



# OPEN Glucodensity functional profiles outperform traditional continuous glucose monitoring metrics

Marcos Matabuena<sup>1✉</sup>, Rahul Ghosal<sup>2</sup>, Javier Enrique Aguilar<sup>3</sup>, Ayya Keshet<sup>4,5</sup>, Robert Wagner<sup>6,7,8</sup>, Carmen Fernández Merino<sup>9</sup>, Juan Sánchez Castro<sup>9</sup>, Vadim Zipunnikov<sup>10</sup>, Jukka-Pekka Onnela<sup>1</sup> & Francisco Gude<sup>11,12,13,14</sup>

Continuous glucose monitoring (CGM) data have revolutionized the management of type 1 diabetes, particularly when integrated with insulin pumps to mitigate clinical events such as hypoglycemia. Recently, there has been growing interest in utilizing CGM devices in clinical studies involving healthy and diabetic populations. However, efficiently exploiting the high temporal resolution of CGM profiles remains a significant challenge. Numerous indices—such as time-in-range metrics and glucose variability measures—have been proposed, but evidence suggests these metrics overlook critical aspects of dynamic glucose homeostasis. As an alternative method, this paper explores the clinical value of glucodensity metrics in capturing glucose dynamics—specifically the speed and acceleration of CGM time series—as new biomarkers for predicting long-term glucose outcomes. Our results demonstrate significant information gains, exceeding 20% in terms of adjusted r-square, in forecasting glycosylated hemoglobin (HbA1c) and fasting plasma glucose (FPG) at five and eight years from baseline AEGIS data, compared to traditional non-CGM and CGM glucose biomarkers. These findings underscore the importance of incorporating more complex CGM functional metrics, such as the glucodensity approach, to fully capture continuous glucose fluctuations across different time-scales.

**Keywords** Glucose dynamic, Continuous glucose monitoring, Glucose metabolism, Digital health, Functional data analysis

Recent technological advances in wearable devices and smartphones now enable near-continuous, time-dependent monitoring of various physiological parameters in humans, such as energy expenditure and heart rate<sup>1</sup>. Modern medical devices have become a significant driving force in the clinical evolution of digital health and personalized medicine<sup>2</sup>. One of the most notable examples of the impact of technological transformation in modern healthcare is the use of continuous glucose monitors (CGMs), a minimally invasive technology that measures interstitial glucose levels every few minutes. Over the past 20 years, CGMs have revolutionized the management of Type 1 diabetes, often integrated with insulin pump systems, significantly reducing high-risk clinical events such as hypoglycemia<sup>3</sup>.

Beyond their use in controlling glucose levels and managing Type 1 diabetes, CGMs have introduced new clinical criteria for assessing the effects of new drugs, including next-generation insulins, in both randomized clinical trials and observational studies<sup>4,5</sup>. Today, they are widely used to define new clinical outcomes in diabetes

<sup>1</sup>Department of Biostatistics, Harvard University, Cambridge, USA. <sup>2</sup>Department of Epidemiology and Biostatistics, University of South Carolina, Columbia, USA. <sup>3</sup>TU Dortmund, Dortmund, Germany. <sup>4</sup>Department of Computer Science and Applied Mathematics, Weizmann Institute of Science, Rehovot, Israel. <sup>5</sup>Department of Molecular Cell Biology, Weizmann Institute of Science, Rehovot, Israel. <sup>6</sup>Department of Endocrinology and Diabetology, Medical Faculty and University Hospital Düsseldorf, Heinrich Heine University Düsseldorf, Moorenstr. 5, 40225 Düsseldorf, Germany. <sup>7</sup>Institute for Clinical Diabetology, German Diabetes Center, Leibniz Center for Diabetes Research at Heinrich Heine University Düsseldorf, Aufm Hennekamp 65, 40225 Düsseldorf, Germany. <sup>8</sup>German Center for Diabetes Research (DZD), Ingolstädter Landstraße 1, 85764 Neuherberg, Germany. <sup>9</sup>Primary Care Center, A Estrada, Santiago, Spain. <sup>10</sup>Department of Biostatistics, Johns Hopkins University, Baltimore, USA. <sup>11</sup>Department of Clinical Epidemiology, Complejo Hospitalario Universitario, Santiago, Spain. <sup>12</sup>Department of Psychiatry, Radiology, Public Health, Nursing and Medicine, University of Santiago de Compostela, Santiago de Compostela, Spain. <sup>13</sup>Concepción Arenal Primary Care Center, SERGAS, Santiago de Compostela, Spain. <sup>14</sup>Research Methods Group (RESMET), Health Research Institute of Santiago de Compostela, Santiago de Compostela, Spain. ✉email: mmatabuena@hsph.harvard.edu

research, reshaping how clinical studies in diabetes are conducted. Additionally, CGM devices are essential tools for developing reliable quantitative methods in the field of personalized nutrition<sup>6</sup>. For example, recent epidemiological studies have revealed considerable variability in individual glucose responses to the same diets, offering new insights into the effects of meal composition—including macronutrient and micronutrient ratios—on glucose responses over time<sup>7</sup>. CGMs are now being used as a formal criterion for personalizing diets based on individual glycemic responses and for promoting nutritional interventions to optimize metabolic capacity<sup>6</sup>.

Despite strong evidence supporting the benefits of incorporating CGM technology into clinical practice to enhance glycemic control in diabetes, the use of CGMs in healthy, normoglycemic, and prediabetic individuals has not been widely explored or validated<sup>8</sup>. However, there is growing interest in using CGMs among healthy populations to estimate diabetes risk and develop interventions aimed at improving metabolic health<sup>9</sup>. Ongoing large-scale studies in the United States and Israel, involving thousands of participants, aim to understand how CGM data can predict diabetes onset, personalize dietary recommendations, and identify early metabolic dysfunctions. These studies seek to fill gaps in the literature regarding the utility of CGMs in non-diabetic populations, providing normative CGM data for glucose variability in healthy individuals<sup>10</sup>.

While these studies are still in their early stages, the population-based Spanish AEGIS study, with a 10-year follow-up, serves as a key reference for exploring CGM use in healthy populations and obtaining new clinical findings to be validated in larger ongoing cohorts. From the practical perspectives of primary care and preventive medicine, the validation of CGMs in non-diabetic populations is of paramount importance for developing preventive strategies and early interventions to reduce the incidence of diabetes and other metabolic disorders.

From a data analysis perspective, high-resolution CGM time series data present significant challenges, particularly in real-world clinical scenarios where patients are monitored in free-living conditions and for periods of different lengths. Direct analysis of these time series is often impractical, necessitating the development of specific CGM metrics<sup>11,12</sup>. Current state-of-the-art CGM metrics rely on compositional data analysis<sup>13</sup>, which defines glucose target ranges and quantifies the proportion of time an individual spends within each range<sup>14</sup>. Other, simpler methods summarize the time series using means, standard deviations, or other statistical moments<sup>15–18</sup>.

Recently, next-generation CGM analyses have been proposed that use functional data analysis<sup>19</sup> to exploit the continuous and functional nature of the data<sup>20</sup>, as seen in glucodensity representations and other distributional data metrics<sup>21,22</sup>. While glucodensity offers a continuous characterization of glucose distributions, it may overlook important dynamic aspects such as glucose variability patterns<sup>23,24</sup>, rapid fluctuations, and the timing of peaks and nadirs<sup>25</sup>—factors critical in understanding glycemic control and predicting complications. Previous studies have shown that glucose variability and excursions are associated with oxidative stress and vascular complications. Despite these limitations, glucodensity remains attractive due to its ability to provide a comprehensive and interpretable summary of glucose data, incorporating traditional CGM metrics automatically into the glucodensity representation.

