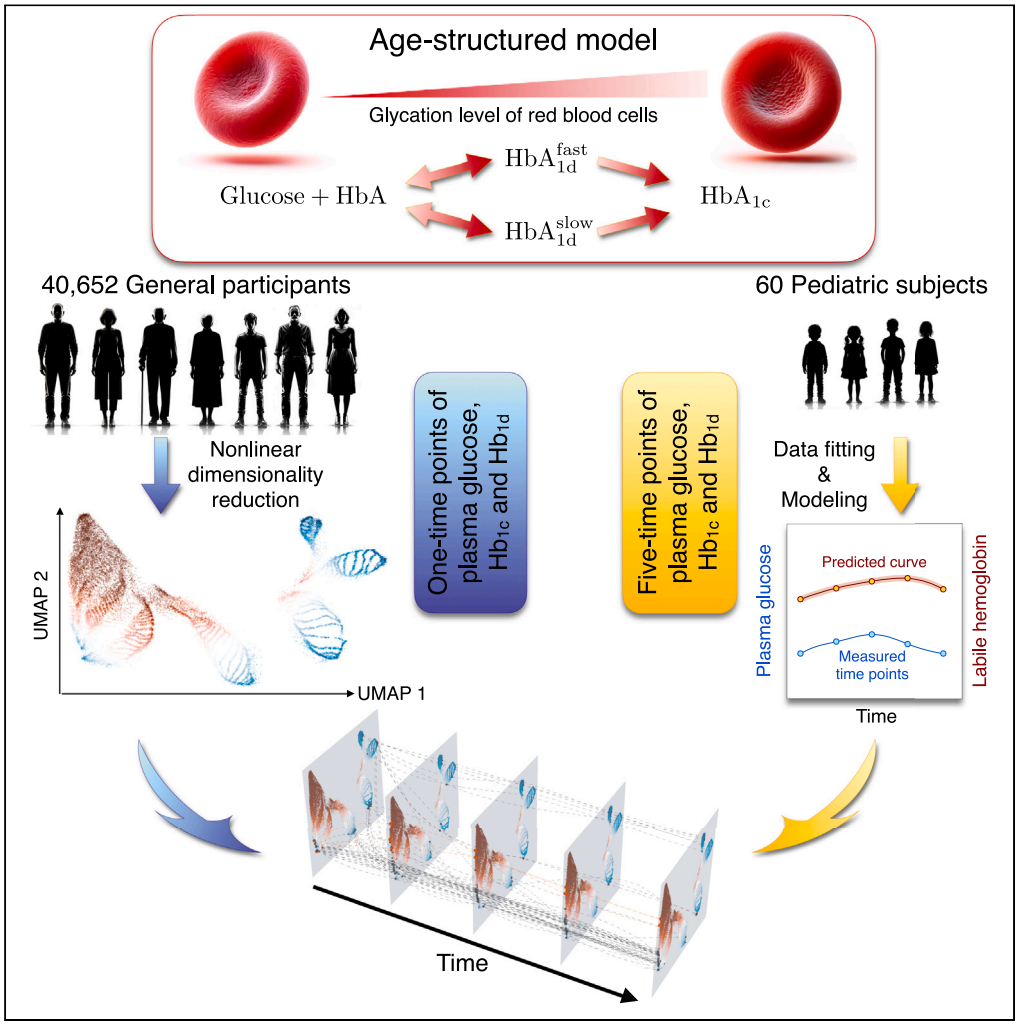
iScience

CellPress
OPEN ACCESS

Article

Integrated modeling of labile and glycated hemoglobin with glucose for enhanced diabetes detection and short-term monitoring



José Antonio Romero-Rosales, David G. Aragones, José Escribano-Serrano, ..., Juan Belmonte-Beitia, María Rosa Durán, Gabriel F. Calvo

gabriel.fernandez@uclm.es

Highlights

Retrospective and prospective study of diabetic biomarkers in over 40,000 individuals

Dimensionality reduction of the large individual dataset reveals diabetic subclusters

Labile hemoglobin identifies diabetic patients undetected by glycated hemoglobin alone

A mathematical model captures the dynamics of glucose, labile, and glycated hemoglobins

Romero-Rosales et al., iScience 27, 109369
April 19, 2024
https://doi.org/10.1016/j.isci.2024.109369

iScience

CellPress
OPEN ACCESS

Article

Integrated modeling of labile and glycated hemoglobin with glucose for enhanced diabetes detection and short-term monitoring

José Antonio Romero-Rosales,[1,9] David G. Aragones,[1,9] José Escribano-Serrano,[2] Marisa González Borrachero,[3] Alfredo Michán Doña,[4,5] Francisco J. Macías López,[6] María Angeles Santos Mata,[6] Inmaculada Naranjo Jiménez,[6] María Jesús Casamitjana Zamora,[6] Hélia Serrano,[7] Juan Belmonte-Beitia,[1] María Rosa Durán,[5,8] and Gabriel F. Calvo[1,10,*]

SUMMARY

Metabolic biomarkers, particularly glycated hemoglobin and fasting plasma glucose, are pivotal in the diagnosis and control of diabetes mellitus. Despite their importance, they exhibit limitations in assessing short-term glucose variations. In this study, we propose labile hemoglobin as an additional biomarker, providing insightful perspectives into these fluctuations. By utilizing datasets from 40,652 retrospective general participants and conducting glucose tolerance tests on 60 prospective pediatric subjects, we explored the relationship between plasma glucose and labile hemoglobin. A mathematical model was developed to encapsulate short-term glucose kinetics in the pediatric group. Applying dimensionality reduction techniques, we successfully identified participant subclusters, facilitating the differentiation between diabetic and non-diabetic individuals. Intriguingly, by integrating labile hemoglobin measurements with plasma glucose values, we were able to predict the likelihood of diabetes in pediatric subjects, underscoring the potential of labile hemoglobin as a significant glycemic biomarker for diabetes research.

INTRODUCTION

Diabetes mellitus encompasses a group of metabolic disorders characterized by high levels of glucose in the blood. In 2021, it was estimated that 537 million adults (aged 20–79) worldwide were living with diabetes, with a projected increase in prevalence of more than 40% over the next two decades.[1] Monitoring glucose levels is crucial for the effective management of diabetes mellitus. In recent years, continuous glucose monitoring (CGM) technologies have provided wearable devices that can record real-time variations in glucose levels. These devices are particularly adept at detecting episodes of hypo- and hyperglycemia under both resting and exercise conditions,[2,3] especially in type 1 diabetes patients.[4] However, the adoption and utility of CGMs in the management of type 2 diabetes is less firmly established, with limited real-world data supporting their use.[5] While CGMs provide invaluable insights for type 2 diabetes management, the high prevalence of this condition makes their widespread employment both logistically and economically challenging. Moreover, most patients can attain satisfactory glycemic control through less technologically intensive means. Alongside the challenges in standardizing CGM readouts, a significant constraint is the associated cost and accessibility of CGM devices for populations heavily impacted by social determinants of health. Many of these more vulnerable groups are located in African, Eastern Mediterranean, and Southeast Asian countries, which have also been found to present the highest age-standardized death rates related to type 2 diabetes mellitus in 2019.[6]

Glycated hemoglobin (HbA_{1c}) is a well-established clinical indicator for the diagnosis of diabetes mellitus.[7] Quantifying the kinetics of HbA_{1c} has been instrumental in revealing its use as a glycemic biomarker, linking glycemia to the emergence of diabetic complications.[8–11] However, several limitations have been noted.[12] Most notably, HbA_{1c} does not correlate well with short-time (i.e., in the range of minutes to a few hours) fluctuations in glucose levels, typically reflecting average glucose alterations over a span of several weeks up to three months. Consequently, it cannot reliably detect episodes of hypo- and hyperglycemia and is susceptible to biological variance.[13,14] Furthermore,

[1]Department of Mathematics, Mathematical Oncology Laboratory (MOLAB), University of Castilla-La Mancha, Ciudad Real, Spain
[2]UGC Primary Care San Roque, Campo de Gibraltar, Cádiz, Spain
[3]Department of Biochemistry, University Hospital of Jerez, Jerez, Cádiz, Spain
[4]UGC Internal Medicine, University Hospital of Jerez and Department of Medicine, University of Cádiz, Cádiz, Spain
[5]Biomedical Research and Innovation Institute of Cadiz (INiBICA), Hospital Universitario Puerta del Mar, Cádiz, Spain
[6]UGC Pediatric, University Hospital of Jerez, Jerez, Cádiz, Spain
[7]Department of Mathematics, Faculty of Chemical Sciences and Technologies, University of Castilla-La Mancha, Ciudad Real, Spain
[8]Department of Mathematics, University of Cádiz, Puerto Real, Cádiz, Spain
[9]These authors contributed equally
[10]Lead contact
*Correspondence: gabriel.fernandez@uclm.es
https://doi.org/10.1016/j.isci.2024.109369

HbA_{1c} is not suitable for use with hospitalized patients suffering from cirrhosis due to their high incidence of anemia, a significant concern given the strong correlation between cirrhosis and diabetes.[15] It may also exhibit analytical problems when dealing with other pathologies, such as elevated triglycerides and hyperbilirubinemia,[16] or yield levels that fail to capture cases of extreme hyperglycemia.[17]

In response to the limitations of HbA_{1c}, labile hemoglobin, HbA_{1d}, has recently emerged as a promising candidate for enhancing the precision and applicability of glycemic control diagnostics.[18] HbA_{1d}, which forms as an intermediate product in the glycation process of hemoglobin, has thus far been primarily regarded as a perturbation parameter for HbA_{1c}.[19] However, accumulating evidence suggests its potential for broader applications. For instance, HbA_{1d} could serve as an indirect measure of the glycation index, adding depth to the current understanding of individual patient glycemic profiles. A salient aspect of HbA_{1d} is its capability to monitor glycemia over shorter periods compared to HbA_{1c},[20] hence being capable of capturing fluctuations that HbA_{1c} may overlook. This could enable more timely adjustments in treatment regimens and improve the prospects for favorable patient outcomes.

In a previous study,[21] a mathematical model that incorporated the interaction of plasma glucose levels with HbA_{1c} and HbA_{1d}, as well as the impact of natural erythrocyte death, was proposed. This model presented a unified view of the temporal alterations in these two hemoglobin forms, suggesting intricate and interconnected biochemical dynamics between HbA_{1c} and HbA_{1d}. Along with other *in silico* frameworks, this methodology serves to complement experimental biology and clinical studies of diabetes, a notion further elaborated in.[22] Such quantitative tools enrich our comprehension of the patient-specific evolution of the glycation process. They facilitate the identification of key influencing elements and elucidate how reactant agents impact this process. The synergy between computational models and experimental or clinical studies underscores the integral role of an integrated approach to enhancing our understanding of diabetes.

In this study, we initially analyzed a large cohort of 40,652 retrospective participants, utilizing single time point measurements for fasting plasma glucose (FPG), HbA_{1d}, and HbA_{1c}. We applied nonlinear dimensionality reduction (NDR) techniques to project the raw and unprocessed dataset into a lower dimensional space, thereby preserving critical structural relationships and facilitating intuitive visualization and interpretation. The generated clusters gave rise to an informative global map comprising both diabetic patients and non-diabetic individuals. Subsequently, we assessed the results of glucose tolerance tests performed on 60 pediatric subjects, with FPG, HbA_{1d}, and HbA_{1c} measured at 30-min intervals over 2 h. A mathematical model, based on a set of ordinary differential equations, was developed to capture the kinetics of FPG, HbA_{1d}, and HbA_{1c} within the time frame of the pediatric subject measurements. By superimposing these dynamics onto the NDR-derived map, we were able to identify pediatric participant subsets that were diabetic. The conceptual strategy presented in this study has shown effectiveness in diabetes research, offering insights that may be relevant to other areas in personalized medicine. While primarily demonstrated in the context of diabetes, our approach in dissecting complex datasets and modeling physiological responses through systems biology frameworks suggests potential for further exploration in conditions like cardiovascular disorders, metabolic syndromes, and cancer.

