adverse Outcomes: A Systematic Review and Meta-analysis. *Diabetes Care* **38**:2354–2369.

Guo W, Zhou Q, Jia Y, Xu J. 2020. Cluster and Factor Analysis of Elements in Serum and Urine of Diabetic Patients with Peripheral Neuropathy and Healthy People. *Biol Trace Elem Res* **194**:48–57.

Hall H, Perelman D, Breschi A, Limcaoco P, Kellogg R, McLaughlin T, Snyder M. 2018. Glucotypes reveal new patterns of glucose dysregulation. *PLoS Biol* **16**:e2005143.

Hill NR, Oliver NS, Choudhary P, Levy JC, Hindmarsh P, Matthews DR. 2011. Normal reference range for mean tissue glucose and glycemic variability derived from continuous glucose monitoring for subjects without diabetes in different ethnic groups. *Diabetes Technol Ther* **13**:921–928.

Hulman A, Vistisen D, Glümer C, Bergman M, Witte DR, Færch K. 2018. Glucose patterns during an oral glucose tolerance test and associations with future diabetes, cardiovascular disease and all-cause mortality rate. *Diabetologia* **61**:101–107.

Keshet A, Shilo S, Godneva A, Talmor-Barkan Y, Aviv Y, Segal E, Rossman H. 2023. CGMap: Characterizing continuous glucose monitor data in thousands of non-diabetic individuals. *Cell Metab* **35**:758-769.e3.

Lakka H-M, Laaksonen DE, Lakka TA, Niskanen LK, Kumpusalo E, Tuomilehto J, Salonen JT. 2002. The metabolic syndrome and total and cardiovascular disease mortality in middle-aged men. *JAMA* **288**:2709–2716.

Larkin ME, Nathan DM, Bebu I, Krause-Steinrauf H, Herman WH, Higgins JM, Tiktin M, Cohen RM, Lund C, Bergenstal RM, Johnson ML, Arends V, GRADE Research Group. 2019. Rationale and Design for a GRADE Substudy of Continuous Glucose Monitoring. *Diabetes Technol Ther* **21**:682–690.

Matsuda M, DeFronzo RA. 1999. Insulin sensitivity indices obtained from oral glucose tolerance testing: comparison with the euglycemic insulin clamp. *Diabetes Care* **22**:1462–1470.

Monnier L, Colette C, Owens DR. 2008. Glycemic variability: the third component of the dysglycemia in diabetes. Is it important? How to measure it? *J Diabetes Sci Technol* **2**:1094–1100.

Oh J-Y, Hong YS, Sung Y-A, Barrett-Connor E. 2004. Prevalence and factor analysis of metabolic syndrome in an urban Korean population. *Diabetes Care* **27**:2027–2032.

Okada K, Hibi K, Gohbara M, Kataoka S, Takano K, Akiyama E, Matsuzawa Y, Saka K, Maejima N, Endo M, Iwahashi N, Tsukahara K, Kosuge M, Ebina T, Fitzgerald PJ, Honda Y, Umemura S, Kimura K. 2015. Association between blood glucose variability and coronary plaque instability in patients with acute coronary syndromes. *Cardiovasc Diabetol* **14**:111.

Otowa-Suematsu N, Sakaguchi K, Komada H, Nakamura T, Sou A, Hirota Y, Kuroda M, Shinke T, Hirata K-I, Ogawa W. 2018. Comparison of the relationship between multiple parameters of glycemic variability and coronary plaque vulnerability assessed by virtual histology-intravascular ultrasound. *J Diabetes Investig* **9**:610–615.

Pei X, Qi D, Liu J, Si H, Huang S, Zou S, Lu D, Li Z. 2023. Screening marker genes of type 2 diabetes mellitus in mouse lacrimal gland by LASSO regression. *Sci Rep* **13**:6862.

Psoma O, Makris M, Tselepis A, Tsimihodimos V. 2022. Short-term Glycemic Variability and Its Association With Macrovascular and Microvascular Complications in Patients With Diabetes. *J Diabetes Sci Technol* 19322968221146810.

Ravaut M, Sadeghi H, Leung KK, Volkovs M, Kornas K, Harish V, Watson T, Lewis GF, Weisman A, Poutanen T, Rosella L. 2021. Predicting adverse outcomes due to diabetes complications with machine learning using administrative health data. *NPJ Digit Med* **4**:24.