Glucodensity<sup>20</sup> extends traditional compositional metrics like “time in range” by providing a continuous characterization of the proportion of time spent above or below each glucose concentration. Practically, glucodensity transforms the glucose time series into a marginal density function, automatically incorporating traditional CGM metrics. In some clinical applications, glucodensity has proven valuable in predicting clinical outcomes while maintaining interpretability—an essential feature in clinical research. However, new methods based on neural networks, such as autoencoders and other deep learning techniques, have been proposed<sup>26</sup> to summarize CGM data and integrate it with other data modalities. While these methods can capture complex temporal patterns and glucose dynamics, they often lack interpretability, making it difficult for clinicians to understand and explain the clinical results. In addition, neural network methods require large sample sizes and extensive computational resources, which may not be practical in clinical studies with small or moderate sample sizes, such as clinical trials. In contrast, glucodensity offers a balance between capturing essential glucose information and maintaining interpretability, making it a practical choice for small participant clinical studies.

Continuous glucose monitoring (CGM) has inspired a wide range of analytical methods, yet most investigations address only a single facet of glycaemic control. Early studies assessed variability using the standard deviation and the mean amplitude of glycaemic excursions (MAGE)<sup>23,27,28</sup>; however, these summaries do not reflect the continuous, temporally ordered nature of CGM profiles<sup>29</sup>. Phase-plane representations—plotting the glucose rate of change against glucose concentration—were subsequently introduced to characterise short-term kinetics<sup>30,31</sup>. Other research has quantified overall exposure cumulative-distribution functions of different CGM metrics across patients<sup>32</sup>, proposed clinically intuitive indices such as Time-in-Tight-Range (TITR)<sup>33</sup>, and shown that glucose distributions are better described by a gamma than by a normal law<sup>34</sup>. Principal functional component analyses have further distilled extensive collections of CGM indices into a few dominant patterns in both type 1 and type 2 diabetes<sup>35</sup>. The recently proposed *glucodensity* framework<sup>20,36</sup> reconciles these contributions by creating an individual CGM profile suitable for analyzing data under free-living conditions. Glucodensity measures the full distribution of an individual's CGM time series over a given period with a non-parametric density estimator; nevertheless, it still omits critical aspects of glucose dynamics. In this paper, we incorporate glucose speed and acceleration into the glucodensity framework to create a unified functional profile that captures both the distributional and dynamic characteristics of CGM data. To validate the new functional CGM index, we analyze baseline data from the AEGIS study, a cohort comprising a random sample of predominantly healthy individuals. We focus on two key biomarkers of diabetes diagnosis, control, and progression: fasting plasma glucose (FPG) and glycosylated haemoglobin (HbA<sub>1c</sub>)—at two future horizons, five and eight years. From a clinical perspective, the primary objective of the article, we provide solid evidence that glucose speed and acceleration, quantified via the glucodensity approach, are promising surrogate biomarkers for glucose control, offering a comprehensive outcome measure that encapsulates CGM information.

The implications of these findings for translational research could have substantial consequences in drug development. For instance, new insulins could be designed to respond more dynamically to rapid changes in glucose levels, optimizing dosing regimens based on individual glucose kinetics rather than static glucose concentrations. This could lead to more effective treatment schedules that minimize glycemic excursions and reduce the risk of complications. We must highlight that even in normoglycemic individuals, rapid glucose excursions—characterized by high velocity—and abrupt transitions—characterized by high acceleration—may signal early  $\beta$ -cell stress<sup>37</sup>. These dynamic patterns have been associated with both a diminished insulin response in the first phase and a blunted incretin effect<sup>38</sup>, two recognized subclinical markers of incipient  $\beta$ -cell dysfunction.

From a methodological point of view, our work opens up new opportunities to define functional clinical biomarkers for human glucose metabolism research and diabetes practice using advanced CGM metrics that capture both the distributional and dynamic aspects of glucose homeostasis metabolism. Recent work by Montaser et al.<sup>39</sup> has similarly highlighted the diagnostic potential of CGM-derived dynamic markers—such as the entropy rate and Poincaré plot metrics—to distinguish diabetes subtypes and detect early-stage dysglycemia. These findings reinforce the clinical relevance of CGM dynamics beyond traditional variability metrics.

This study offers two principal contributions. First, we extend the original concept of glucodensity<sup>20</sup> to a multidimensional setting by incorporating glucose velocity and acceleration, thereby capturing the full dynamical profile of continuous glucose monitoring (CGM) data. Second, we provide formal evidence—both theoretical and empirical—that multivariate glucodensity encodes substantially richer information than its marginal counterpart and than established CGM summary metrics. In the following, we detail these specific contributions.

- We introduce novel CGM data analysis methods based on glucodensity approach and distributional data analysis to capture glucose dynamics at different time scales, focusing on the speed and acceleration of glucose levels over time.
- We validate the proposed CGM metrics by predicting long-term glucose evolution at five and eight years, considering the biomarkers glycosylated hemoglobin (HbA1c) and fasting plasma glucose (FPG)—the primary biomarkers for diabetes control and diagnosis.
- We employ functional data analysis methods that efficiently leverage CGM data and distributional representations. This approach allows us to capture glucose dynamics in terms of speed and acceleration, translating into greater accuracy in predicting clinical outcomes.
- Our results show a 20% increase in  $R^2$  for predicting the continuous outcomes FPG and HbA1c at different temporal points when incorporating glucose marginal distributional patterns, speed, and acceleration as predictors in regression models, compared to regression models that only involve traditional CGM and non-CGM metrics. The traditional CGM and non-CGM models include: (i) baseline FPG and HbA1c; and (ii) baseline FPG and HbA1c, along with CGM metrics like the Area Under the Curve (AUC), Mean Amplitude of Glycemic Excursions (MAGE), Continuous Overall Net Glycemic Action (CONGA), and hyperglycemia time-in-range.

## Methods

In this section, we introduce the formal framework for analyzing continuous glucose monitor (CGM) data using the glucodensity (distributional data analysis) approach. This method incorporates the dynamics of glucose in terms of speed and acceleration, framed within the context of distributional glucose representations. We begin by presenting the existing concept of the distributional representation of glucose data.

Next, we extend this concept by proposing a novel multidimensional glucodensity approach that aggregates multiple physiological signals, specifically incorporating the speed (first derivative) and acceleration (second derivative) of the continuous-time glucose process. We then discuss the formal methods used to integrate these derivatives into the glucodensity representations, employing smoothing techniques for accurate estimation.

Finally, we introduce specific additive semi-parametric regression models that accommodate both marginal and multivariate distributional representations to predict continuous glucose scalar outcomes.

### Distributional representations for CGM-data analysis

#### Univariate Glucodensity approach

Following<sup>20</sup>, for an individual  $i^{th}$ , denote the glucose monitoring data by pairs  $(t_{ij}, G_{ij})$ ,  $j = 1, \dots, n_i$ , where the  $t_{ij}$  represent recording times that are typically equally spaced across the observation interval, and  $G_{ij}$  is the glucose level at time  $t_{ij} \in \mathcal{S}_i = [0, T_i]$ . Note that the number of records  $n_i$ , the spacing between them, and the overall observation length  $T_i$  can vary by individual. One can think of these data as discrete observations of a continuous latent process  $Y_i(t)$ , with  $G_{ij} = Y_i(t_{ij})$ . The glucodensity for this patient is defined in terms of this latent process as  $f_i(x) = F'_i(x)$ , where

$$F_i(x) = \frac{1}{T_i} \int_0^{T_i} \mathbf{I}(Y_i(t) \leq x) dt \quad \text{for} \quad \inf_{t \in [0, T_i]} Y_i(t) \leq x \leq \sup_{t \in [0, T_i]} Y_i(t) \quad (1)$$

is the proportion of the observation interval in which the glucose levels remain below  $x$ . Since  $F_i$  is increasing from 0 to 1, the data to be modeled are a set of probability density functions  $f_i$ ,  $i = 1, \dots, n$ .