RESULTS

Nonlinear dimensionality reduction unveils distinctive diabetic subclusters through labile hemoglobin

First, we applied an NDR technique to retrospective data from 40,652 participants using the Euclidean metric (see Figure 1), as described in.[23] Further details about the NDR method employed can be found in the STAR methods. These datasets incorporated the diabetic/non-diabetic status of the participants, along with levels of FPG, HbA_{1c}, and HbA_{1d} (detailed baseline characteristics are displayed in Table 1). Using NDR helped us uncover possible correlations which were not as apparent in the original direct 3D spatial plot (see Figure S1). Figure 1A delineates two distinct large clusters: the left one primarily comprising diabetic patients (indicated by orange dots), and the right one predominantly non-diabetic participants (highlighted by blue dots). The impact of these variables on the NDR is depicted through color scales. Figures 1B and 1C represent the levels of FPG and HbA_{1c}, respectively, for each participant. A strong correlation between these two indicators can be observed within the left large cluster. This correlation, however, diminishes significantly in the lower right subcluster, indicating a subset of diabetic patients with disparate FPG and HbA_{1c} values. In contrast, Figure 1D, corresponding to HbA_{1d}, exhibits a clear progression from low to high fractions, transitioning from the right large cluster to the left one (and also within each subcluster). Notably, HbA_{1d} seems effective in identifying subsets of individuals in a prediabetic state, especially within the lower subcluster of the left large cluster. The predictive capability of HbA_{1d} for identifying prediabetic states stems from the substantial correlation existing between diagnostic values (Figure 1A) and HbA_{1d} levels (Figure 1D). This correlation constituted initial compelling evidence, highlighting a significant association between diagnostic status and HbA_{1d} levels.

Labile hemoglobin identifies diabetic patients undetected by HbA_{1c} alone

The NDR analysis of the dataset encompassing 40,652 general participants allowed us to confirm that each of the three variables studied (FPG, HbA_{1c}, and HbA_{1d}) held predictive power for diabetes, despite their nonzero correlations. Figure 2 presents logistic regression analyses for all the variables. These statistical models estimate the probability of a subject being classified as diabetic or non-diabetic based on their respective variable values. Figure 2 A differentiates individuals (diabetic/nondiabetic) and potential biomarkers using color codes: (1) A prominent green cluster corresponds to diabetic patients identified by high HbA_{1c} (> 6.5%); (2) a smaller red cluster represents diabetic patients with high HbA_{1d} (> 1.8%); (3) a yellow cluster denotes patients identified by high FPG (> 110 mg/dL); (4) patients who are not identifiable by any of the aforementioned variables are colored in orange; and (5) non-diabetic participants are depicted in blue. Interestingly, HbA_{1d} proves effective in identifying diabetic patients undetectable by HbA_{1c} alone, with only a minimal number of patients remaining undetected by the studied variables. Figure 2B shows the probabilities of being diabetic for each individual, derived from a logistic model integrating all

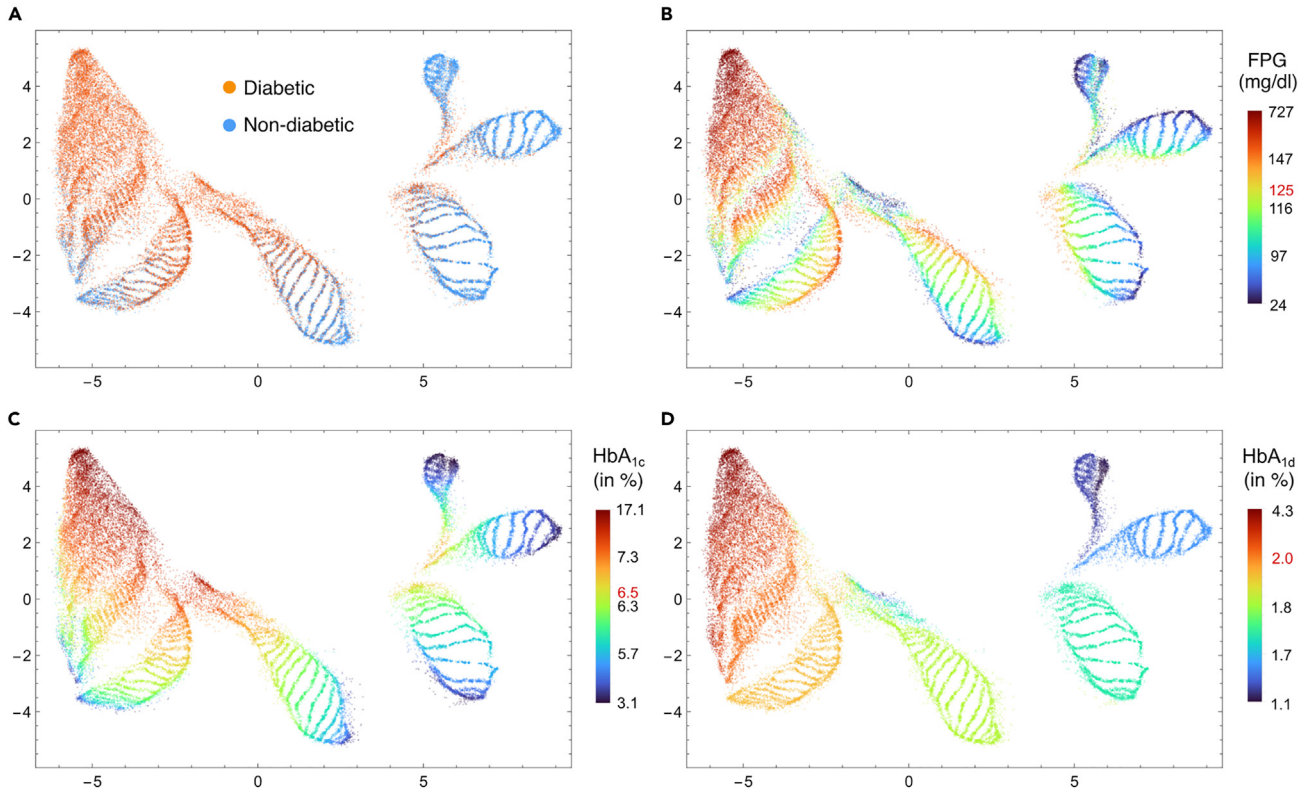


Figure 1. Uniform Manifold Approximation and Projection (UMAP) nonlinear dimensionality reduction (NDR) applied to 40,652 general participants who were simultaneously assayed for plasma glucose, HbA_{1c}, and HbA_{1d} under fasting conditions

(A) Diabetic (23,866) and non-diabetic (16,786) participants are color-coded, revealing their correspondence to two distinct major clusters following NRD.

(B) Fasting plasma glucose (FPG) values.

(C) Glycated hemoglobin (HbA_{1c}) fractions (expressed in %).

(D) Labile hemoglobin (HbA_{1d}) fractions (expressed in %). The color scales are linear on the percentiles of the data. The values highlighted in red represent prediabetic-diabetic threshold references.

three variables (for comprehensive statistical details, refer to Tables S1, S2, and Figure S3 in the supplemental information). A discernible large cluster on the left, represented in red, exhibits high probabilities of diabetes, whereas varying subclusters, from the lower left corner to the right, primarily in blue, account for low probabilities (typically smaller than 0.2 or lower). A white transition region in the center may encompass subjects likely in a prediabetic state. Finally, Figures 2C–2E display the McFadden likelihood ratio index (R^2_{McF}) for each variable, with threshold values shown in red. It is worth noting that R^2_{McF}, a normalized value ranging between [0,1], indicates a superior classification capability over a uniformly random discrete classification model as its value increases.

Red blood cell glycation can be described by a dynamic-based model

Based on the previous work,[21] we put forward a mathematical model (see Box 1) to capture the dynamic interplay between glucose levels in the blood plasma, HbA_{1c}, and HbA_{1d}. Our approach, predicated on the law of mass action, results in a set of linear, time-varying ordinary differential equations, with glucose being a given time-dependent function. These equations account for HbA_{1d} formation and dissociation,

Table 1. Baseline characteristics of the participant cohorts

	Age [a] (years)	Number of participants	HbA_{1d} [b] (%)	HbA_{1c} [b] (%)	FPG [b] (mg/dL)	Sex Male/Female
General	65 (0–105)	40,652	1.89 ± 0.30	6.69 ± 1.43	129.2 ± 49.95	19,102/21,550
Training	65 (0–105)	32,521	1.89 ± 0.30	6.68 ± 1.43	129.0 ± 49.81	15,264/17,257
Validation	65 (0–105)	8,131	1.90 ± 0.30	6.71 ± 1.44	129.7 ± 50.51	3,838/4,293
Pediatric	11 (2–14)	60	1.78 ± 0.24	5.19 ± 0.33	83.25 ± 8.33	26/34

[a]Values expressed as median (range).

[b]Values expressed as mean ± standard deviation.

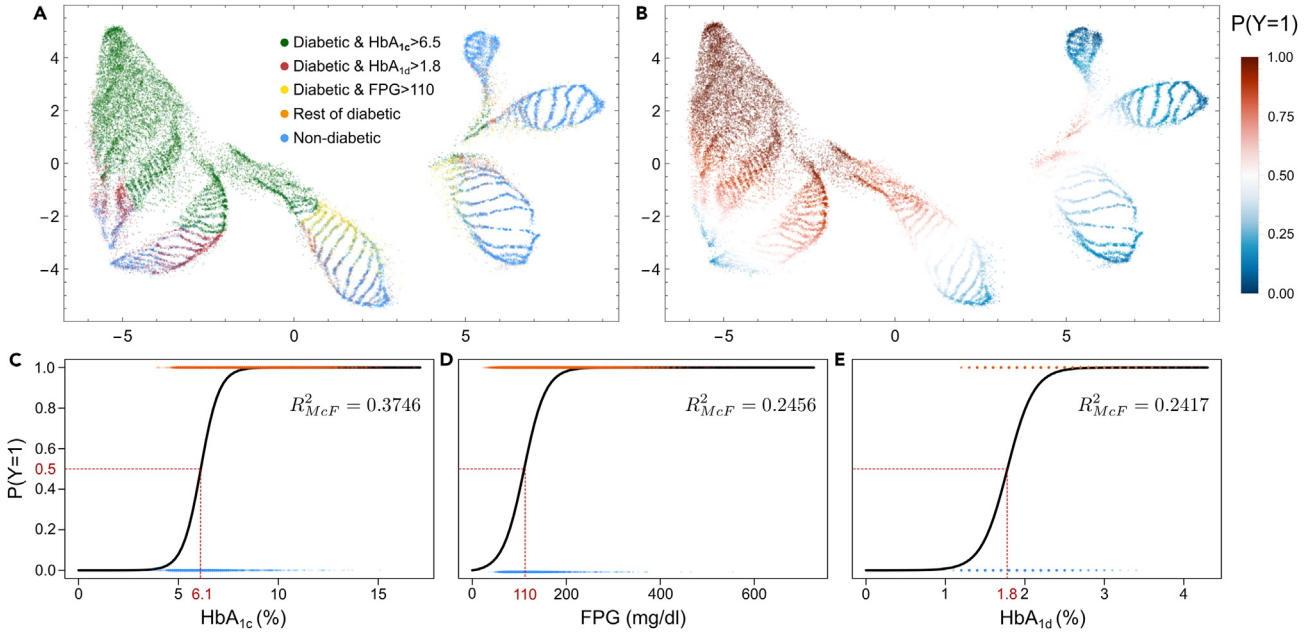


Figure 2. Logistic regression analyses for the three studied glycemic variables using the retrospective general participant dataset

(A) Participant groups classified by diagnosis (diabetic/non-diabetic) as predicted by different thresholds of the variables.

(B) Probabilities of being diabetic shown over the UMAP computed in Figure 1.