Rousseeuw PJ. 1987. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *J Comput Appl Math* **20**:53–65.

Selvin E, Steffes MW, Zhu H, Matsushita K, Wagenknecht L, Pankow J, Coresh J, Brancati FL. 2010. Glycated hemoglobin, diabetes, and cardiovascular risk in nondiabetic adults. *N Engl J Med* **362**:800–811.

Service FJ. 2013. Glucose variability. *Diabetes* **62**:1398–1404.

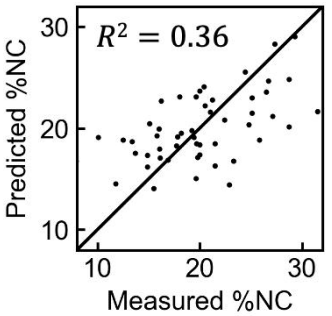
Su G, Mi S, Tao H, Li Z, Yang H, Zheng H, Zhou Y, Ma C. 2011. Association of glycemic variability and the presence and severity of coronary artery disease in patients with type 2 diabetes. *Cardiovasc Diabetol* **10**:19.

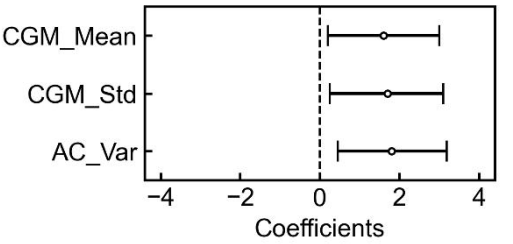
Sugimoto H, Hironaka K-I, Nakamura T, Yamada T, Miura H, Otowa-Suematsu N, Fujii M, Hirota Y, Sakaguchi K, Ogawa W, Others. 2023. Improved Detection of Decreased Glucose Handling Capacities via Novel Continuous Glucose Monitoring-Derived Indices: AC_Mean and AC_Var. *medRxiv* 2023–2009.

Tang X, Li S, Wang Y, Wang M, Yin Q, Mu P, Lin S, Qian X, Ye X, Chen Y. 2016. Glycemic variability evaluated by continuous glucose monitoring system is associated with the 10-y cardiovascular risk of diabetic patients with well-controlled HbA1c. *Clin Chim Acta* **461**:146–150.

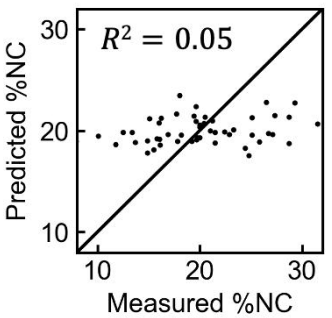
694 Tibshirani R. 1996. Regression shrinkage and selection via the lasso. *J R Stat Soc*
695 **58**:267–288.
696 Utzschneider KM, Prigeon RL, Faulenbach MV, Tong J, Carr DB, Boyko EJ, Leonetti DL,
697 McNeely MJ, Fujimoto WY, Kahn SE. 2009. Oral disposition index predicts the
698 development of future diabetes above and beyond fasting and 2-h glucose levels.
699 *Diabetes Care* **32**:335–341.
700 Virtanen P, Gommers R, Oliphant TE, Haberland M, Reddy T, Cournapeau D, Burovski E,
701 Peterson P, Weckesser W, Bright J, van der Walt SJ, Brett M, Wilson J, Millman KJ,
702 Mayorov N, Nelson ARJ, Jones E, Kern R, Larson E, Carey CJ, Polat İ, Feng Y,
703 Moore EW, VanderPlas J, Laxalde D, Perktold J, Cimrman R, Henriksen I, Quintero
704 EA, Harris CR, Archibald AM, Ribeiro AH, Pedregosa F, van Mulbregt P, SciPy 1.0
705 Contributors. 2020. SciPy 1.0: fundamental algorithms for scientific computing in
706 Python. *Nat Methods* **17**:261–272.
707 Wang C, Kong H, Guan Y, Yang J, Gu J, Yang S, Xu G. 2005. Plasma phospholipid
708 metabolic profiling and biomarkers of type 2 diabetes mellitus based on
709 high-performance liquid chromatography/electrospray mass spectrometry and
710 multivariate statistical analysis. *Anal Chem* **77**:4108–4116.
711 Wei H, Sun J, Shan W, Xiao W, Wang B, Ma X, Hu W, Wang X, Xia Y. 2022.
712 Environmental chemical exposure dynamics and machine learning-based prediction of
713 diabetes mellitus. *Sci Total Environ* **806**:150674.
714 Wold S, Sjöström M, Eriksson L. 2001. PLS-regression: a basic tool of chemometrics.
715 *Chemometrics Intellig Lab Syst* **58**:109–130.
716 Young LH, Wackers FJT, Chyun DA, Davey JA, Barrett EJ, Taillefer R, Heller GV,
717 Iskandrian AE, Wittlin SD, Filipchuk N, Ratner RE, Inzucchi SE, DIAD Investigators.
718 2009. Cardiac outcomes after screening for asymptomatic coronary artery disease in
719 patients with type 2 diabetes: the DIAD study: a randomized controlled trial. *JAMA*
720 **301**:1547–1555.
721 Zhao Q, Zhu J, Shen X, Lin C, Zhang Y, Liang Y, Cao B, Li J, Liu X, Rao W, Wang C. 2023.
722 Chinese diabetes datasets for data-driven machine learning. *Sci Data* **10**:35.
723 Zhou JJ, Schwenke DC, Bahn G, Reaven P, VADT Investigators. 2018. Glycemic Variation
724 and Cardiovascular Risk in the Veterans Affairs Diabetes Trial. *Diabetes Care*
725 **41**:2187–2194.