Of course, neither  $F_i$  nor the glucodensity  $f_i$  is directly observed in practice. However, one can construct an approximation using a density estimate  $\hat{f}_i(\cdot)$  obtained from the observed sample. In the case of CGM data, the glucodensities may have different supports and shapes. Therefore, we suggest using a non-parametric approach, such as a kernel-type estimator, to estimate each density function. Alternatively, for individual  $i$ , we can use the quantile function  $\hat{\mathcal{Q}}_i(p) = \inf\{s \in \mathbb{R} : \hat{F}_i(s) \geq p\}$  as a representation, where  $\hat{F}_i(s) = \frac{1}{n_i} \sum_{j=1}^{n_i} \mathbf{I}(G_{ij} \leq s)$

Figure 1 illustrates graphically the process of collapsing CGM information from the raw CGM series into the quantile and density distributions. Density functions offer more interpretability than quantile functions because they provide a surrogate measure for the distribution of individuals across each glucose concentration. In contrast, quantile functions represent the profile of an individual patient. They are linear statistical objects that take values in vector-valued spaces and can be analyzed using standard linear functional data methods.

In metabolic disorders, such as diabetes, it is biologically plausible to assume that disease remains stable over certain periods. Consequently, we can hypothesize that physiological responses, including short-term glucose variation during these periods, conform to a stationary distribution, specifically over intervals  $\mathcal{S}_i$ . Therefore, the glucodensity approach is well-suited to capture individual distributional patterns within the examined temporal periods of glucose values.

### Multivariate glucodensity approach

We now extend the notion of glucodensity to cases where multiple physiological parameters are observed simultaneously—the multidimensional case.

For an individual  $i$ , we observe  $n_i$  measurements over the time interval  $\mathcal{S}_i = [0, T_i]$  from a medical device at the temporal instants  $\Gamma_i = \{t_{ij}\}_{j=1}^{n_i} \subset \mathcal{S}_i$  for  $m$  physiological variables recorded simultaneously at the same sampling frequency. The measurements are denoted as  $G_{ijk} = Y_{ik}(t_{ij})$ , where  $Y_{ik}(\cdot)$  is the underlying continuous model, and  $k = 1, \dots, m, i = 1, \dots, n, j = 1, \dots, n_i$ .

Throughout this paper, we consider  $m = 3$  physiological variables: glucose concentration (measured in mg/dL) and its first and second derivatives with respect to time, representing the speed and acceleration of glucose change. In the AEGIS study, the CGM device is continuously worn until the end of the monitoring period; thus, there is no missing data in the CGM time series collected. We then smooth the raw functional data trajectories via  $B$ -spline basis expansion and define

$$G_{ij1} = \hat{Y}_{i1}(t_{ij}),$$

where

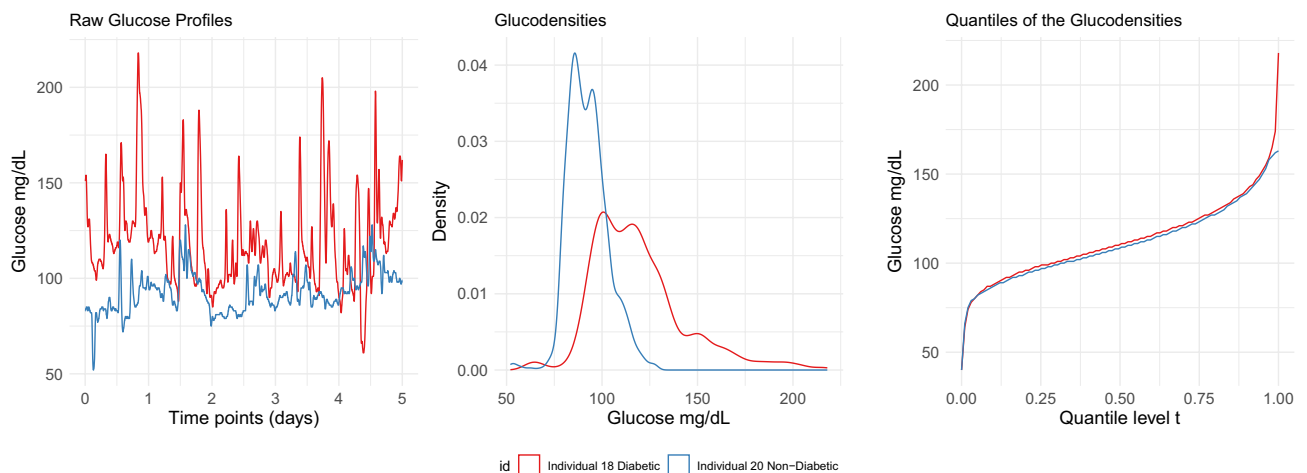
$$\hat{Y}_{i1}(t_{ij}) = \sum_{l=1}^{r_i} c_{il} \phi_l(t_{ij}), \quad i = 1, \dots, n, j = 1, \dots, n_i,$$

$\phi_l(t_{ij})$  are the basis functions predefined at the beginning of functional data smoothing, and  $c_{il}$  are the individual coefficients. The smoothing criterion consists of minimizing the mean squared error while introducing a quadratic penalty term to control the level of smoothness.

Formally, we define the sequences  $G_{ij2}$  and  $G_{ij3}$  as the evaluations at times  $t_{ij}$  of the first and second derivatives of the smoothing representation  $\hat{Y}_{i1}(t_{ij})$ , that is,

$$G_{ij2} = \hat{Y}'_{i1}(t_{ij}) \quad \text{and} \quad G_{ij3} = \hat{Y}''_{i1}(t_{ij}).$$

From a population standpoint, we define the cumulative distribution function (c.d.f.) of the multivariate process  $Y_i(t) = (Y_{i1}(t), \dots, Y_{im}(t))$  over the interval  $[0, T_i]$  as  $F_i(p_1, \dots, p_m)$ , described by the following equation:



**Fig. 1.** (Left) Continuous glucose monitor (CGM) time-series profiles for two non-diabetic individuals over six days, following<sup>20</sup>. (Center) Corresponding univariate glucodensity representations using the methodology proposed in<sup>20</sup>. (Right) Corresponding quantile functions derived from the glucodensity representations.

$$F_i(p_1, \dots, p_m) = \frac{1}{T_i} \int_0^{T_i} \mathbb{I}\{Y_i(t) \leq p\} dt.$$

The corresponding individual density function  $f_i(p_1, \dots, p_m)$  is defined as the gradient of  $F_i(p_1, \dots, p_m)$ :

$$f_i(p_1, \dots, p_m) = \nabla F_i(p_1, \dots, p_m).$$

From an empirical perspective, assuming the original process  $Y_{ik}(t)$  is smooth, we estimate the underlying marginal density functions using smoothing methods as follows:

$$\hat{f}_H(p) = \frac{1}{n} \sum_{i=1}^n K_H(p - p_i),$$

where  $p = (p_1, p_2, \dots, p_m)^\top$ ,  $p_i = (p_{i1}, p_{i2}, \dots, p_{im})^\top$  for  $i = 1, 2, \dots, n$  are  $m$ -vectors;  $H$  is a symmetric and positive definite  $m \times m$  matrix that serves as the bandwidth; and  $K$  is the kernel function, a symmetric multivariate density defined as:

$$K_H(p) = |H|^{-1/2} K(H^{-1/2}p).$$

To select the bandwidth matrix  $H$ , we can use rules of thumb based on asymptotic Gaussian processes or finite-sample data-driven approaches such as cross-validation (see more details in<sup>40</sup>).

## Regression modeling

### Modelling framework

We denote the scalar outcome of interest as  $R_i$ . Let us denote the scalar covariates e.g., age and other demographic predictors as  $X_i$ . The subject-specific marginal distributional representation using quantile function of glucose concentration, speed and acceleration is denoted by  $\mathcal{Q}_{i1}(\cdot)$ ,  $\mathcal{Q}_{i2}(\cdot)$  and  $\mathcal{Q}_{i3}(\cdot)$  respectively. The quantile function representation has been previously used in distributional data analysis<sup>22</sup> and offers attractive mathematical advantages. We observe a random sample  $D_i = (R_i, X_i, \mathcal{Q}_{i1}(\cdot), \mathcal{Q}_{i2}(\cdot), \mathcal{Q}_{i3}(\cdot))$ ,  $i = 1, \dots, n$ , which are assumed to be independent and identically distributed for each  $i$ .