(C) Logistic model of HbA_{1c}.

(D) Logistic model of FPG.

(E) Logistic model of HbA_{1d}. In subfigures (C), (D) and (E), the McFadden likelihood ratio index (R^2_{McF}) is also shown for each of the variables, with threshold values highlighted in red corresponding to a probability of 0.5. As in Figure 1, orange and blue dots represent diabetic and non-diabetic participants, respectively.

caused by interactions between free hemoglobin and glucose, and glycation through Amadori rearrangement and cell death, reflecting the non-enzymatic nature of glycation. The uptake of glucose by RBCs occurs via SLC2 membrane transporters, facilitating rapid glycation and the formation of unstable HbA_{1d}, which later transforms into stable HbA_{1c} through an Amadori rearrangement.[25] This stable molecule persists until RBC apoptosis, with a lifespan averaging 100–120 days (mean lifetime of 50–60 days) in healthy adults.[26] The typical duration of the Amadori rearrangement is approximately 2 h, but can be modulated by labile hemoglobin. The presence of HbA_{1c}, during the lifespan of RBCs, underpins the use of the ADAG formula[27] as a biomarker of average plasma glucose levels. In contrast, the rapid dynamics of labile hemoglobin justifies studying its potential as a biomarker for shorter-term glucose variations. Integrating both biomarkers could refine the estimation of blood plasma glucose fluctuations.

A strong correlation exists between increments of labile hemoglobin and glucose

Equipped with the UMAP template derived from the 40,652 general participants, we turned our attention to the second cohort, comprising 60 pediatric subjects, who had undergone a glucose tolerance test (refer to Tables 1 and 2 for comprehensive information on the consecutive measurements). Figure 3A depicts the differences between successive labile hemoglobin and glucose measurements, taken at 30-min intervals over a 2-h period. Our analysis revealed a significant correlation between the consecutive increments of HbA_{1d} and glucose, obtained at times t and $t - 30$ min. Figure 3B shows the same correlation using the mathematical model described in Box 1. The correlations generated from the mathematical model closely mirror those observed experimentally. These findings, unattainable with the mathematical model from,[21] provided a first test to assess our model's ability to accurately fit the data, using a combination of patient-specific data and common population parameters, thus indicating an accurate interpretation of glycation dynamics in the studied subjects. This agreement was further confirmed by the high coefficients of determination in both instances (see Figures 3A and 3B). Moreover, both Figures 3A and 3B suggest that, as time advanced, the dispersion of the data points along the regression line displaced from the upper right corner toward the origin (where both Δ Glucose and Δ HbA_{1d} are equal to zero), thereby acting as a fixed point attractor.

Labile hemoglobin provides additional information of monophasic and biphasic blood glucose curves

We applied again the mathematical model described in Box 1 to capture the temporal dependence of the specific HbA_{1d} curve for each pediatric subject. The parameters entering into the model equations were adjusted to each individual participant, as detailed in the STAR methods and supplemental information. The consistency of our model results with the single point dataset was further scrutinized in Figure S2. To compare the model results for HbA_{1d} and HbA_{1c} with those from the pediatric patients, we fed the interpolated glucose profile into model

Box 1. Biomathematical model for the kinetics of glycemic biomarkers

To provide a simple conceptual mathematical framework capturing the temporal variations of glucose, HbA_{1c}, and HbA_{1d} recorded from the tolerance tests of the pediatric subjects, we extended the model presented in.[21] Our model describes the glycation kinetics of HbA_{1c} and HbA_{1d} within red blood cells (RBCs). During their lifespan, RBC hemoglobin undergoes progressive glycation, a process primarily influenced by blood plasma glucose concentration and the presence of $GLUT_1$ transporters on the RBC membrane. Typically, glycation involves two main steps; first, a glucose molecule condenses with an amino group of free hemoglobin to form a Schiff's base (aldimine), leading to HbA_{1d}. In a second step, an Amadori rearrangement of the aldimine linkage results in the stable ketoamine form, giving rise to HbA_{1c}. While HbA_{1d} can dissociate back into free hemoglobin and glucose due to the unstable first reaction, once HbA1c is formed it remains irreversibly inside the RBC until the cell dies. In our study, we considered two distinct labile hemoglobin forms, each with unique reaction, dissociation, and glycation rates, as previously reported in.[24] These forms, represented as the *fast* and *slow* labile hemoglobins HbA_{1d}^{fast} and HbA_{1d}^{slow}, respectively, create new interdependencies, as the concentration of free hemoglobin relies on both complexes. Given that the total hemoglobin within RBCs remains practically constant throughout their lifespan, free hemoglobin can be calculated in terms of HbA_{1c}, HbA_{1d}^{fast}, and HbA_{1d}^{slow}. Thus, we may express these variables as fractions of total hemoglobin, which we rename for convenience as h_S, h_{L_1} and h_{L_2} respectively. The resulting system of ordinary differential equations for these three types of hemoglobins, dependent on the time variation of blood plasma glucose concentration, $G(t)$, is

$$\frac{dh_{L_1}}{dt} = k_{FL_1} G(t)\left(1 - h_{L_1} - h_{L_2} - h_S\right) - k_{LF_1} h_{L_1} - k_{LS_1} h_{L_1} - \frac{h_{L_1}}{\tau_{RBC}}, \quad \text{(Equation 1)}$$

$$\frac{dh_{L_2}}{dt} = k_{FL_2} G(t)\left(1 - h_{L_1} - h_{L_2} - h_S\right) - k_{LF_2} h_{L_2} - k_{LS_2} h_{L_2} - \frac{h_{L_2}}{\tau_{RBC}}, \quad \text{(Equation 2)}$$

$$\frac{dh_S}{dt} = k_{LS_1} h_{L_1} + k_{LS_2} h_{L_2} - \frac{h_S}{\tau_{RBC}}. \quad \text{(Equation 3)}$$

The key parameters entering into (1)_(3) are:

- k_{FL_1} is the rate constant that accounts for the transformation of free hemoglobin into the fast labile form, a process reliant on the interaction of free hemoglobin and plasma glucose concentration. Thus, the reaction in Equation 1 is proportional to the concentration of both. Similarly, k_{FL2} is the rate constant for the transformation of free hemoglobin into the slow labile form, as displayed in Equation 2.
- The rate constants k_{FL_1} and k_{FL_2}, arising in Equations 1 and 2, respectively, represent the inverse reaction process.
- k_{LS_1} in Equation 1 is the rate constant determining the Amadori rearrangement that locks the molecule into its HbA_{1c} form. This transformation leads to a reduction in labile hemoglobin, as it either reverts back to free hemoglobin or becomes HbA1c. These terms have their equivalent version for slow hemoglobin in Equation 2. The counterpart of k_{LS_1} for the second and third equation is defined by the rate $2k_{LS_2}$.
- Finally, τ_{RBC} denotes the mean lifetime of an RBC. Upon apoptosis of the RBC, all hemoglobin forms disappear.

Other relevant details can be found in the STAR methods, as well as a diagram of the biological process in Figure S4.

Equations 1, 2, and 3. Moreover, we analyzed the stability of the obtained solutions from the model system (1)-(3). Using the kinetic parameters that best fitted the subject dataset along with the interpolated glucose function, we found that the solutions were asymptotically stable. A more detailed explanation of the procedure employed for the stability analysis is provided in the STAR methods and the supplemental information.

Figure 3C showcases a representative sample of 16 out of the total 60 pediatric subjects. The blue curve represents a cubic spline interpolation of the glucose measurements, while the red curve represents the model-obtained HbA_{1d} responses. The recorded data points for glucose and HbA_{1d} are marked with solid circles. Our model successfully replicated aspects that other single labile models could not, such as the rapid response of HbA_{1d} to glucose changes, accommodating the different glucose levels, as well as the ages, physiology and health conditions of each individual. In addition, each subfigure includes the range values of HbA_{1c} during the procedure for each participant. As expected, during the course of the tolerance test, HbA_{1c} showed relatively small variations (less than 10%) in each subject. This was also a feature confirmed by our model.

Oral glucose tolerance tests typically yield monophasic and biphasic blood glucose curves. These response patterns are instrumental in distinguishing physiologically distinct groups of individuals based on variations in insulin secretion and sensitivity.[28] Among the 60 pediatric

Table 2. Mean measurements and standard deviations of pediatric subjects at different intervals

Measurement	HbA_{1d} (%)	HbA_{1c} (%)	Glucose (%)
Pediatric 1 mean (sd)	1.78 (0.24)	5.20 (0.33)	83.17 (8.34)
Pediatric 2 mean (sd)	2.02 (0.30)	5.17 (0.32)	131.00 (31.57)
Pediatric 3 mean (sd)	2.05 (0.33)	5.19(0.32)	126.58 (36.78)
Pediatric 4 mean (sd)	2.06 (0.33)	5.18(0.31)	120.68 (33.70)
Pediatric 5 mean (sd)	2.05 (0.35)	5.16 (0.31)	117.76 (31.05)

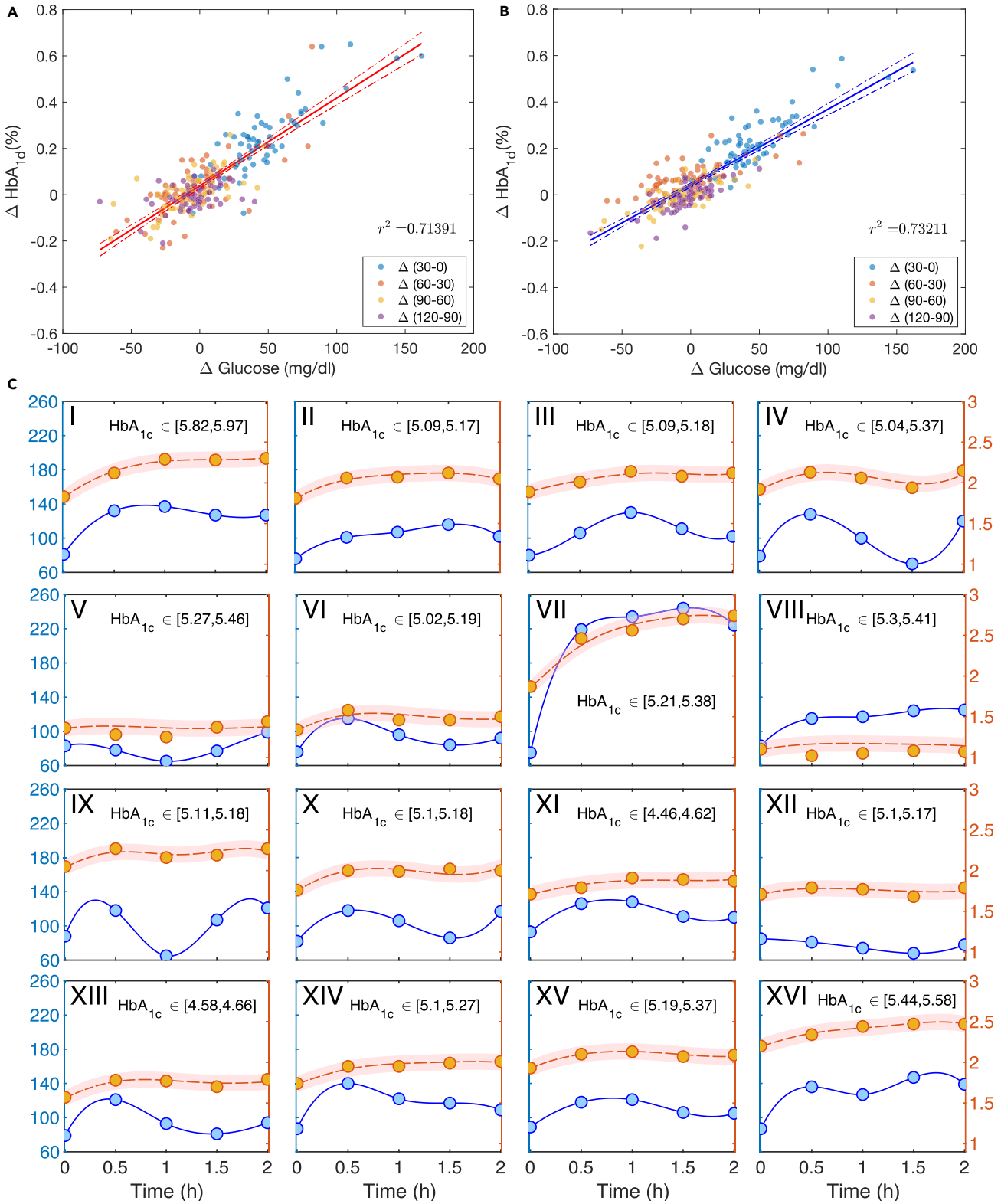


Figure 3. Modeling the glucose and labile hemoglobin of the pediatric subject cohort during the glucose tolerance test

(A) Experimental correlation between increments of HbA_{1d} and glucose for the 60 pediatric subject dataset. The obtained slope and intercept are 0.0038 and 0.0385, respectively.