A


$R^2 = 0.36$
Predicted %NC
Measured %NC
30
20
10
10
20
30


CGM_Mean
CGM_Std
AC_Var
−4
−2
0
2
4
Coefficients

B


$R^2 = 0.05$
Predicted %NC
Measured %NC
30
20
10
10
20
30

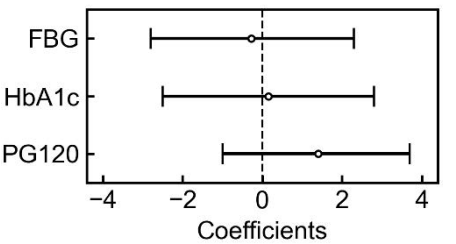

FBG
HbA1c
PG120
−4
−2
0
2
4
Coefficients

Fig. 1

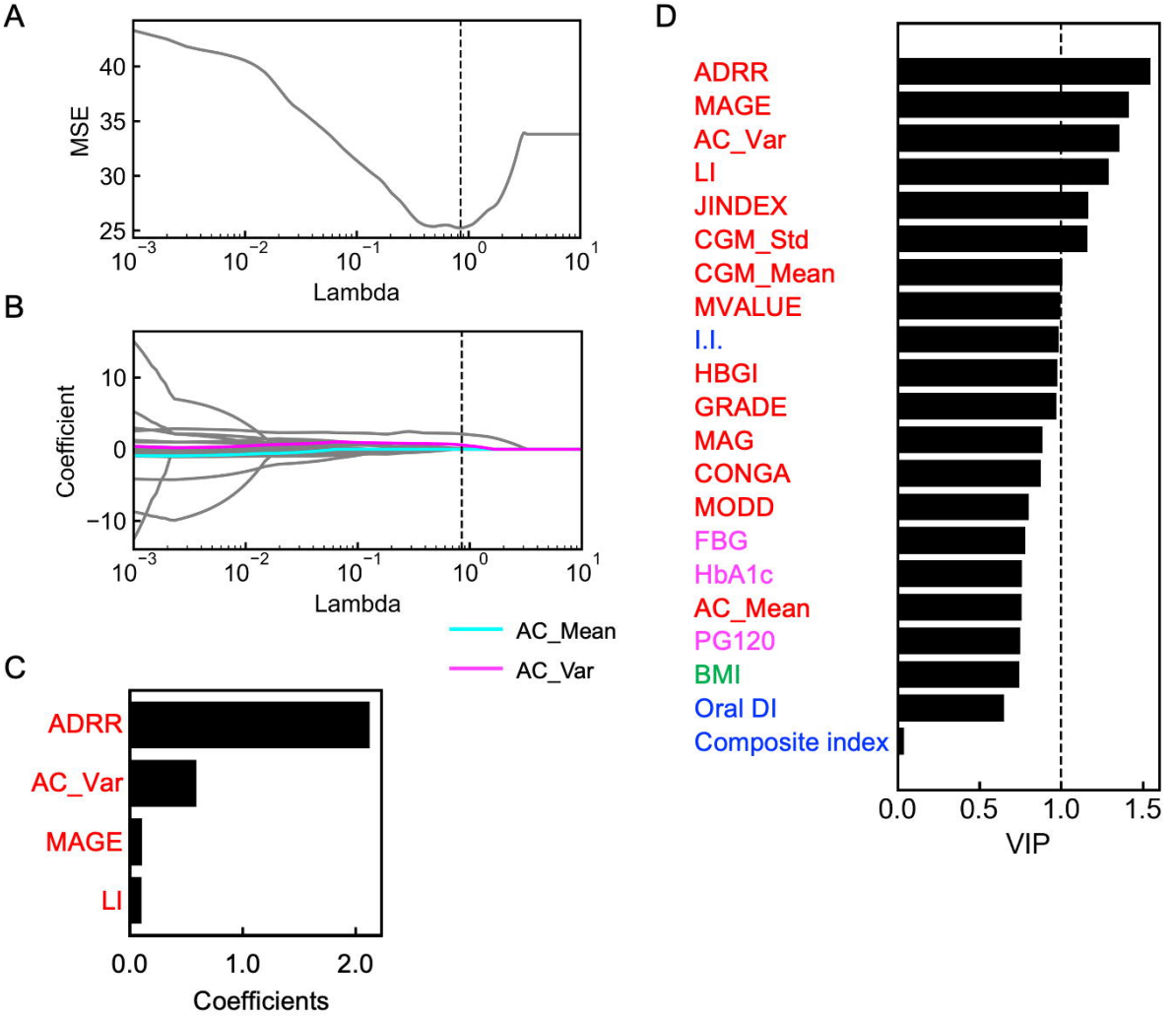

A
MSE
Lambda
B
Coefficient
Lambda
AC_Mean
AC_Var
C
ADRR
AC_Var
MAGE
LI
Coefficients
D
ADRR
MAGE
AC_Var
LI
JINDEX
CGM_Std
CGM_Mean
MVALUE
I.I.
HBGI
GRADE
MAG
CONGA
MODD
FBG
HbA1c
AC_Mean
PG120
BMI
Oral DI
Composite index
VIP