### Scalar on distribution regression

We use the following additive scalar on distribution regression approach<sup>21,22</sup> for modelling the scalar outcome of interest based on multivariate distributional representations. A functional generalized additive model<sup>41</sup> is used for capturing additive nonlinear distributional effects of glucose concentration, speed and acceleration. The model is given by,

$$R_i = \alpha_0 + X_i^T \alpha + \int_{P_1} F_1(\mathcal{Q}_{i1}(p_1), p_1) dp_1 + \int_{P_2} F_2(\mathcal{Q}_{i2}(p_2), p_2) dp_2 + \int_{P_3} F_3(\mathcal{Q}_{i3}(p_3), p_3) dp_3 + \epsilon_i, \quad (2)$$

where  $F_j(\cdot, \cdot)$ ,  $j = 1, \dots, 3$  denote a unknown bivariate function capturing the additive effect of  $Q_{ij}(p_j)$  at index  $p_j$ . The unknown parameters of interest in the above model, which needs to be estimated are given by  $\Theta = (\alpha_0, \alpha, F_1(\cdot, \cdot), F_2(\cdot, \cdot), F_3(\cdot, \cdot))$ . Note that we do not need directly observed these distribution valued covariates, rather they are empirically estimated.

### Estimation

We employ a semiparametric penalized estimation approach to estimate the model parameters. The unknown bivariate functional effects  $F_j(\cdot, \cdot)$  ( $j=1, \dots, 3$ ) are modelled using tensor product of cubic B-spline basis functions as

$$F_j(u, p) = \sum_{k=1}^{K_0} \sum_{l=1}^{L_0} \theta_{j,kl} B_{j,U}(u) B_{j,\mathcal{P}}(p). \quad (3)$$

Here  $\{B_{j,U}(u)\}_{k=1}^{K_0}$ ,  $\{B_{j,\mathcal{P}}(p)\}_{l=1}^{L_0}$  are sets of known basis functions over  $u$  and  $p$  arguments respectively. We denote  $\theta_j = \{\theta_{j,kl}\}_{k=1, l=1}^{K_0, L_0}$  to be the unknown basis coefficients. In this article, we use cubic B-spline basis functions, however, other basis functions can be used as well. Plugging in these basis expansions in model (3) we have,

$$\begin{aligned}
R_i &= \alpha_0 + X_i^T \alpha + \sum_{j=1}^3 \int_{P_j} \sum_{k=1}^{K_0} \sum_{l=1}^{L_0} \theta_{j,kl} B_{j,U}(\mathcal{Q}_{ij}(p_j)) B_{j,P}(p_j) dp_j. \\
&= \alpha_0 + X_i^T \alpha + \sum_{j=1}^3 \sum_{k=1}^{K_0} \sum_{l=1}^{L_0} \theta_{j,kl} \int_{P_j} B_{j,U}(\mathcal{Q}_{ij}(p_j)) B_{j,P}(p_j) dp_j. \\
&= \alpha_0 + X_i^T \alpha + \sum_{j=1}^3 W_{ij}^T \theta_j.
\end{aligned} \tag{4}$$

Here  $\hat{W}_{ij} = \{\int_{P_j} B_{j,U}(\mathcal{Q}_{ij}(p_j)) B_{j,P}(p_j) dp_j\}_{k=1, l=1}^{K_0, L_0}$  denotes the  $K_0 L_0$ -dimensional stacked vectors which can be approximated using Riemann sum. Denote the unknown parameter of interest as  $\psi = (\alpha_0, \alpha^T, \theta_1^T, \theta_2^T, \theta_3^T)^T$ . We use the following penalized least square criterion to estimate the model parameters which simultaneously estimate the basis coefficients and enforces smoothness in the additive functional effects  $F_j(\cdot, \cdot)$  ( $j = 1, \dots, 3$ ).

$$\begin{aligned}
R_p(\psi|\lambda) &= R(\psi|\lambda) + \sum_{j=1}^3 \theta_j^T P_j \theta_j \\
&= \sum_{i=1}^n (R_i - \alpha_0 - X_i^T \alpha - \sum_{j=1}^3 \hat{W}_{ij}^T \theta_j)^2 + \sum_{j=1}^3 \theta_j^T \hat{P}_j \theta_j,
\end{aligned} \tag{5}$$

where  $P_1$  and  $P_2$  are roughness penalty matrices<sup>42</sup> to introduce smoothness in  $u$  and  $p$  direction respectively for each of the function  $F_j(\cdot, \cdot)$ . In this article, we have used a second order difference penalty<sup>43</sup> which introduces smoothness in both arguments  $u$  and  $p$ . The penalty matrices  $P_j$  ( $j = 1 \dots 3$ ) are given by  $\hat{P}_j = \lambda_{U,j} \hat{D}_{U,j}^T \hat{D}_{U,j} \otimes \hat{I}_{L_0} + \lambda_{P,j} \hat{I}_{K_0} \otimes \hat{D}_{P,j}^T \hat{D}_{P,j}$ , where  $\hat{I}_{K_0}, \hat{I}_{L_0}$  are identity matrices with dimension  $K_0, L_0$  and  $\hat{D}_{U,j}, \hat{D}_{P,j}$  are matrices corresponding to the row and column of second order difference penalties. The penalty parameters  $\lambda_{U,j}, \lambda_{P,j}$  ( $j = 1 \dots 3$ ) are the corresponding smoothing parameters, controlling smoothness of  $F_j(\cdot, \cdot)$  over  $u$  and  $p$  respectively. For fixed values of  $\lambda$ , Newton-Raphson algorithm can be employed for estimating  $\psi$  as the maximizer of the above penalized least square. We used the `gam` function within `mgcv` package in R<sup>44</sup> for estimating the parameter vector  $\psi$ . The smoothing parameters  $\lambda$  were selected with data-driven criteria, namely generalized cross-validation (GCV;<sup>45</sup>) and restricted maximum likelihood (REML;<sup>46</sup>). All smoothing terms were fitted using the `mgcv: :gam()` function in R, with smoothing parameters estimated by REML and term-wise shrinkage enabled via `select = TRUE`. This double-penalty strategy automatically balances goodness-of-fit and smoothness, shrinking non-informative smooths to (effectively) zero degrees of freedom<sup>47</sup>.

## Results

### Data description

The AEGIS population study<sup>48</sup> is a ten-year longitudinal study that focuses on changes in circulating glucose and its connections to inflammation and obesity. These factors are critically linked to the potential development of comorbidities and diabetes mellitus. Unlike other epidemiological studies, AEGIS incorporates continuous glucose monitoring (CGM) for a random subsample, providing detailed glucose profiles at various time points over a period of five years. In this random AEGIS CGM subsample, most individuals are non-diabetic (normoglycemic and prediabetic), offering a unique opportunity to evaluate the clinical value of CGM in healthy populations. The diabetic individuals in the sample have Type II diabetes and are very well-controlled, with a minimal progression of the disease at baseline. For this reason, we refer that with analysis, we are characterizing the long-term risk and progression of Diabetes Mellitus disease. We analyze approximately one week of data for each participant. Over such a period, it is biologically plausible that glucose measurements remain stationary. In diabetes—or when metabolism changes—alterations typically occur gradually over months (see for example<sup>7</sup>), so assuming weekly stationarity is reasonable.

### Sample and procedures

**Study design** A subset of the subjects in the A Estrada Glycation and Inflammation Study (AEGIS; trial NCT01796184 at [www.clinicaltrials.gov](http://www.clinicaltrials.gov)) provided the sample for the present work. In the latter cross-sectional study, an age-stratified random sample of the population (aged  $\geq 18$ ) was drawn from Spain's National Health System Registry. A detailed description has been published elsewhere<sup>48</sup>. For a one-year period beginning in March, subjects were periodically examined at their primary care center where they (*i*) completed an interviewer-administered structured questionnaire; (*ii*) provided a lifestyle description; (*iii*) were to biochemical measurements, and (*iv*) were prepared for CGM (lasting 6 days). The subjects who made up the present sample were the 581 (361 women, 220 men) who completed at least 2 days of monitoring, out of an 622 persons who consented to undergo a 6-day period of CGM. Another 41 original subjects were withdrawn from the study due to non-compliance with protocol ( $n = 4$ ) or difficulties in handling the device ( $n = 37$ ). The characteristics of the participants are shown in the Table 1.