Figure 3. *Continued*

(B) Simulated correlation using the mathematical model. The found slope and intercept are 0.0033 and 0.0424, respectively.

(C) Measured and calculated glucose and HbA_{1d}. Blue and orange circles represent the recorded subject values for glucose and HbA_{1d}, respectively. The blue solid curves correspond to cubic spline interpolation of the glucose measurements, while the red curves to the model-computed HbA_{1d} profiles (see the STAR methods for an explanation of how the curves were computed). The salmon color bands describe a 10% variability in the model's parameters. The root-mean-square error (RSME) between the measured and computed values was of 0.036 % for HbA_{1d}. The pre-existing pathologies of the 16 subjects, prior to the oral glucose tolerance test, were: (I) Celiac disease; (II) Obesity; (III) Obesity; (IV) Altered Basal Glucose; (V) Obesity; (VI) Obesity and mild insulin resistance; (VII) Cystic fibrosis; (VIII) Obesity; (IX) Obesity; (X) Obesity; (XI) Obesity and mild insulin resistance; (XII); Obesity; (XIII) Obesity; (XIV) Obesity; (XV) Obesity; (XVI) Obesity.

subjects considered in our study, none had a prior diagnosis of diabetes or prediabetes, although other known pathologies were prevalent, such as obesity, mild insulin resistance, and fatty liver disease (see Table 3). Notice that a number of participants had several disorders. Our analysis of the tolerance test results revealed both monophasic and biphasic glucose curves (refer to Figure 3C). In most instances, but not all, the HbA_{1d} profiles mimicked the glucose trends, albeit with less pronounced variability. Interestingly, three patients exhibited monophasic glucose and HbA_{1d} profiles, a feature shared with 17 other participants. These three patients, however, had notably higher peak values (exceeding 200 mg/dL), as exemplified for one of them (patient VII) in Figure 3C. This observation was indicative of a potential diabetic condition among these three patients. Yet, all 60 pediatric cases examined had HbA_{1c} values well below the 6.5% set by the ADAG formula as the threshold level for the diagnosis of diabetes,[27] as indicated in each of the subfigures corresponding to a sample of 16 subjects in Figure 3C. Our results are compatible with the previously explored smoothing and delaying effect of labile hemoglobin with respect to subject glucose profiles, allowing for a more complete analysis of resistance glucose tests and better patient monitoring.

Moreover, we probed the applicability and sensitivity of our model in alternate scenarios. For this, we generated synthetic data to establish glucose profiles that aligned with extant results (refer to supplemental information for details). Our analysis revealed that the labile reaction rates k_{LF_1} and k_{LF_2} are the most important parameters for our dual HbA_{1d} complexes. We further determined that the mean lifetime of RBCs, τ_{RBC}, significantly influences HbA_{1c}. Utilizing Floquet theory, we established the stability of our equations within a realistic range of HbA_{1d} functions. Lastly, we examined the evolution of the simulated HbA_{1d} when aligned with real data from diabetic patients. Our findings substantiate that it imparts a smoothing effect on the glucose data, introducing a slight delay. This observation supports the use of HbA_{1d} as a reliable indicator of past glucose states.

Persistently high values of HbA_{1d} and glucose on the UMAP predict diabetes in pediatric patients

Following the administration of oral glucose tolerance tests to the 60 pediatric participants, we sought to determine whether by combining these test results with the UMAP template, derived from the 40,652 general participants, could facilitate a definitive diabetes diagnosis. This

Table 3. Characteristics and pre-existing pathologies of the 60 pediatric subjects studied, prior to the oral glucose tolerance test

Table 2- Pathologies

Pathology	Number of subjects (%)
Obesity	46 (76.7)
Mild insulin resistance	34 (56.7)
Fatty liver disease	9 (15.0)
Adrenarche	9 (15.0)
Dyslipidemia	5 (8.3)
Overweight	5 (8.3)
Early puberty	4 (6.7)
Altered basal glucose	3 (5.0)
Celiac disease	2 (3.3)
Sleep apnea	2 (3.3)
Adrenal hyperplasia	2 (3.3)
Mild child obesity	1 (1.7)
Ketosis	1 (1.7)
Hyperglycemia	1 (1.7)
Cystic fibrosis	1 (1.7)
Down syndrome	1 (1.7)
Hypoglycemia	1 (1.7)
Gynecomastia	1 (1.7)

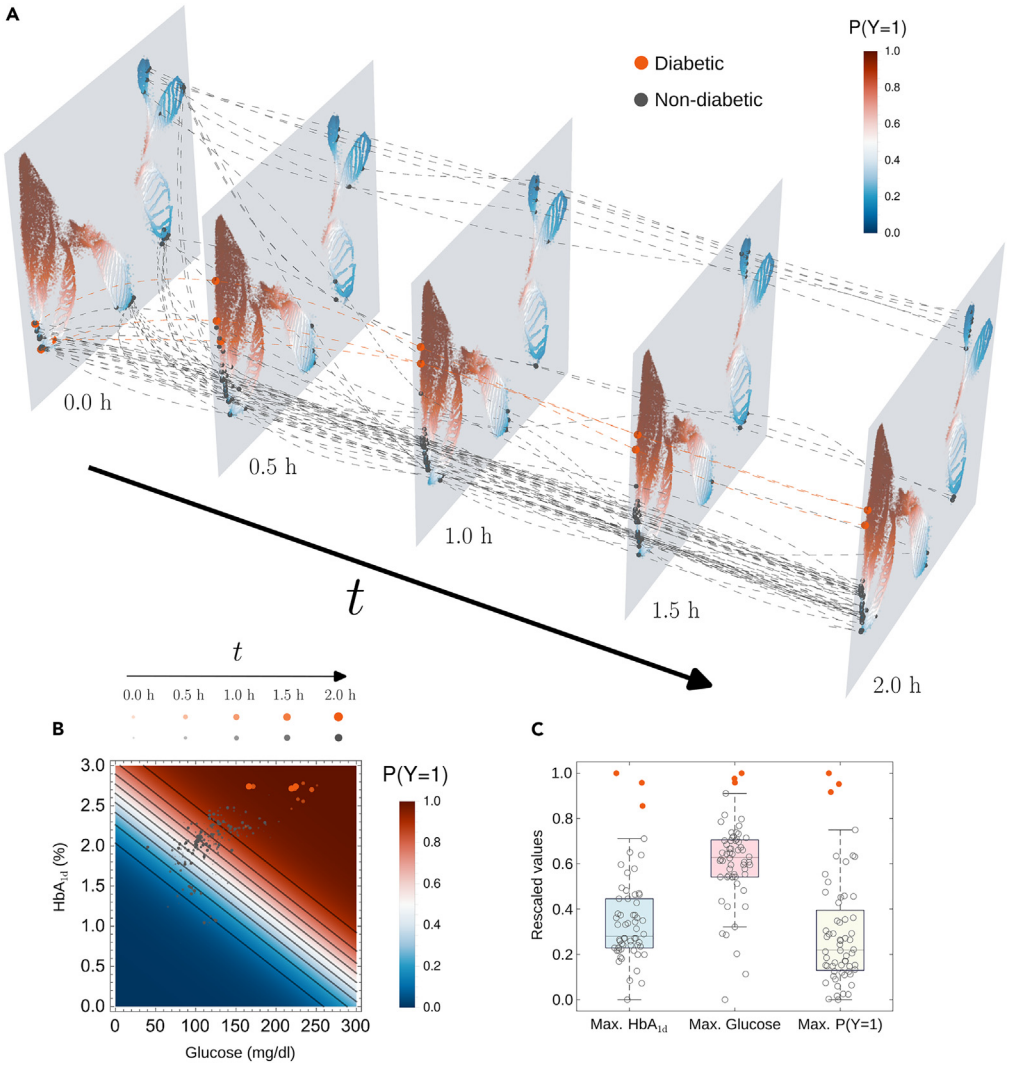


Figure 4. Fate mappings for predicting diabetes in pediatric subjects during the glucose tolerance test

(A) Pediatric subjects have been projected over the UMAP performed on the 40,652 general participant dataset by means of nearest neighbors. Their position changes in time in different probability areas. Pediatric subjects diagnosed as diabetic have been marked in orange, while non-diabetic ones have been marked in dark gray. Probabilities calculated for adult subjects are shown in a red-blue scale. The trajectories, inferred from the numerical solution of the mathematical model presented in Box 1 and highlighted as dashed gray lines, reveal the evolving dynamics.

(B) Position of pediatric subjects over a density plot of the logistic model developed using two variables (glucose and HbA_{1d}). For longer times, pediatric subjects diagnosed as diabetic remain in higher probability areas (orange dots).

(C) Maximum values of HbA_{1d}, glucose and probability of being diabetic (predicted using the logistic model) for the pediatric subjects, rescaled into a [0, 1] range. Orange dots represent the subjects diagnosed as diabetic, which are outliers in the higher range (values close to 1).

was particularly relevant for three patients who exhibited potential diabetic conditions. To this end, we tracked the glucose and HbA_{1d} levels in all pediatric individuals, as depicted in Figure 4. At each time point recorded during the tolerance tests, we projected the values of HbA_{1c}, glucose, and HbA_{1d} onto the UMAP template, representing them as trajectories. Notably, our analysis revealed a distinct divergence of these trajectories among different subject subgroups (see Figure 4A). Specifically, diabetic patients consistently clustered within the high-probability region (depicted in dark red, where the probability $P > 0.8$), thereby distinguishing them from non-diabetic subjects. The driving force behind this behavior becomes clearer when examining Figure 4B. It is evident that sustained elevations of both glucose and HbA_{1d} levels significantly contribute to the tendency of diabetic patients to cluster within the specified UMAP region. This association becomes particularly striking when these elevated biomarker levels persist over time. For further clarity, we rescaled the maximum values of glucose, as it is also a marker for diabetes,[29] HbA_{1d}, and diabetes probability (predicted using the logistic model) to fall within the range [0,1], as illustrated in Figure 4C. Through this procedure, a distinctive pattern emerged. Remarkably, those pediatric subjects that were later diagnosed as diabetic

stood out as discernible outliers residing in the upper range of the rescaled values, with probabilities close to 1. This finding constitutes a significant hallmark, demonstrating robust predictive power for diabetes diagnosis.

DISCUSSION

Over the past few years, the development of quantitative methods to forecast glucose metabolism levels has attracted considerable attention in biomedical research.[3,21,22,30–34] The advent of wearable monitoring technologies has further stimulated the advancement of increasingly complex mathematical models and efficient data processing algorithms capable of handling an influx of real-time datasets.[12,14,35] While these developments provide significant benefits for following distinct patient groups in clinical trials and validating emerging hypotheses and mechanisms, their wide accessibility remains a challenge. This highlights the imperative need for innovative methodologies that can evaluate glucose levels using resources that are readily available to patients globally.