Fig. 2

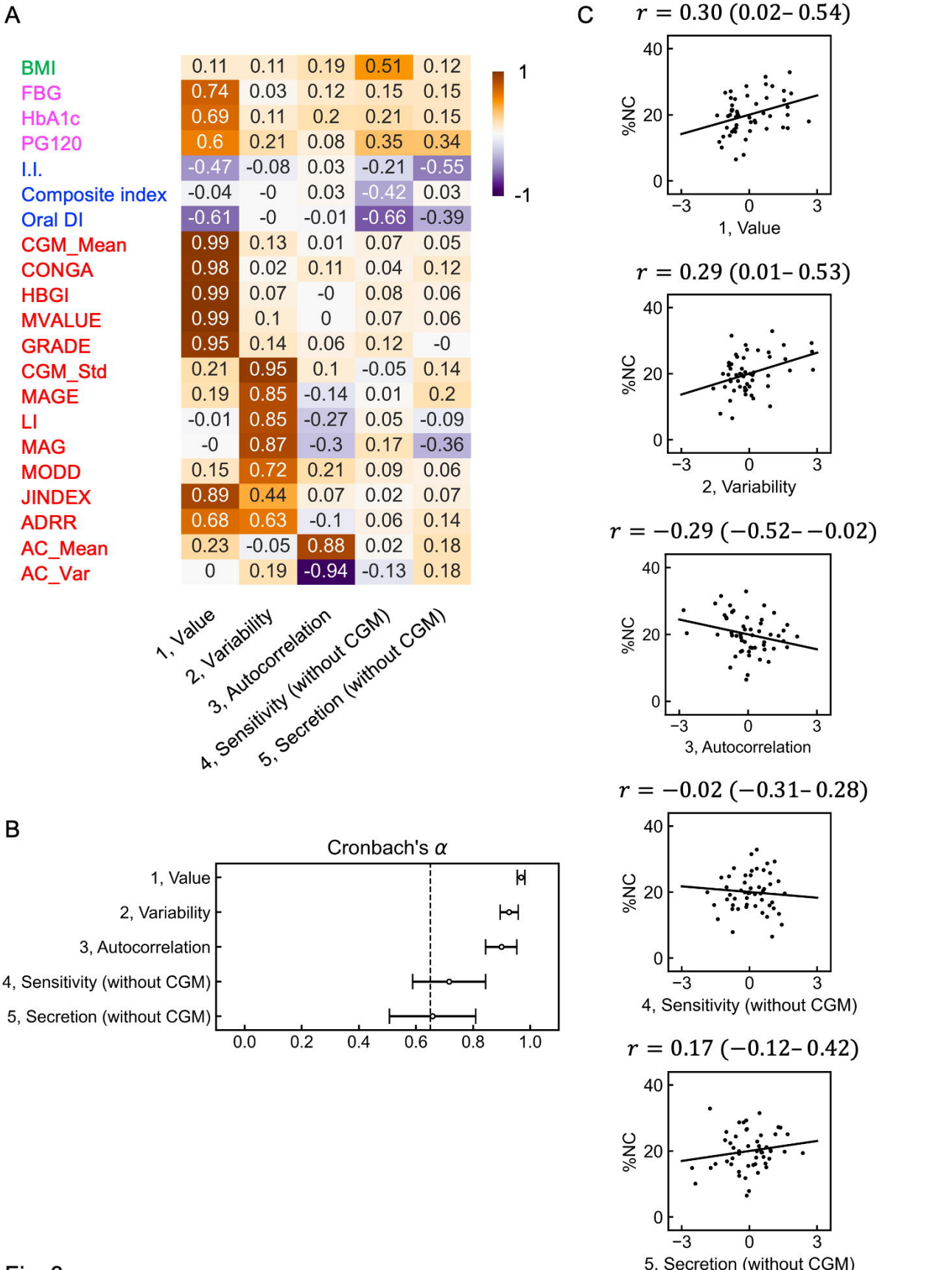

A
BMI 0.11 0.11 0.19 0.51 0.12
FBG 0.74 0.03 0.12 0.15 0.15
HbA1c 0.69 0.11 0.2 0.21 0.15
PG120 0.6 0.21 0.08 0.35 0.34
I.I. -0.47 -0.08 0.03 -0.21 -0.55
Composite index -0.04 -0 0.03 -0.42 0.03
Oral DI -0.61 -0 -0.01 -0.66 -0.39
CGM_Mean 0.99 0.13 0.01 0.07 0.05
CONGA 0.98 0.02 0.11 0.04 0.12
HBGI 0.99 0.07 -0 0.08 0.06
MVALUE 0.99 0.1 0 0.07 0.06
GRADE 0.95 0.14 0.06 0.12 -0
CGM_Std 0.21 0.95 0.1 -0.05 0.14
MAGE 0.19 0.85 -0.14 0.01 0.2
LI -0.01 0.85 -0.27 0.05 -0.09
MAG -0 0.87 -0.3 0.17 -0.36
MODD 0.15 0.72 0.21 0.09 0.06
JINDEX 0.89 0.44 0.07 0.02 0.07
ADRR 0.68 0.63 -0.1 0.06 0.14
AC_Mean 0.23 -0.05 0.88 0.02 0.18
AC_Var 0 0.19 -0.94 -0.13 0.18
1
-1
1, Value
2, Variability
3, Autocorrelation
4, Sensitivity (without CGM)
5, Secretion (without CGM)
B
Cronbach's α
1, Value
2, Variability
3, Autocorrelation
4, Sensitivity (without CGM)
5, Secretion (without CGM)
0.0
0.2
0.4
0.6
0.8
1.0
C
r = 0.30 (0.02– 0.54)
%NC
1, Value
r = 0.29 (0.01– 0.53)
%NC
2, Variability
r = −0.29 (−0.52– −0.02)
%NC
3, Autocorrelation
r = −0.02 (−0.31– 0.28)
%NC
4, Sensitivity (without CGM)
r = 0.17 (−0.12– 0.42)
%NC
5, Secretion (without CGM)