	Men (n = 220)	Women (n = 361)	p-value (t-test)
Age, years	47.8 ± 14.8	48.2 ± 14.5	0.04
A1c, %	5.6 ± 0.9	5.5 ± 0.7	0.26
FPG mg/dL	97 ± 23	91 ± 21	< 0.001
HOMA-IR mg/dL·μUI/m	3.97 ± 5.56	2.74 ± 2.47	< 0.001
BMI kg/m <sup>2</sup>	28.9 ± 4.7	27.7 ± 5.3	< 0.001
CONGA mg/dL	0.88 ± 0.40	0.86 ± 0.36	0.03
MAGE mg/dL	33.6 ± 22.3	31.2 ± 14.6	< 0.001
MODD	0.84 ± 0.58	0.77 ± 0.33	< 0.001

**Table 1.** Characteristics of AEGIS study participants by sex. Mean ± standard deviation are shown. *BMI* - body mass index; *FPG* - fasting plasma glucose; *A1c* - glycated haemoglobin; *HOMA-IR* - homeostasis model assessment-insulin resistance; *CONGA* - glycemic variability in terms of continuous overall net glycemic action; *MODD* - mean of daily differences; *MAGE* - mean amplitude of glycemic excursions.

**Laboratory determinations** Glucose was determined in plasma samples from fasting participants by the glucose oxidase peroxidase method. A1c was determined by high performance liquid chromatography in a Menarini Diagnostics HA-8160 analyzer; all A1c values were converted to DCCT-aligned values<sup>49</sup>. Insulin resistance was estimated using the homeostasis model assessment method (HOMA-IR) as the fasting concentration of plasma insulin ( $\mu$  units/mL)  $\times$  plasma glucose (mg/dL)/ 405.

**CGM Procedures** At the start of each monitoring period, a research nurse inserted a sensor (Enlite<sup>TM</sup>, Medtronic, Inc, Northridge, CA, USA) subcutaneously into the subject's abdomen and instructed him/her in the use of the iPro<sup>TM</sup> CGM device (Medtronic, Inc, Northridge, CA, USA). The sensor continuously measures the interstitial glucose level 40 – 400 (range mg/dL) of the subcutaneous tissue, recording values every 5 min. The participants calibrated the iPro<sup>TM</sup> CGM with 3 premeal capillary tests per day using a OneTouch<sup>®</sup>Verio<sup>®</sup>Pro meter. Days with < 3 calibrations, > 2 h of signal loss, or the initial 24 h of wear were discarded. The resulting dataset—581 participants, 9 980 paired readings—showed a global MARD of 7.9 %, with 87 % of values within the ISO 15197 15/15 % limits, a level deemed clinically accurate for retrospective CGM analysis.

**Glycaemic variability** After downloading the recorded data, the following glycemic variability indexes were analyzed: area under the curve (AUC), mean amplitude of glycemic excursions (MAGE)<sup>27</sup>, continuous overlapping net glycemic action at 1 h (CONGA1)<sup>50</sup>, the mean of daily differences (MODD)<sup>51</sup>, and the time above range (TAR). The mathematical formulae of the methods of assessment for glucose variability were taken from their original publications for inclusion in an R program, which is freely available<sup>48</sup>.

**Advanced CGM-derived dynamic metrics, including semi-Markov (e.g., entropy rate) and Poincaré-based measures** In addition to standard CGM indices, we evaluated two information-theoretic metrics that characterize glucose dynamics: the entropy rate (ER) and the Poincaré plot ellipse area (S), recently introduced by Montaser et al.<sup>39</sup>. Both statistics were computed in R using an implementation that reproduces the original procedure described by<sup>39</sup>.

**Ethical approval and informed consent** The present study was reviewed and approved by the Clinical Research Ethics Committee from Galicia, Spain (CEIC2012-025). Written informed consent was obtained from each participant in the study, which was in accordance with the current Declaration of Helsinki. of Helsinki.

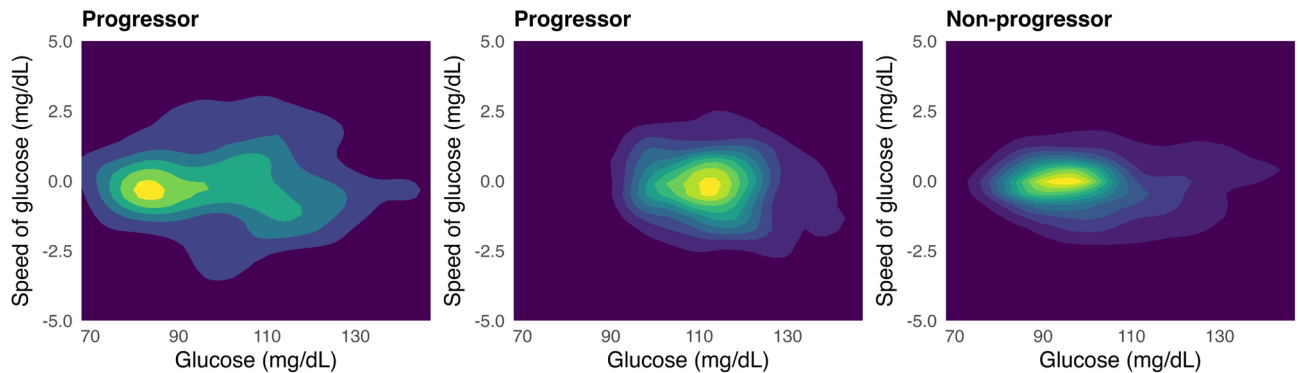
### Descriptive analysis speed of glucose dynamics and patient evolution

To illustrate the clinical relevance of fluctuations on different time scales, we examine the velocity of glucose within the glucodensity framework from a descriptive perspective. Figure 2 presents two-dimensional heat map estimates of glucose and its first derivative (velocity) for three non-diabetic individuals at the start of the study.

In the right panel—corresponding to the individual who remained non-diabetic—the density is tightly concentrated within the normal glucose range, and the velocity exhibits lower variability. In contrast, the two individuals who progressed to diabetes have more dispersed densities and greater velocity fluctuations. These descriptive plots suggest that incorporating velocity information into CGM representations may reveal dynamic patterns indicative of diabetes risk and progression. In Supplemental Material B, Figure 3, we show, for the same three individuals, a plot of the classical CGM-based dynamic markers/indices (see<sup>39</sup> for further details).

### Statistical association results for validation of multivariate glucodensity in AEGIS

The primary objective of this Results section is to provide evidence of the predictive superiority of the CGM distributional metrics introduced in section “Distributional representations for CGM-data analysis”, which incorporate glucose speed and acceleration to predict clinical outcomes. To evaluate these metrics, we compare models of varying complexity that include both CGM and non-CGM variables from the AEGIS baseline data, using the regression methodology described in section “Regression modeling”. Given the limited sample size and the presence of missing data in the five- and eight-year outcomes analyzed, which further reduces the final sample size, we focus on introducing glucose speed and acceleration in a marginal, yet additive manner within the regression models. For each clinical outcome, we fit the model independently using the individual data available, as the number of missing responses varies between the variables examined. The baseline data contain no missing values and the observed missing pattern appears to be completely missing at random.



**Fig. 2.** Heatmap of the two-dimensional density for glucose concentration and its first derivative (speed) estimated from the CGM time series for three non-diabetic individuals at baseline of the AEGIS study. After 8 years, the individuals in the left and middle panels developed diabetes, while the individual in the right panel did not.

We evaluated the predictive performance of these models using adjusted  $R^2$ , which allows us to fairly compare the models while adjusting for complexity and avoiding overfitting. The CGM distributional metrics, including glucose speed and acceleration, are compared with traditional CGM and non-CGM biomarkers.