In this study, we have leveraged both a large retrospective cohort of 40,652 general participants and a prospective pediatric cohort of subjects to provide substantial evidence supporting the diagnostic and monitoring capabilities of labile hemoglobin (HbA_{1d}), which reflects glycemic levels in the hours prior to sampling.[18,21] This marker's relation to glycated and labile hemoglobin has been studied previously.[36] Our methodology combines NDR techniques with mathematical modeling based on a set of ordinary differential equations, capturing the kinetics of glucose, HbA_{1c}, and HbA_{1d}. Our strategy involved constructing a probability density map using the large retrospective general participant group. This map represented the probabilities of a diabetes diagnosis, encapsulating the topological structure of fasting plasma glucose, HbA_{1c}, and HbA_{1d} in an integrated format. Furnished with this comprehensive template, we tracked the dynamic responses of the prospective pediatric cohort following the oral administration of a glucose tolerance test. This approach enabled the unambiguous identification in the pediatric cohort of diabetic subjects, even those exhibiting HbA_{1c} values significantly below the threshold levels set by the widely used ADAG formula.[27] Our work also revealed strong correlations between increments of HbA_{1d} and glucose in the 60-subject pediatric dataset, corroborating these findings through the proposed mathematical model.

A distinctive feature of our mathematical model is its ability to capture both short-term, ranging from minutes to few hours, and long-term, spanning several weeks up to three months, kinetics of glucose, HbA_{1c}, and HbA_{1d}. This dual functionality distinguishes our model from previous ones, which could not accurately reproduce the correlation between HbA_{1d} and plasma glucose over short timescales typical of oral glucose tolerance tests. To overcome this limitation, we decomposed HbA_{1d} into two separate complexes, *fast* and *slow*, motivated by an earlier study that examined the formation and dissociation of unstable hemoglobin complexes, which arise from spontaneous reactions between glucose and hemoglobin amino groups.[24] We hypothesized that swift, minor alterations in labile hemoglobin might originate from a less abundant but faster component, HbA_{1d}^{fast}. This division allows a minor proportion of HbA_{1d}^{fast} to change rapidly. In contrast, when averaging these rapid changes in the slower component, HbA_{1d}^{slow}, the overall rate of labile hemoglobin alteration is too slow to account for the observed correlation. This explains the inadequacy of models based solely on a single HbA_{1d} compartment in short-term analysis. By focusing primarily on the faster component, our model establishes a more robust correlation with blood plasma glucose fluctuations, thereby providing a plausible explanation for the observations displayed in Figures 3A and 3B. However, while this study demonstrates the utility of the mathematical model in encompassing these kinetic aspects, further validation of its predictive capabilities is crucial. This includes addressing longer time ranges and incorporating additional data points to assess its predictive accuracy. Future research should concentrate on validating and refining the model using a more diverse patient dataset, considering also different stages of diabetes progression and treatment regimes.

The role of NDR has played an important role in the present work, particularly in discerning the differential patterns between diabetic and non-diabetic individuals. Leveraged extensively in recent studies,[37] this computational approach has allowed us to unveil the intricate influence of fasting plasma glucose (FPG), HbA_{1c}, and HbA_{1d} on subject clusters and subclusters. While HbA_{1c} is a well-established diabetes biomarker, here we have tried to carry out a synergistic analysis of FPG, HbA1c, and HbA_{1d}, harmonizing these three key indicators through NDR. Our analysis of the general participant dataset revealed significant correlations among the chosen variables and offered valuable insights into the dispersion patterns of diabetic and non-diabetic individuals based on each of those three variables. Utilizing logistic models, we quantified the classification ability of these indicators using the McFadden likelihood ratio index. Interestingly, the threshold values identified were in close alignment with the standards posited in existing literature.[38,39] Moreover, beyond the scope of the retrospective general participants, we extended our research to a pediatric cohort exhibiting elevated levels of plasma glucose and HbA_{1d}. Portraying these data points onto the UMAP uncovered a clear pattern: pediatric subjects persistently demonstrating high levels of plasma glucose and HbA were at an increased risk of being diagnosed with diabetes. This observation underscores the potential of HbA_{1d} as a promising biomarker in early diabetes detection and highlights the significance of our approach in diabetes research.

Our investigation demonstrates the utility of HbA_{1d} as a robust biomarker, when used in conjunction with HbA_{1c} and FPG, for identifying individuals in a pre-diabetic or diabetic state. Multiple lines of evidence substantiate this assertion: (1) Figure 1 illustrates a significant correlation between diagnostic values (Figure 1A) and HbA_{1d} levels (Figure 1D), signifying a profound association between diagnostic status and HbA_{1d}; (2) As depicted in Figure 2, HbA_{1d} displays substantial predictive efficacy, as indicated by the McFadden likelihood ratio index (Figure 2E) and the corresponding p value in Table S1. Notably, it effectively identifies diabetic patients who might evade detection by HbA_{1c} alone, elucidated in Figure 2A (highlighted in red). This underscores the complementary role of HbA_{1d} in identifying patients overlooked by conventional diagnostic methods; (3) Figure 4 reveals a consistent pattern where elevated HbA_{1d} values precisely align with pediatric subjects subsequently diagnosed as diabetic (depicted in orange). These children consistently occupied the higher probability and HbA_{1d} range in Figures 4A and 4B. Furthermore, their maximum HbA_{1d} values emerged as outliers in Figure 4C, emphasizing the unique potential of HbA_{1d} in identifying and characterizing diabetic patients. Taken together, these findings underscore the clinical value of measuring HbA_{1d} in

predicting diabetes. Regular monitoring of HbA_{1d} levels, especially when combined with other diagnostic markers, could significantly enhance the precision of diabetes detection, enabling timely interventions and improving patient outcomes. Its use as a biomarker does not entail additional costs. Given that its fraction is concurrently measured with HbA_{1c} via chromatography, its clinical integration is both feasible and economical. Moreover, even if the proportion of diabetic patients detected solely by HbA_{1d} would appear as relatively small (refer to the red cluster in Figure 2A), this percentage, when juxtaposed with global diabetes incidence rates, translates to a significant number of patients potentially missed otherwise.

In conclusion, this work has not only highlighted the potential of labile hemoglobin (HbA_{1d}) as a relevant biomarker for diabetes research but has also proposed a proof of concept for its synergistic use with fasting plasma glucose (FPG) and HbA_{1c} to create a comprehensive diagnostic tool. By employing nonlinear dimensionality reduction techniques on single-time point measurements from a large retrospective dataset of 40,652 general participants, we generated a probabilistic structural map that distinguished between diabetic and non-diabetic individuals. This method was then successfully validated with data from glucose tolerance tests conducted on a cohort of 60 pediatric subjects. Our mathematical model, grounded in a set of ordinary differential equations, captured the kinetic intricacies of FPG, HbA_{1d}, and HbA_{1c}, facilitating the projection of these dynamic responses onto the structural map, thereby identifying subsets of pediatric subjects predisposed to diabetes. The innovative strategy presented in this work, merging mathematical modeling, advanced data visualization, and integrative biomarker analysis, offers broad applicability that extends beyond diabetes, potentially paving the way for novel diagnostic and monitoring methodologies in biomedical research. It could be applied, for instance, in oncology, where understanding the complex interactions between diverse genetic/phenotypic markers and treatment responses is crucial. NDR techniques have already been instrumental in tumor characterization.[40,41] By integrating NDR, which typically relies on sparse data snapshots, with differential equation-based mathematical models,[42–44] there is potential to dynamically identify distinct patient subgroups based on their genetic/phenotypic profiles and treatment outcomes. This could greatly enhance personalized therapeutic interventions.

Limitations of the study

To further validate the predictions of our ordinary differential equation–based mathematical model put forward in the present study, it would be necessary to incorporate several elements. These include the glycation index, which measures the gap between the average glucose of patients and their HbA_{1c} standard, and generally requires continuous glucose monitoring. We considered only pediatric subjects in the prospective stage of our study, but this should be extended to larger cohorts including adults. Moreover, the model was applied to a temporal window of 2 h (five time points), that corresponded to the available data from the pediatric subjects. This window should be expanded through longer explorations of subjects, spanning days rather than hours, to have a more robust validation of its predictive capability. Additionally, it would be advisable to have constant glucose monitoring data in confirmed type I diabetic patients with accompanying regularly measured labile hemoglobin data.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCES AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Cohort of retrospective general participants assayed for plasma glucose, glycated hemoglobin, and labile hemoglobin under fasting conditions
  - Cohort of pediatric subjects subjected to an oral glucose tolerance test
- METHODS DETAILS
  - Dimensionality reduction
  - General subject dataset visualization
  - Logistic models
  - Linear regression analysis
  - Kernel density estimation
  - Density histograms
  - Mathematical model
  - Model derivation
  - Numerical solution of the system of differential equations
  - Sensitivity and stability analysis
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Statistical analysis of linear correlation of difference of labile hemoglobin and glucose levels in pediatric subjects
  - Statistical analysis in logistic regression

- Kolmogorov-Smirnov tests and regression of general subject cohort
- ADDITIONAL RESOURCES

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.isci.2024.109369.

ACKNOWLEDGMENTS

We thank three anonymous reviewers for very helpful comments that greatly improved the manuscript. This work was supported by the Spanish Ministerio de Ciencia e Innovación, MCIN/AEI/10.13039/501100011033 (grant numbers PID2019-110895RB-I00 and PID2022-142341OB-I00) and by the Junta de Comunidades de Castilla-La Mancha (grant SBPLY/19/180501/000211). J.A.R.-R. and D.G.A. were supported, respectively, by the AECC-Asociación Española Contra el Cáncer (contract 2023-PRED-35671) and by the Spanish Ministerio de Ciencia e Innovación (contract 2023-CDT-11616 from a project with grant number TED2021-132296B-C55).

AUTHOR CONTRIBUTIONS

Conceptualization, G.F.C. and H.S.; model development and simulation, J.A.R.-R., D.G.A., and G.F.C.; formal analysis, J.A.R.-R., D.G.A., and G.F.C.; data acquisition, J.E.-S., M.G.B., A.M.D., F.J.M.L., M.A.S.M., I.N.J., M.J.C.Z., and M.R.D.; data curation, J.A.R.-R. and D.G.A.; software, J.A.R.-R. and D.G.A.; visualization, D.G.A., J.A.R.-R., and G.F.C.; writing–original draft, G.F.C., J.A.R.-R., and D.G.A.; writing-review and editing, J.A.R.-R., D.G.A., H.S., J.B.B., M.R.D., and G.F.C.; supervision, G.F.C. and M.R.D.; funding acquisition: G.F.C. and J.B.B.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: July 31, 2023
Revised: February 16, 2024
Accepted: February 26, 2024
Published: March 1, 2024