Fig. 3

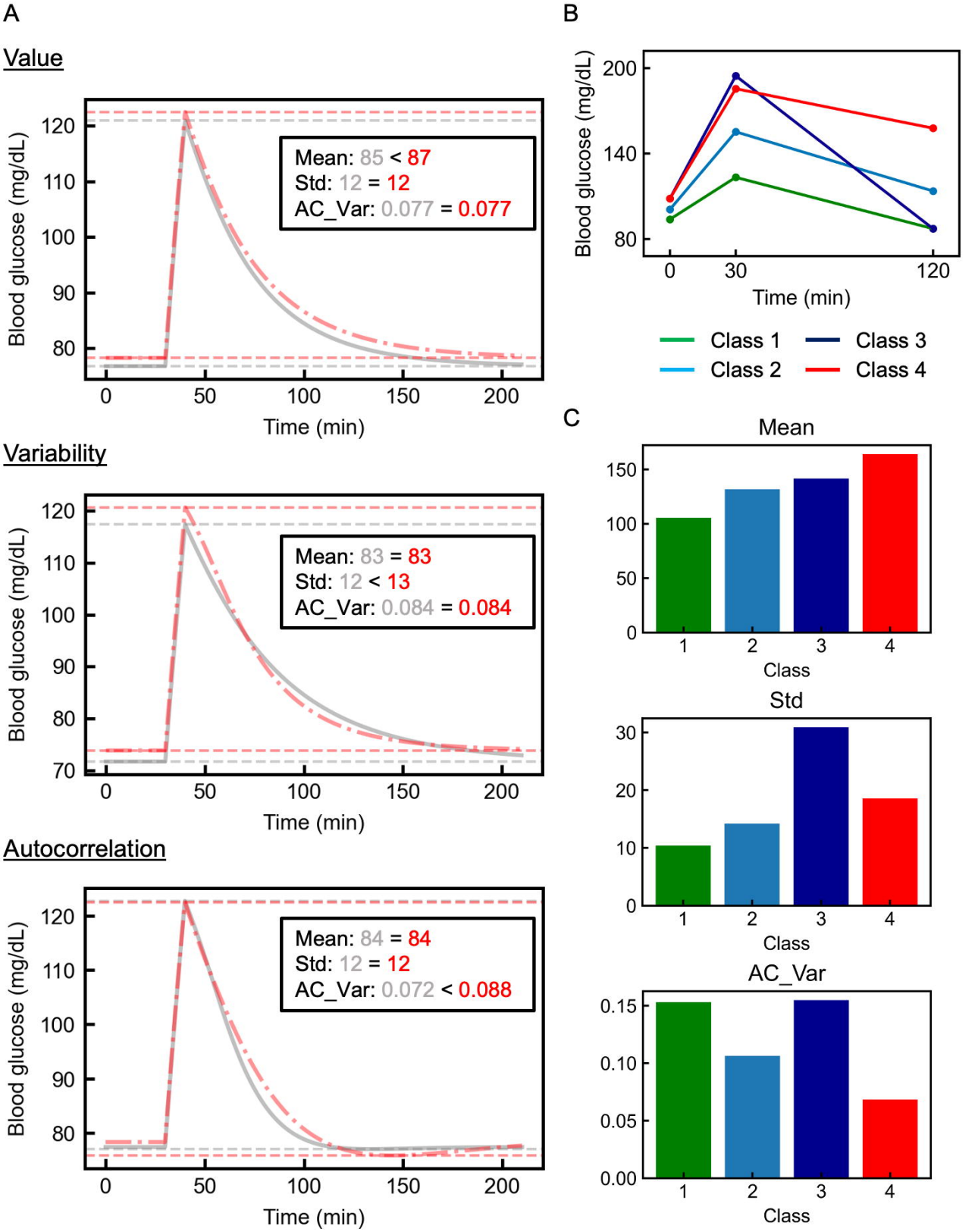

A
Value
Blood glucose (mg/dL)
Time (min)
Mean: 85 < 87
Std: 12 = 12
AC_Var: 0.077 = 0.077
Variability
Mean: 83 = 83
Std: 12 < 13
AC_Var: 0.084 = 0.084
Autocorrelation
Mean: 84 = 84
Std: 12 = 12
AC_Var: 0.072 < 0.088
B
Blood glucose (mg/dL)
Time (min)
Class 1
Class 2
Class 3
Class 4
C
Mean
Std
AC_Var
Class

Fig. 4