Table 1 displays the baseline patient characteristics of the AEGIS population sample by sex, indicating that the sample is mainly composed of non-diabetic individuals. In cases where patients have diabetes, they are very well controlled, with glucose values very close to the prediabetes range. For the characterization of baseline data, focus on monitoring the risk and progression of diabetes in a primarily healthy cohort. Among individuals with diabetes, the progression of the disease remains minimal.

We consider four clinical outcomes: (i) HbA1c at 5 years, (ii) HbA1c at 8 years, (iii) FPG at 5 years, and (iv) FPG at 8 years. Six models based on baseline data of increasing complexity are evaluated:

- **Model (1):** Includes age, FPG, and HbA1c.
- **Model (2):** Adds traditional CGM metrics, such as MAGE, CONGA, and Hyper, to age, FPG, and HbA1c.
- **Model (3):** Adds CGM-based dynamic markers—entropy rate (ER) and Poincaré ellipse area (S)—to age, FPG, and HbA1c.
- **Model (4):** Incorporates age, FPG, HbA1c, and the glucodensity quantile marginal distributional representation.
- **Model (5):** Extends Model (4) by including quantiles of glucose speed.
- **Model (6):** Extends Model (5) by adding quantiles of glucose acceleration.

To predict the clinical outcomes, denoted as  $R^j$  where  $j \in \{\text{HbA1c-5}, \text{HbA1c-8}, \text{FPG-5}, \text{FPG-8}\}$ , we employ semi-parametric additive generalized models, as described in section “[Scalar on distribution regression](#)”. The univariate quantile functions for glucose, speed, and acceleration are estimated on a grid of 100 points, following the smoothing methodology from section “[Multivariate glucodensity approach](#)”. Each individual distributional-quantile representation  $\mathcal{Q}_{fij}$ ,  $j \in \{1, 2, 3\}$ , is modeled using a bivariate spline approach outlined in McLean et al. (2014)<sup>41</sup>. Continuous scalar predictors are incorporated into the regression model as linear terms.

Table 2 presents the predictive performance of the five models across the four clinical outcomes, using adjusted  $R^2$ . This metric accounts for the number of parameters, allowing for a balanced comparison across models with different complexities.

Model (6), which incorporates glucose, glucose speed, and acceleration, consistently outperforms the other models, demonstrating significant information gains compared to Models (1) and (2). For example, Model (6) shows a 23% increase in adjusted  $R^2$  for HbA1c at 5 years compared to Model (2), and an 8.7% increase compared to Model (5). For HbA1c at 8 years, Model (6) achieves a 17.4% improvement over Model (2) and an 8.3% increase compared to Model (5). The gains are even more pronounced for fasting plasma glucose (FPG) predictions: at 5 years, Model (6) shows a 40.9% improvement over Model (2) and a 19.1% increase over Model (5). Similarly, for FPG at 8 years, Model (5) achieves a 21.4% improvement over Model (2) and a 21.7% increase relative to Model (5). Models (1) and (2) perform similarly, with Model (2) including traditional CGM metrics. The modest information gain observed for Model (2) is plausibly explained by the fact that the conventional CGM summary statistics used in this specification primarily capture short-term glycaemic variability, whereas the target outcomes HbA1c and FPG—reflect longer-term average glycaemia. In addition, FPG is intrinsically noisier than HbA1c: it is derived from a single fasting measurement that is susceptible to day-to-day fluctuations, while HbA1c reflects average circulating glucose levels over approximately 120 days, making it a much more stable biomarker<sup>52</sup>. These classical models may show greater differentiation with larger sample sizes or additional data. Glucodensity-based models achieved higher adjusted  $R^2$  in predicting long-term HbA1c and FPG levels compared to Model (3), which uses CGM-based dynamic markers such as entropy rate and Poincaré ellipse area. Typically, Model (3) predicts slightly better than Model (2) and Model (1). Importantly, the marginal

	Model (1)	Model (2)	Model (3)	Model (4)	Model (5)	Model (6)
HbA1c-5 years ( $n = 209$ )	0.640	0.645	0.650	0.732	0.743	0.796
HbA1c-8 years ( $n = 250$ )	0.584	0.579	0.589	0.628	0.656	0.680
FPG-5 years ( $n = 408$ )	0.434	0.438	0.474	0.518	0.587	0.617
FPG-8 years ( $n = 223$ )	0.339	0.426	0.385	0.425	0.445	0.517

**Table 2.** Adjusted  $R^2$  summary for clinical outcomes across six predictive models of varying complexity. Model (1) includes baseline age, FPG, and HbA1c. Model (2) adds CGM metrics such as Area Under the Curve (AUC), Mean Amplitude of Glycemic Excursions (MAGE), Continuous Overall Net Glycemic Action (CONGA), and hyperglycemia time-in-range. Model (3) adds CGM-based dynamic markers—entropy rate (ER) and Poincaré ellipse area (S)—to age, FPG, and HbA1c. Model (4) adds the quantile function for the glucose profile to age, FPG, and HbA1c. Model (5) adds quantile functions for glucose speed to Model (4), and Model (6) adds quantile functions for glucose acceleration to Model (5).

glucodensity, as well as the speed and acceleration functional variables, show statistical significance in the global  $F$ -test, with  $p$  values below 0.02 for the four results in models (4), (5) and (6), respectively.

Additional results, including linear term coefficients and other performance metrics such as log-likelihood and UBRE, are available in the Supplemental Material (see Supplemental Material A, Tables 3–6).

## Discussion

In this paper, we explore the clinical value of glucose speed and acceleration for long-term glucose prediction using novel functional CGM metrics. Our focus is on two primary clinical outcomes: HbA1c and FPG—key biomarkers for diagnosing diabetes and monitoring disease progression. We utilize baseline data from the AEGIS study as predictors, with outcomes collected five and eight years from baseline, employing a novel functional distributional regression models.

The results demonstrate that incorporating functional glucose fluctuations through marginal distributional representations of speed and acceleration increases the statistical association—in terms of adjusted  $R^2$ —by over 20% compared to using classical CGM and non-CGM metrics alone. This significant improvement underscores the additional time-dependent information captured by these novel CGM metrics. It suggests that CGM can predict diabetes risk in healthy populations and monitor disease progression in diabetic individuals with higher precision than existing biomarkers.

From a practical standpoint, the findings indicate that not just the intensity of glucose, but also the direction (speed) and magnitude (acceleration) of glucose concentration changes are crucial. These metrics, derived from CGM time series, are closely linked to glucose variability<sup>53</sup>—a key factor in dysglycemia and a driver of cellular damage<sup>54</sup>. Our findings support the hypothesis that not only hyperglycemic states, but also speed and acceleration at which glucose levels fluctuate, may predict the deterioration of functional  $\beta$ -cell mass in Type 2 diabetes. In manifest diabetes, pancreatic  $\beta$ -cells produce insufficient amounts of insulin to compensate for insulin resistance, resulting in a relative insulin deficiency and leading to hyperglycemia<sup>55</sup>.

While glucose speed and acceleration can be intuitively derived from a time-series of glucose measurements, the physiological traits that govern these dynamics are complex. Glucose speed could closely reflect physiological processes such as glucose absorption and disposal that are in turn dependent on insulin secretion and insulin sensitivity. Glucose acceleration could be related to physiologic modulators of these processes such as incretin secretion and action after meals or the effect of glucagon enhancing endogenous glucose production to decelerate falling glucose after meals.

Our analysis offers a novel perspective on glucose variability by incorporating the glucodensity approach, which includes both marginal and multivariate functional metrics. These advanced metrics provide a deeper understanding of CGM time series beyond traditional measures such as Continuous Overall Net Glycemic Action (CONGA) and Mean Amplitude of Glycemic Excursions (MAGE), which primarily focus on glucose variability without capturing the full dynamics<sup>23,52</sup>. We anticipate that future clinical trials will adopt glucose variability as a key outcome measure, moving beyond reliance on average glucose values alone. Our novel CGM metrics can serve as a foundation for developing new quantitative methods in this field.