REFERENCES

1. Sun, H., Saeedi, P., Karuranga, S., Pinkepank, M., Ogurtsova, K., Duncan, B.B., Stein, C., Basit, A., Chan, J.C.N., Mbanya, J.C., et al. (2022). Idf diabetes atlas: Global, regional and country-level diabetes prevalence estimates for 2021 and projections for 2045. Diabetes Res. Clin. Pract. *183*, 109119.
2. Kim, J., Campbell, A.S., de Ávila, B.E.F., and Wang, J. (2019). Wearable biosensors for healthcare monitoring. Nat. Biotechnol. *37*, 389–406.
3. Tyler, N.S., Mosquera-Lopez, C., Young, G.M., El Youssef, J., Castle, J.R., and Jacobs, P.G. (2022). Quantifying the impact of physical activity on future glucose trends using machine learning. iScience *25*, 103888.
4. Beck, R.W., and Bergenstal, R.M. (2021). Beyond a1c—standardization of continuous glucose monitoring reporting: Why it is needed and how it continues to evolve. Diabetes Spectr. *34*, 102–108.
5. Karter, A.J., Parker, M.M., Moffet, H.H., Gilliam, L.K., and Dlott, R. (2021). Association of real-time continuous glucose monitoring with glycemic control and acute metabolic events among patients with insulin-treated diabetes. JAMA, J. Am. Med. Assoc. *325*, 2273–2284.
6. Chew, N.W.S., Ng, C.H., Tan, D.J.H., Kong, G., Lin, C., Chin, Y.H., Lim, W.H., Huang, D.Q., Quek, J., Fu, C.E., et al. (2023). The global burden of metabolic disease: Data from 2000 to 2019. Cell Metabol. *35*, 414–428.e3.
7. Lenters-Westra, E., Schindhelm, R.K., Bilo, H.J., and Slingerland, R.J. (2013). Haemoglobin a1c: Historical overview and current concepts. Diabetes Res. Clin. Pract. *99*, 75–84.
8. Koenig, R.J., Peterson, C.M., Jones, R.L., Saudek, C., Lehrman, M., and Cerami, A. (1976). Correlation of glucose regulation and hemoglobin aic in diabetes mellitus. N. Engl. J. Med. *295*, 417–420.
9. Beach, K.W. (1979). A theoretical model to predict the behavior of glycosylated hemoglobin levels. J. Theor. Biol. *81*, 547–561.
10. Treviño, G. (2006). On a1c and its dependence on pg level. Diabetes Res. Clin. Pract. *73*, 111–112.
11. Malka, R., Nathan, D.M., and Higgins, J.M. (2016). a. Mechanistic modeling of hemoglobin glycation and red blood cell kinetics enables personalized diabetes monitoring. Sci. Transl. Med. *8*, 359ra130.
12. Chehregosha, H., Khamseh, M.E., Malek, M., Hosseinpanah, F., and Ismail-Beigi, F. (2019). A view beyond hba1c: role of continuous glucose monitoring. Diabetes Ther. *10*, 853–863.
13. Zhong, V.W., Juhaeri, J., Cole, S.R., Shay, C.M., Gordon-Larsen, P., Kontopantelis, E., and Mayer-Davis, E.J. (2019). Proximal hba1c level and first hypoglycemia hospitalization in adults with incident type 2 diabetes. J. Clin. Endocrinol. Metab. *104*, 1989–1998.
14. Lu, J., Ma, X., Zhang, L., Mo, Y., Lu, W., Zhu, W., Bao, Y., Jia, W., and Zhou, J. (2020). Glycemic variability modifies the relationship between time in range and hemoglobin a1c estimated from continuous glucose monitoring: a preliminary study. Diabetes Res. Clin. Pract. *161*, 108032.
15. Sehrawat, T., Jindal, A., Kohli, P., Thour, A., Kaur, J., Sachdev, A., and Gupta, Y. (2018). Utility and limitations of glycated hemoglobin (hba1c) in patients with liver cirrhosis as compared with oral glucose tolerance test for diagnosis of diabetes. Diabetes Ther. *9*, 243–251.
16. Radin, M.S. (2014). Pitfalls in Hemoglobin A1c Measurement: When Results may be Misleading. J. Gen. Intern. Med. *29*, 388–394.
17. Kato, S., Otaka, I., Toyama, H., Kusumi, R., Takahashi, K., Nara, M., Suganuma, Y., Sato, T., Morii, T., Fujita, H., and Waki, H. (2022). Cases of fulminant type 1 and type 2 diabetes mellitus whose hba1c levels were unmeasurable due to increased labile hba1c. Diabetol. Int. *13*, 698–703.
18. Delanghe, J.R., Lambrecht, S., Fiers, T., and Speeckaert, M.M. (2022). Labile glycated hemoglobin: an underestimated laboratory marker of short term glycemia. Clin. Chem. Lab. Med. *60*, 451–455.
19. Koga, M., Inada, S., and Miyazaki, A. (2016). Identification of the Presence of Variant Hemoglobin Using a Measurement of the Labile HbA1c (#C) Fraction. Ann. Clin. Lab. Sci. *46*, 387–392. http://www.annclinlabsci.org/content/46/4/387.long.
20. Pradhan, R., Batabyal, S.K., Maiti, S., and Chatterjee, S. (2013). Short-term changes in plasma glucose affect glycated hemoglobin measurement. Diabetes Res. Clin. Pract. *100*, e17–e19.
21. León-Triana, O., Calvo, G.F., Belmonte-Beitia, J., Rosa Durán, M., Escribano-Serrano, J., Michan-Doña, A., and Pérez-García, V.M. (2018). Labile haemoglobin as a glycaemic

biomarker for patient-specific monitoring of diabetes: mathematical modelling approach. J. R. Soc. Interface *15*, 20180224.

22. Kurata, H. (2021). Virtual metabolic human dynamic model for pathological analysis and therapy design for diabetes. iScience *24*, 102101.
23. McInnes, L., Healy, J., Saul, N., and Großberger, L. (2018). UMAP: Uniform Manifold Approximation and Projection. J. Open Source Softw. *3*, 861.
24. Hempe, J.M., McGehee, A.M., Hsia, D., and Chalew, S.A. (2012). Characterization of unstable hemoglobin A1c complexes by dynamic capillary isoelectric focusing. Anal. Biochem. *424*, 149–155.
25. Mueckler, M., and Thorens, B. (2013). The SLC2 (GLUT) Family of Membrane Transporters. Mol. Aspect. Med. *34*, 121–138.
26. Franco, R.S. (2012). Measurement of Red Cell Lifespan and Aging. Transfus. Med. Hemotherapy *39*, 302–307.
27. Klonoff, D.C. (2014). ADAG Study Group Data Links A1C Levels with Empirically Measured Blood Glucose Values - New Treatment Guidelines Will Now be Needed. J. Diabetes Sci. Technol. *8*, 439–443.
28. Kim, J.Y., Michaliszyn, S.F., Nasr, A., Lee, S., Tfayli, H., Hannon, T., Hughan, K.S., Bacha, F., and Arslanian, S. (2016). The shape of the glucose response curve during an oral glucose tolerance test heralds biomarkers of type 2 diabetes risk in obese youth. Diabetes Care *39*, 1431–1439.
29. American Diabetes Association (2020). 2. Classification and Diagnosis of Diabetes: *Standards of Medical Care in Diabetes—2020*. Diabetes Care *43*, S14–S31. https://diabetesjournals.org/care/article/43/Supplement_1/S14/30640/2-Classification-and-Diagnosis-of-Diabetes.
30. Ajmera, I., Swat, M., Laibe, C., Le Novère, N., and Chelliah, V. (2013). The impact of mathematical modeling on the understanding of diabetes and related complications. CPT Pharmacometrics Syst. Pharmacol. *2*, e54.
31. Malka, R., Nathan, D.M., and Higgins, J.M. (2016). b. Mechanistic modeling of hemoglobin glycation and red blood cell kinetics enables personalized diabetes monitoring. Sci. Transl. Med. *8*.
32. Faruqui, S.H.A., Du, Y., Meka, R., Alaeddini, A., Li, C., Shirinkam, S., and Wang, J. (2019). Development of a deep learning model for dynamic forecasting of blood glucose level for type 2 diabetes mellitus: secondary analysis of a randomized controlled trial. JMIR Mhealth Uhealth *7*, e14452.
33. Kameyama, M., Okumiya, T., Tokuhiro, S., Matsumura, Y., Matsui, H., Ono, Y., Iwasaka, T., Hiratani, K., and Koga, M. (2021). Estimation of the hemoglobin glycation rate constant. Sci. Rep. *11*, 986.
34. Wang, P., Yuan, D., Zhang, C., Jia, S., Song, Y., Tang, X., Zhao, X., Gao, R., Xu, B., and Yuan, J. (2023). Association between cumulative lipoprotein(a) exposure and adverse cardiovascular outcomes in patients with prediabetes or diabetes. iScience *26*, 106117.
35. Rodbard, D. (2017). Continuous glucose monitoring: A review of recent studies demonstrating improved glycemic outcomes. Diabetes Technol. Ther. *19*, S25–S37.
36. Loh, T.P., Peng, W.K., Chen, L., and Sethi, S.K. (2014). Application of smoothed continuous labile haemoglobin A1c reference intervals for identification of potentially spurious HbA1c results. J. Clin. Pathol. *67*, 712–716.
37. Bej, S., Sarkar, J., Biswas, S., Mitra, P., Chakrabarti, P., and Wolkenhauer, O. (2022). Identification and epidemiological characterization of Type-2 diabetes sub-population using an unsupervised machine learning approach. Nutr. Diabetes *12*, 27.
38. American Diabetes Association (2013). Standards of Medical Care in Diabetes—2013. Diabetes Care *36*, S11–S66.
39. International Expert Committee (2009). International Expert Committee Report on the Role of the A1C Assay in the Diagnosis of Diabetes. Diabetes Care *32*, 1327–1334.
40. Cantini, L., Zakeri, P., Hernandez, C., Naldi, A., Thieffry, D., Remy, E., and Baudot, A. (2021). Benchmarking joint multi-omics dimensionality reduction approaches for the study of cancer. Nat. Commun. *12*, 124.
41. Trajanoska, K., Bhérer, C., Taliun, D., Zhou, S., Richards, J.B., and Mooser, V. (2023). From target discovery to clinical drug development with human genetics. Nature *620*, 737–745.
42. Jiménez-Sánchez, J., Bosque, J.J., Jiménez Londoño, G.A., Molina-García, D., Martínez, Á., Pérez-Beteta, J., Ortega-Sabater, C., Honguero Martínez, A.F., García Vicente, A.M., Calvo, G.F., and Pérez-García, V.M. (2021). Evolutionary dynamics at the tumor edge reveal metabolic imaging biomarkers. Proc. Natl. Acad. Sci. USA *118*, e2018110118.
43. Bosque, J.J., Calvo, G.F., Molina-García, D., Pérez-Beteta, J., García Vicente, A.M., and Pérez-García, V.M. (2023). Metabolic activity grows in human cancers pushed by phenotypic variability. iScience *26*, 106118.
44. Pillai, M., Hojel, E., Jolly, M.K., and Goyal, Y. (2023). Unraveling non-genetic heterogeneity in cancer with dynamical models and computational tools. Nat. Comput. Sci. *3*, 301–313.
45. R Core Team. *R* (2023). A Language and Environment for Statistical Computing (R Foundation for Statistical Computing). https://www.R-project.org/.
46. Wickham, H. (2016). ggplot2: Elegant Graphics for Data Analysis (Springer-Verlag New York). https://ggplot2.tidyverse.org.
47. Becht, E., McInnes, L., Healy, J., Dutertre, C.-A., Kwok, I.W.H., Ng, L.G., Ginhoux, F., and Newell, E.W. (2019). Dimensionality reduction for visualizing single-cell data using UMAP. Nat. Biotechnol. *37*, 38–44.
48. (2023). What is Sensitivity Analysis? - MATLAB & Simulink. https://www.mathworks.com/help/sldo/ug/what-is-sensitivity-analysis.html.
49. Chervoneva, I., Freydin, B., Hipszer, B., Apanasovich, T.V., and Joseph, J.I. (2014). Estimation of nonlinear differential equation model for glucose-insulin dynamics in type I diabetic patients using generalized smoothing. Ann. Appl. Stat. *8*, 886–904. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8025877/.
50. Chicone, C. (1999). Ordinary Differential Equations with Applications. In Texts in Applied Mathematics, *vol 34*Texts in Applied Mathematics (Springer). http://link.springer.com/10.1007/b97645.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Software and algorithms		
MATLAB (R2023a)	The MathWorks, Inc. Natick, MA	https://mathworks.com/products/matlab.html
MATLAB Optimization Toolbox	The MathWorks, Inc. Natick, MA	https://mathworks.com/products/optimization.html
MATLAB Statistics & Machine Learning Toolbox	The MathWorks, Inc. Natick, MA	https://mathworks.com/products/statistics.html
R programming language	R Core Team. R et al.[45]	https://www.r-project.org/
ggplot2	Wickham et al.[46]	https://ggplot2.tidyverse.org
D-100 Software	BioRad	https://www.https://www.bio-rad.com/en-es/a
Original Code	This manuscript	https://doi.org/10.5281/zenodo.10374159

RESOURCES AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Prof. Gabriel F. Calvo (gabriel.fernandez@uclm.es).

Materials availability

This study did not generate new physical material or unique reagents.

Data and code availability

- The complete original data reported in this study cannot be deposited in a public repository because these data are confidential medical records. To request access, contact Prof. Gabriel F. Calvo (gabriel.fernandez@uclm.es).
- All original code has been deposited at the github repository Diabetes applications.git and is publicly available as of the date of publication. DOIs are listed in the key resources table.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Cohort of retrospective general participants assayed for plasma glucose, glycated hemoglobin, and labile hemoglobin under fasting conditions

A cohort of 40,652 human subjects, examined during routine medical checkups, was retrospectively gathered from Hospital Universitario Puerta del Mar, Cádiz, Spain, between November 2013 and December 2016. All participants provided written informed consent before enrollment, following the ethical approval granted by the Institutional Review Board of the hospital (the reference of the research study was A1d14). The collected dataset comprised key variables, including fasting plasma glucose (FPG), glycated hemoglobin (HbA_{1c}), labile hemoglobin (HbA_{1d}), as well as age, sex, and diabetic status of each participant. Samples were collected under fasting conditions during regular medical check-ups, followed by analyses at the "Punta Europa" healthcare facilities in Algeciras and La Línea (Spain). These facilities employed an Arkray Adams HA-8160 HbA1c analyzer (Menarini Diagnostics, Barcelona), a high-performance liquid chromatography cation exchange method, for automatic quantification of both HbA_{1c} and HbA_{1d}. Furthermore, serum fasting plasma glucose levels were determined using the hexokinase method, executed with Roche Diagnostics in "Punta Europa" and Beckman Coulter AU in La Línea. This study was conducted in accordance with the guidelines of the Spanish National Program of Glycohemoglobin Normalization and received approval from the Cádiz Research Ethics Committee. The criteria for classifying subjects within this cohort were applied by a panel of expert medical doctors, independent from our data analysis team. Specifically, a participant in that cohort was deemed diabetic if at least one of these conditions was satisfied: (i) An HbA_{1c} level over 6.5%; (ii) An FPG level exceeding 126 mg/dL; (iii) Participants undergoing treatment with glucose-reducing medications.

Data on the sex and age of the participants are available in Table 1. Both sex and age were obtained from the patients' medical records, and gender identity was not recorded in the original study. Neither age nor sex was found to be significant in this context.

Cohort of pediatric subjects subjected to an oral glucose tolerance test

Sixty pediatric subjects were prospectively enrolled at the Hospital Universitario de Jerez, Jerez, Spain, forming the pediatric cohort of our study. The ethical approval for the study, granted by the Institutional Review Board of the hospital (the reference of the research study was

DIAMATH with number 44.23), was followed by participants providing their informed consent for participation (parents or tutors). Children were chosen, as they were deemed a more vulnerable group, with childhood diabetes management being an important and challenging issue for the parents or guardians. Inclusion criteria for this group involved children aged between 2 and 14 years who had undergone an oral glucose tolerance test (OGTT), and the availability of their corresponding blood glucose data along with HbA_{1c} and HbA_{1d} values. The samples were systematically collected between March 2018 and December 2020, ensuring they closely aligned with the BIO RAD system records maintained in the hospital laboratory.

Exclusion criteria were set to maintain the reliability of results and included the ingestion of food or liquids 8 h prior to the test, complications during OGTT such as nausea, signs of fever or infection, recent surgical interventions (less than two weeks ago), administration of glucose-altering medication like corticoids, and fasting blood glucose levels above a high threshold of 200 mg/dL. Additionally, patients following very low carbohydrate diets, experiencing severe emotional stress, or with coagulation difficulties that presented risks associated with obtaining blood samples were also excluded from the study.

The collected clinical data encompassed variables such as age, sex, weight, height, body mass index, basal insulin levels, and plasma glucose data from the OGTT. HbA_{1c} and HbA_{1d} fractions were determined using the D-100 HbA1c test or BIO RAD system, an ion-exchange high-performance liquid chromatography method. In this procedure, sample dilution in the D-100 was followed by injection into the test cartridge. A programmed buffer gradient of increasing ionic strength separated the hemoglobins, which were subsequently measured for changes in absorbance at 415 nm. The D-100 software used raw data from each test to calculate HbA1c values based on a two-level calibration curve and generated a report for each sample. The analysis time per sample averaged approximately 45 s, demonstrating the efficiency of the technique. As in the previous cohort, data on the sex and age of the participants are available in Table 1. Both sex and age were obtained from the patients' medical records, and gender identity was not recorded in the original study. Neither age nor sex was found to be significant in this context.

Prior to the administration of the OGTT, none of the pediatric subjects had been diagnosed with diabetes, as they did not meet the aforementioned criteria. For an individual in this cohort to be diagnosed as diabetic, based on the OGTT results, their glucose measurements during the procedure must exceed a threshold of 200 mg/dL.

METHODS DETAILS

Dimensionality reduction

We employed Uniform Manifold Approximation and Projection (UMAP) for dimensionality reduction, transforming the high-dimensional data from adult participants into a 2D space.[23,47] The purpose of this analysis was to visualize and understand significant properties related to variations in FPG, HbA_{1d}, and HbA_{1c} and their implications for diabetes diagnosis. UMAP, a nonlinear and nonparametric method, posits that the data is uniformly distributed on a locally connected Riemannian manifold with an approximately locally constant Riemannian metric. This technique involves creating a high-dimensional graph representation of the data and subsequently optimizing a low-dimensional graph to match the structure of the high-dimensional one as closely as possible. We conducted these analyses using the R programming language.[45] Prior to this, we cleaned the data, eliminating empty and non-numeric values. Furthermore, we rescaled the data to the interval $[0, 1]$ based on the original minimum and maximum values of each variable, before embarking on the dimensionality reduction.

General subject dataset visualization

Figure S1 provides a comprehensive visualization of our dataset comprising the 40,652 general participants, allowing for a raw understanding of the relationships within the three variables under investigation (HbA_{1c}, HbA_{1d}, and FPG). In Figure S1A, we present a 3D scatterplot, which offers an immediate and intuitive representation of the raw data in its original three-dimensional space. Each point in the plot corresponds to an individual subject, and the axes represent the respective variables. This visualization yields a direct examination of the distribution of all the participants. Furthermore, in Figure S1B, we present a Principal Component Analysis (PCA) linear dimensionality reduction. PCA effectively captures the variance in the data and projects it onto a two-dimensional plane. Together, these visualizations give a first overview of our data from both a raw and a simpler dimensionally-reduced perspective than the UMAP used in the main text.

Logistic models

In our analysis of the general participant dataset, we utilized logistic models to assess the classification capability of each variable involved (FPG, HbA_{1d} and HbA_{1c}). These statistical models estimate the probability of an individual being diabetic/non-diabetic according to the data available. Once the equations were determined, threshold values were calculated considering a probability of 0.5. To quantify the classification capability of the models, we employed the McFadden likelihood ratio index (R^2_{McF}), which is a normalized measure in the range $[0,1]$, with higher values meaning a better capability. The full model developed using the three variables together was determined after splitting the data available between a training set (80%) and a validation set (20%). The validity of the model was rigorously examined. First, a confusion matrix and a ROC curve were calculated from the validation set, providing a comprehensive understanding of the predictive performance of the model. Furthermore, 10-fold cross-validation was employed to thoroughly assess the stability and robustness of the model under different data partitions. This approach ensured that the performance of the model remained consistent.

In Table S1, we show the statistical findings obtained after performing a logistic regression training. Additionally, the validation results are detailed in Figure S3. Figure S3A provides insights into the model performance through a confusion matrix, while Figure S3 B illustrates the

ROC curve, offering a overview of the validation process. For a more comprehensive evaluation, Table S2 showcases the coefficients derived through a 10-fold cross-validation, including ranges of model parameters and area under the ROC curve (AUC) values. All variables included in the model (HbA_{1c}, FPG and HbA_{1d}) proved useful in predicting diabetes, with a confidence level of 99.9%.

Linear regression analysis

To ascertain correlations between several variables of interest within both pediatric and the general groups, and to assess the significance of the effects, we employed linear regression analysis. Initially, we applied it to the differences in labile glucose and fasting plasma glucose at successive times in pediatric data. Subsequently, we also examined the correlation between pairs of variables (FPG, HbA_{1d} and HbA_{1c}) in the general participant group. We estimated the unknown parameters using ordinary least squares. The linear regression analyses were performed using MATLAB, specifically the fitlm function.

Kernel density estimation

We applied kernel density estimation to approximate the probability density function for both diabetic and non-diabetic subjects, utilizing a standard Gaussian density function as the kernel. The bandwidth was selected as 0.3 times the standard deviation of the sample. Subsequently, to gauge the resemblance between distribution functions, we computed the Kolmogorov-Smirnov statistics. This nonparametric test was applied directly between the samples.

In Figure S2, we present a statistical analysis of the general participants dataset. The main diagonal Figures S2A, S2E, and S2I), showcases kernel density estimations for each of the three studied variables (HbA_{1c}, FPG, and HbA_{1d}), distinguishing between 23,866 diabetic and the 16,786 non-diabetic cases (depicted in orange and blue, respectively). The three variables under study provide a degree of differentiation between non-diabetic and diabetic subjects, as the density estimations differ significantly in all instances. To quantify these differences, we performed Kolmogorov-Smirnov tests, with the results, denoted as D_{KS} in the figure, substantially deviating from zero – a value representing perfect similarity between diabetic and non-diabetic distributions. These results are significant at the 99.9% confidence level. Figures S2B, S2C, and S2F, illustrate density histograms for various variable pairs. Figures S2D, S2G, and S2H display the correlations between variables. They are expressed in terms of coefficients of determination, r^2 (the square of the Pearson correlation coefficient). These coefficients were computed via linear regression, with a black straight line indicating the fitted function. Consistent with the color code employed in the main text, non-diabetic participants are represented in blue, while diabetic patients appear in orange.

Density histograms

Density histograms were compiled to emphasize the data concentration in scatterplots of adult participants. We adopted the Freedman-Diaconis rule to determine the number of bins. This rule sets the bin width as twice the interquartile range divided by the cube root of the sample size.

Mathematical model

In Box 1, we offer an overview of the key features of the mathematical model developed in this study. We incorporate two distinct labile hemoglobin forms (termed as *fast* and *slow*), each having independent reaction, disassociation and glycation rates. This introduces a new coupling, as the free hemoglobin now depends on these two complexes. The ensuing system of three interlinked ordinary differential equations incorporates both fast and slow types of labile hemoglobin HbA_{1d}^{fast} and HbA_{1d}^{slow}, together with HbA_{1c}. This model was partly motivated by the experimental results that lend support to this hypothesis.[24] To arrive at the final ordinary differential equations, the free hemoglobin could be exactly removed from the initial set as the total amount of hemoglobin inside an RBC is approximately constant throughout its lifetime. By doing so, we coupled both HbA_{1d} equations taking into account the rate at which both the fast and slow types transform into HbA_{1c} (see Box 1).

Model derivation

For the formulation of the model in Box 1, we expanded the theoretical framework presented in[21] to formulate a system of partial differential equations having an age-structure. The model in[21] was never tested against time-dependent data corresponding to each participant, potentially rendering it unreliable for short-time (occurring in the range of minutes to few hours) changes in labile hemoglobin levels. Here we incorporate two types of labile hemoglobin radicals, which were reported experimentally in,[24] to account for both short-time and long-time (in the range of weeks to a few months) dynamics.