The traditional marginal glucodensity approach in diabetes research has the clear advantage of simultaneously capturing low, high, and mid-range glucose concentrations. Furthermore, regression models based on these outcomes are highly interpretable. Our previous work on modeling CGM data in clinical trials<sup>7</sup> using the glucodensity approach and multilevel models emphasizes the clinical relevance of this methodology for better understanding of glucose profile evolution during interventions. This provides a more nuanced and sophisticated perspective. In this context, an observational study using glucodensity further highlights the modeling advantages of this approach compared to traditional CGM metrics when comparing next-generation insulins<sup>4</sup>.

In another recent study, we demonstrated that the glucodensity approach could predict time-to-diabetes<sup>56</sup> (survival model) more accurately than traditional biomarkers, achieving over 13% improvement in the area under the curve (AUC). Here, we focus on regression modeling because it is a more appropriate statistical method for detecting stronger statistical associations with functional models, especially given the sample

sizes we are considering. For this reason, this paper presents multivariate glucodensity analyses primarily for descriptive purposes.

With larger sample sizes, the multivariate glucodensity model, which simultaneously analyzes glucose speed and acceleration, addresses the limitations of marginal models by offering a more comprehensive functional representation. This model has the potential to be more powerful and more interpretable, as evidenced by the analyses we performed. With a massively increasing use of continuous glucose monitoring in early prevention our approach has the potential to meaningfully extend the predictive capabilities of this diagnostic method. Incorporating glucose acceleration may facilitate early detection of  $\beta$ -cell dysfunction, support digital phenotyping and risk stratification, and inform novel intervention strategies to improve metabolic capacity of individuals.

A key next step is to replicate our clinical validation in larger and more diverse cohorts of individuals with diabetes, including pediatric populations. In these settings, using marginal glucodensity as the clinic outcome — instead of relying solely on mean glucose metrics — will enable a more nuanced assessment of hypoglycemic and hyperglycemic excursions driven by glucose speed and acceleration. Validation in large cohorts of healthy individuals is also important; however, ongoing population studies such as *All of Us Research Program* and *Human Phenotype Project* currently provide less than four years of follow-up, limiting their usefulness for long-term glycemic evaluation.

We emphasise that the present study investigates *statistical* associations, not causal relationships. Although the strong effects identified in our regression models may hint at plausible causal mechanisms, confirming such pathways would require dedicated causal-inference analyses beyond the scope of this work.

In conclusion, this work highlights glucose speed and acceleration as novel biomarkers to predict long-term glucose levels and diabetes outcomes. This vision aligns with Montaser et al.<sup>39</sup>, who showed that dynamic CGM markers detect early dysglycemia and differentiate diabetes phenotypes. Our findings also emphasize the need for sophisticated analytical tools to leverage CGM data effectively<sup>57</sup>. For example, the use of functional data methods can be instrumental in order to analyze continuous digital health data including CGM data<sup>19</sup>.

## Data availability

The datasets used and/or analysed during the current study available from the corresponding author on reasonable request.

Received: 22 March 2025; Accepted: 29 August 2025

Published online: 29 September 2025

## References

- Li, X. et al. Digital health: tracking physiomes and activity using wearable biosensors reveals useful health-related information. *PLoS Biol.* **15**(1), 1–30 (2017).
- Steinhubl, S. R. & Topol, E. J. Digital medicine, on its way to being just plain medicine (2018).
- Cryer, P. E., Davis, S. N. & Shamooin, H. Hypoglycemia in diabetes. *Diabetes Care* **26**(6), 1902–1912 (2003).
- Gomez-Peralta, F. et al. Insulin glargine 300 u/ml versus insulin degludec 100 u/ml improves nocturnal glycaemic control and variability in type 1 diabetes under routine clinical practice: A glucodensities-based post hoc analysis of the onecare study. *Diabetes Obes. Metab.* **26**(5), 1993–1997 (2024).
- Battelino, T. et al. Continuous glucose monitoring and metrics for clinical trials: an international consensus statement. *Lancet Diabetes Endocrinol.* **11**(1), 42–57 (2023).
- Zeevi, D. et al. Personalized nutrition by prediction of glycemic responses. *Cell* **163**(5), 1079–1094 (2015).
- Matabuena, M. & Crainiceanu, C. M. Multilevel functional distributional models with application to continuous glucose monitoring in diabetes clinical trials. arXiv preprint [arXiv:2403.10514](https://arxiv.org/abs/2403.10514) (2024).
- Oganesova, Z., Pemberton, J. & Brown, A. Innovative solution or cause for concern? the use of continuous glucose monitors in people not living with diabetes: a narrative review. *Diabet. Med.* **41**(9), e15369 (2024).
- Hall, H. et al. Glucotypes reveal new patterns of glucose dysregulation. *PLoS Biol.* **16**(7), 1–23 (2018).
- Shilo, S. et al. Continuous glucose monitoring and intrapersonal variability in fasting glucose. *Nat. Med.* **30**(5), 1424–1431 (2024).
- Bergental, R. M. et al. Recommendations for standardizing glucose reporting and analysis to optimize clinical decision making in diabetes: The ambulatory glucose profile (agp). *Diabetes Technol. Therapeut.* **15**(3), 198–211 (2013) (PMID: 23448694).
- Nguyen, M. et al. A review of continuous glucose monitoring-based composite metrics for glycemic control. *Diabetes Technol. Therapeut.* (2025).
- Dumuid, D. et al. Compositional data analysis in time-use epidemiology: what, why, how. *Int. J. Environ. Res. Public Health* **17**(7), 2220 (2020).
- Biagi, L. et al. Individual categorisation of glucose profiles using compositional data analysis. *Stat. Methods Med. Res.* **28**(12), 3550–3567 (2019).
- Beck, R. W., Connor, C. G., Mullen, D. M., Wesley, D. M. & Bergenstal, R. M. The fallacy of average: How using hba1c alone to assess glycemic control can be misleading. *Diabetes Care* **40**(8), 994–999 (2017).
- Beyond A1C Writing Group. Need for regulatory change to incorporate beyond a1c glycemic metrics. *Diabetes Care* **41**(6), e92–e94 (2018).
- Nathan, D. M., Turgeon, H. & Regan, S. Relationship between glycosylated haemoglobin levels and mean glucose levels over time. *Diabetologia* **50**(11), 2239–2244 (2007).
- Beck, R. W. et al. Validation of time in range as an outcome measure for diabetes clinical trials. *Diabetes Care* (2018).
- Crainiceanu, C. M., Goldsmith, J., Leroux, A. & Cui, E. *Functional data analysis with R* (CRC Press, 2024).
- Matabuena, M., Petersen, A., Vidal, J. C. & Gude, F. Glucodensities: a new representation of glucose profiles using distributional data analysis. *Stat. Methods Med. Res.* **30**(6), 1445–1464 (2021).
- Matabuena, M. & Petersen, A. Distributional data analysis of accelerometer data from the nhanes database using nonparametric survey regression models. *J. Roy. Stat. Soc.: Ser. C (Appl. Stat.)* **72**(2), 294–313 (2023).
- Ghosal, R. et al. Distributional data analysis via quantile functions and its application to modeling digital biomarkers of gait in alzheimer's disease. *Biostatistics* **24**(3), 539–561 (2023).
- Monnier, L., Colette, C. & Owens, D. R. Glycemic variability: the third component of the dysglycemia in diabetes. is it important? how to measure it? *J. Diabetes Sci. Technol.* **2**(6), 1094–1100 (2008).

24. Kovatchev, B. P., Otto, E., Cox, D., Gonder-Frederick, L. & Clarke, W. Evaluation of a new measure of blood glucose variability in diabetes. *Diabetes Care* **29**(11), 2433–2438 (2006).
25. Boyne, M. S., Silver, D. M., Kaplan, J. & Saudek, C. D. Timing of changes in interstitial and venous blood glucose measured with a continuous subcutaneous glucose sensor. *Diabetes* **52**(11), 2790–2794 (2003).
26. Lutsker, G. et al. From glucose patterns to health outcomes: a generalizable foundation model for continuous glucose monitor data analysis. arXiv preprint [arXiv:2408.11876](https://arxiv.org/abs/2408.11876) (2024).
27. McDonnell, C. M., Donath, S. M., Vidmar, S. I., Werther, G. A. & Cameron, F. J. A novel approach to continuous glucose analysis utilizing glycemic variation. *Diabetes Technol. Therapeut.* **7**(2), 253–263 (2005) (PMID: 15857227).
28. Kovatchev, B. & Cobelli, C. Glucose variability: timing, risk analysis, and relationship to hypoglycemia in diabetes. *Diabetes Care* **39**(4), 502–510 (2016).
29. Zaccardi, F. & Khunti, K. Glucose dysregulation phenotypes - time to improve outcomes. *Nat. Rev. Endocrinol.* **14**(11), 632–633 (2018).
30. Kovatchev, B. P., Clarke, W. L., Breton, M., Brayman, K. & McCall, A. Quantifying temporal glucose variability in diabetes via continuous glucose monitoring: mathematical methods and clinical application. *Diabetes Technol. Therapeut.* **7**(6), 849–862 (2005).
31. Kovatchev, B. P., Shields, D. & Breton, M. Graphical and numerical evaluation of continuous glucose sensing time lag. *Diabetes Technol. Therapeut.* **11**(3), 139–143 (2009).
32. Rodbard, D. New and improved methods to characterize glycemic variability using continuous glucose monitoring. *Diabetes Technol. Therapeut.* **11**(9), 551–565 (2009).
33. Beck, R. W. Is it time to replace time-in-range with time-in-tight-range? maybe not (2024).
34. Yongjin, X. et al. Time in range, time in tight range, and average glucose relationships are modulated by glycemic variability: Identification of a glucose distribution model connecting glycemic parameters using real-world data. *Diabetes Technol. Therapeut.* **26**(7), 467–477 (2024) (PMID: 38315505).
35. Montaser, E., Fabris, C. & Kovatchev, B. Essential continuous glucose monitoring metrics: the principal dimensions of glycemic control in diabetes. *Diabetes Technol. Therapeut.* **24**(11), 797–804 (2022) (PMID: 35714355).
36. Matabuena, M., Félix, P., Ditzhaus, M., Vidal, J. & Gude, F. Hypothesis testing for matched pairs with missing data by maximum mean discrepancy: an application to continuous glucose monitoring. *Am. Stat.* **77**(4), 357–369 (2023).
37. Godsland, I. F., Jeffs, J. A. R. & Johnston, D. G. Loss of beta cell function as fasting glucose increases in the non-diabetic range. *Diabetologia* **47**, 1157–1166 (2004).
38. Drucker, D. J. The biology of incretin hormones. *Cell Metab.* **3**(3), 153–165 (2006).
39. Montaser, E., Farhy, L. S. & Kovatchev, B. P. Novel detection and progression markers for diabetes based on continuous glucose monitoring data dynamics. *J. Clin. Endocrinol. Metabol* **110**(1), 254–262 (2024).
40. Chacón, J. E. & Duong, T. *Multivariate Kernel Smoothing and its Applications* (Chapman and Hall/CRC, 2018).
41. McLean, M. W., Hooker, G., Staicu, A.-M., Scheipl, F. & Ruppert, D. Functional generalized additive models. *J. Comput. Graph. Stat.* **23**(1), 249–269 (2014).
42. Happ, C., Greven, S. & Schmid, V. J. The impact of model assumptions in scalar-on-image regression. *Stat. Med.* **37**, 4298–4317 (2018).
43. Marx, B. D. & Eilers, P. H. C. Multidimensional penalized signal regression. *Technometrics* **47**(1), 13–22 (2005).
44. Wood, S. Package 'mgcv'. *R package version 1* (29), 729 (2015).
45. Ruppert, D., Wand, M. P. & Carroll, R. J. *Semiparametric Regression. Number 12* (Cambridge University Press, 2003).
46. Kim, J., Maity, A. & Staicu, A.-M. Additive nonlinear functional concurrent model. *Stat. Interface* **11**, 669–685 (2018).
47. Wood, S. N. *Generalized Additive Models: An Introduction with R* (Chapman and hall/CRC, 2017).
48. Gude, F. et al. Glycemic variability and its association with demographics and lifestyles in a general adult population. *J. Diabetes Sci. Technol.* **11**(4), 780–790 (2017).
49. Hoelzel, W. et al. Ifcc reference system for measurement of hemoglobin a1c in human blood and the national standardization schemes in the united states, japan, and sweden: A method-comparison study. *Clin. Chem.* **50**(1), 166–174 (2004).
50. FJ Service et al. Mean amplitude of glycemic excursions, a measure of diabetic instability. *Diabetes* **19**(9), 644–655 (1970).
51. Molnar, G. D., Taylor, W. F. & Ho, M. M. Day-to-day variation of continuously monitored glycaemia: a further measure of diabetic instability. *Diabetologia* **8**(5), 342–348 (1972).
52. Selvin, E., Crainiceanu, C. M., Brancati, F. L. & Coresh, J. Short-term variability in measures of glycemia and implications for the classification of diabetes. *Arch. Intern. Med.* **167**(14), 1545–1551 (2007).
53. Kovatchev, B. & Cobelli, C. Glucose variability: timing, risk analysis, and relationship to hypoglycemia in diabetes. *Diabetes Care* **39**(4), 502–515 (2016).
54. Rolo, A. P. & Palmeira, C. M. Diabetes and mitochondrial function: role of hyperglycemia and oxidative stress. *Toxicol. Appl. Pharmacol.* **212**(2), 167–178 (2006).
55. Škrha, J., Šoupal, J. & Prázný, M. Glucose variability, hba1c and microvascular complications. *Rev. Endocr. Metab. Disord.* **17**(1), 103–110 (2016).
56. Matabuena, M., Díaz-Louza, C., Ghosal, R., & Gude, F. Personalized imputation in metric spaces via conformal prediction: Applications in predicting diabetes development with continuous glucose monitoring information. arXiv preprint [arXiv:2403.18069](https://arxiv.org/abs/2403.18069) (2024).
57. Klonoff, D. C. et al. Cgm data analysis 2.0: functional data pattern recognition and artificial intelligence applications. arXiv preprint [arXiv:2505.07885](https://arxiv.org/abs/2505.07885) (2025).

## Author contributions

MM: Designed research Performed research, Contributed new reagents/analytic tools, Analyzed data, Wrote the paper. RG: Contributed new reagents/analytic tools, Wrote the paper. JA: Contributed new reagents/analytic tools, Wrote the paper. AK: Contributed new reagents/analytic tools, Wrote the paper. RW: Contributed new reagents/analytic tools, Wrote the paper. CM: Contributed new reagents/analytic tools, Wrote the paper. JC: Contributed new reagents/analytic tools, Wrote the paper. VZ: Contributed new reagents/analytic tools. Wrote the paper. JO: Contributed new reagents/analytic tools. Wrote the paper. FG: Data management. Contributed new reagents/analytic tools, Wrote the paper.

## Funding

This study has been funded by Instituto de Salud Carlos III (ISCIII) through the projects PI11/02219, PI16/01395, PI20/01069 and by the European Union (Programme for Research and Innovation, 2021–2027; Horizon Europe, EIC Pathfinder: 101161509-GLUCOTYPES); the Network for Research on Chronicity, Primary Care, and Health Promotion, ISCIII, RD24/0005/0010, co-funded by the European Union (UE); and the Galician Innovation Agency-Competitive Benchmark Groups (GAIN-GRC/IN607A2025-09/Xunta de Galicia). The A Estrada

Glycation and Inflammation Study (AEGIS) group acknowledges the efforts of the participants and thanks them for participating.

### Competing interests

The authors declare no competing interests.

### Additional information

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1038/s41598-025-18119-2>.

**Correspondence** and requests for materials should be addressed to M.M.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025