Let $G(t)$ represent the glucose concentration in blood plasma. Assuming that the RBCs are subjected to a spatially uniform bath of free glucose, the system of partial differential equations capturing the interplay of $G(t)$ with the concentrations of $\mathcal{C}_F$, $\mathcal{C}_{L_1}$, $\mathcal{C}_{L_2}$, and $\mathcal{C}_S$, denoting the free, fast-labile, slow-labile and glycated hemoglobin, respectively, reads as:

$$\frac{\partial \mathcal{C}_F}{\partial t}+\frac{\partial \mathcal{C}_F}{\partial a}= -\left(k_{FL_1}+k_{FL_2}\right)G(t)\mathcal{C}_F+k_{LF_1}\mathcal{C}_{L_1}+k_{LF_2}\mathcal{C}_{L_2}-\frac{\mathcal{C}_F}{\tau(a)}, \qquad \text{(Equation 4)}$$

$$\frac{\partial C_{L_1}}{\partial t}+\frac{\partial C_{L_1}}{\partial a}=k_{FL_1}G(t)C_F-k_{LF_1}C_{L_1}-k_{LS_1}C_{L_1}-\frac{C_{L_1}}{\tau(a)}, \quad \text{(Equation 5)}$$

$$\frac{\partial C_{L_2}}{\partial t}+\frac{\partial C_{L_2}}{\partial a}=k_{FL_2}G(t)C_F-k_{LF_2}C_{L_2}-k_{LS_2}C_{L_2}-\frac{C_{L_2}}{\tau(a)}, \quad \text{(Equation 6)}$$

$$\frac{\partial C_S}{\partial t}+\frac{\partial C_S}{\partial a}=k_{LS_1}C_{L_1}+k_{LS_2}C_{L_2}-\frac{C_S}{\tau(a)} \quad \text{(Equation 7)}$$

with a being the age of a red blood cell (RBC), and where the parameters represent.

- k_{FL_1} and k_{FL_2} are the rate constants of free hemoglobin interacting with glucose to create HbA_{1d}^{fast} and HbA_{1d}^{slow}, respectively.
- k_{LF_1} and k_{LF_2} are the rate constants of HbA_{1d}^{fast} and HbA_{1d}^{slow} disassociating back into free hemoglobin and glucose, respectively.
- k_{LS_1} and k_{LS_2} are the rate contant of HbA_{1d}^{fast} and HbA_{1d}^{slow} turning into HbA_{1c}.
- $\tau(a)$ is the death time of RBCs, which generally depends on a.

The meaning of each of the terms entering into the above system is the following.

- In 4, the first term on the right-hand side indicates the interaction of free hemoglobin and glucose to produce HbA_{1d}. The second term corresponds to the reverse reaction and the third term takes into account the natural death of RBCs.
- In both (5) and (6), the first term on the right-hand side accounts for the production of HbA_{1d}^{fast} and HbA_{1d}^{slow}, which comes from Equation 4. The second term corresponds to the rate of HbA_{1d}^{fast} and HbA_{1d}^{slow} going back to free hemoglobin and the third term takes into account the Amadori rearrangement which makes glycation stable. The last term represents the disappearance of the labile forms due to the death of RBCs.
- Finally, in the fourth equation, the first two terms take into account the Amadori rearrangement of the labile forms to become HbA_{1c}. The third term captures the disappearance of the glycated hemoglobin due to the death of RBCs.

These equations can be integrated over the age a of the RBCs, with $a \in [0,\infty)$, to obtain a time-dependent version of the previous system, regardless of RBC age. Using the mean value theorem, the rates at which RBCs loose free, glycated, and labile hemoglobins can be replaced by constant rates, although we approximate all four using the half-life of RBCs, as it should be the primary factor contributing to their decay. In addition, two facts were considered to reduce the dimension of the system. The first one is that total hemoglobin tends to be uniform over the lifetime of an RBC. The second is that it is reasonable to assume that RBCs start their life in the bloodstream with 100% free hemoglobin (although this is probably an approximation and further investigation is deserved). Therefore, using a similar derivation procedure as the one provided in,[21] we arrive at the following set of ordinary differential equations (expressed now as percentages of the total hemoglobin)

$$\frac{dh_{L_1}}{dt}=k_{FL_1}G(t)\left(1-h_{L_1}-h_{L_2}-h_S\right)-k_{LF_1}h_{L_1}-k_{LS_1}h_{L_1}-\frac{h_{L_1}}{\tau_{RBC}}, \quad \text{(Equation 8)}$$

$$\frac{dh_{L_2}}{dt}=k_{FL_2}G(t)\left(1-h_{L_1}-h_{L_2}-h_S\right)-k_{LF_2}h_{L_2}-k_{LS_2}h_{L_2}-\frac{h_{L_2}}{\tau_{RBC}}, \quad \text{(Equation 9)}$$

$$\frac{dh_S}{dt}=k_{LS_1}h_{L_1}+k_{LS_2}h_{L_2}-\frac{h_S}{\tau_{RBC}}. \quad \text{(Equation 10)}$$

Here, h_{L_1}, h_{L_2}, and h_S denote the percentages of HbA_{1d}^{fast}, HbA_{1d}^{slow}, and HbA_{1c}, respectively. We model the dataset of prospective pediatric subjects with these equations, together with the initial conditions $h_{L_1}(0)$, $h_{L_2}(0)$ and $h_S(0)$. The left-hand-side of the equations represents the rate at which each of the intervening quantities varies, whereas the right-hand side accounts for the different mechanisms that contribute to these changes. This is the model summarized in Box 1.

Numerical solution of the system of differential equations

To solve the system of ordinary differential equations (ODEs) given by 1-3 (see Box 1), it was necessary to first estimate the time-dependence of the glucose function entering into the model ODEs (1) and (2). To accomplish this, the glucose profile was interpolated from the measured pediatric dataset for each subject. This was carried out by minimizing, via an active-set method, the squared difference between the simulated and measured values at the corresponding timepoints for each subject. As indicated in Table 1, the subjects were 2–14 years old and had different medical conditions (see Table 3). We then employed a numerical scheme, based on the Runge-Kutta method, specifically the ode45 solver in MATLAB. Parameter fitting was performed via nonlinear regression using the MATLAB function fmincon, available from the optimization toolbox. In particular, we utilized an active-set algorithm with a function tolerance of 10^{-12}. We minimized the square sum of differences between the HbA_{1d} and HbA_{1c} values computed by our model and the measured values, achieved via a standard least squares method. Some terms were standardized across all subjects, while others were tailored to each individual.

In Table S3, we show the three calculated kinetic parameters using said nonlinear least square fitting for the mathematical model to the values of the kinetic constants for each of the individual pediatric subjects from which we had five HbA_{1d}, glucose and HbA_{1c} data points. We

chose to adjust three variables to each patient as it was the number that minimized the Akaike Information Criterion. The three variables were selected based on their significant impact on improving the model's performance when individually adjusted for each subject. This selection process involved an iterative testing, wherein only one variable was adjusted at a time while keeping the other nine variables fixed. The variables that resulted in the least amount of error during this process were ultimately employed in the combined model. Furthermore, to check that our model gives a good overall fit to all subjects and to avoid overfitting, we obtained the parameters that would best fit all of the pediatric subjects simultaneously. We did this both for the whole group and for two subsets of 42 optimization and 18 testing subjects. Both groups had a similar gender ratio of 57%. These values are displayed in Table S4. The relative errors for each of these sets were approximately equal to 0.23% (in the three of them).

To avoid overfitting, the less sensitive parameters were fixed using the aforementioned "common" parameters for all subjects. This approach involved an iterative calculation of the squared total error by fixing all parameters except one. This provided an estimation of the combined importance and patient variability of each kinetic parameter. Accordingly, we could fix each parameter based on the criteria mentioned previously. This one-factor-at-a-time (OFAT) sensitivity analysis enabled us to evaluate the effect of each parameter on the model and its variability across the subject population.[48] This provided us with a hierarchy of parameters, and we eliminated them one by one until we identified the optimal number of parameters. The kinetic parameters are listed in the first seven columns of Table S5. The last four columns detail the percentage of initial fast HbA_{1d} ($h_{I1}(0)$%), the error in initial HbA_{1d} and HbA_{1c} ($\Delta h_I(0)$, $\Delta h_g(0)$), and the sum of squared total errors (ϵ), respectively.

The combination of population and subject-specific parameters is not rare,[49] and it allows us to reproduce relatively small perturbations in the labile curves of individual subjects. Moreover, the utility of this model is evident, given that diabetes is a chronic condition subject to regular monitoring. This continuous monitoring implies a potentially substantial number of available data points for a given patient, further supporting the effectiveness and practicality of our approach.

To assess the consistency of our results, additional parameters were obtained for the general subject datasets as presented in Table S5. Furthermore, results were obtained for interpolated continuous glucose monitoring data from three patients Figure S5.

Sensitivity and stability analysis

We conducted a sensitivity analysis of the model's parameters and confirmed the stability of the system's solutions through the use of Floquet multipliers. Sensitivity analysis involved the computation of Sobol' indices for a range of each parameter involved, wherein we varied each constant from 50% to 200% of its average value and determined the corresponding response over a 24-h range for each of the three outcomes. To ascertain the stability of the solutions, that is, ensuring that the solution always tends toward a periodic solution for a periodic non-homogeneous function (the case of the glucose profile), we obtained the numerical solution over a range of sinusoidal functions for orthogonal initial conditions, corresponding to one period of the equivalent homogeneous equations. Figure S6 shows different graphs of the sensitivity analysis computed for fast HbA_{1d}, slow HbA_{1d}, and HbA_{1c}.

The most important parameter entering into ordinary differential equation for each variable is.

- k_{LF_1}: Fast HbA_{1d} conversion rate to free hemoglobin for fast HbA_{1d} [see Equation 8].
- k_{LF_2}: Slow HbA_{1d} conversion rate to free hemoglobin for slow HbA_{1d} [see Equation 9].
- τ_{RBC}: Red blood cell death rate for HbA_{1c} [see Equation 10].

In order to check the stability of the solutions of the system of differential equations, we calculate the eigenvalues of the monodromy matrix. All the values obtained were less than one: $\lambda_1 = 0.2123$, $\lambda_2 = 0.9810$, $\lambda_3 = 2.3645 \times 10^{-6}$ where these values were the maxima of each one for a range of reasonable sinusoidal glucose functions, confirming the asymptotic stability of the solutions.[50] Thus, from these results, we can also infer that the system of differential equations will only have one periodic solution, which will be asymptotically stable according to Floquet theory.[50] We carried out these model simulations and data analyses using MATLAB (R2023a, The MathWorks, Inc., Natick, MA, USA).

QUANTIFICATION AND STATISTICAL ANALYSIS

Statistical analysis of linear correlation of difference of labile hemoglobin and glucose levels in pediatric subjects

Two linear estimates were performed on empirical and simulated data as shown in Figures 3A and 3B, where N = 240 in both cases. These estimates represent the difference between subsequent measurements of HbA1d and glucose in all 60 pediatric subjects in A, and their simulated HbA1d and measured glucose in B. Both estimates were deemed significant through an F-test on the regression model, which tests whether our model fits the data notably better than a degenerate model adjusting only a constant term. This analysis was conducted using MATLAB (R2023a) and the function fitlm. Additionally, the coefficient of determination (r^2), which provides an estimate of the variance of the data predicted by our linear fit, obtained with the same function, was also deemed sufficient.

Statistical analysis in logistic regression

In Figure 2, three logistic regressions were performed on 32,521 subjects from the general dataset, constituting the training cohort. HbA1c was measured in C, FPG in D, and HbA1d in E. Their corresponding McFadden likelihood ratio index (R^2_{McF}), which is a type of pseudo-R squared where a value of 1 indicates a perfect fit and 0 indicates a completely random fit (though a value between 0.2 and 0.4 represents a good fit, rather than the higher values you tend to see in linear estimates), was calculated. A Wald test was also used to check whether

the slopes are significantly different from zero, indicating that the values of these measurements are statistically associated with a higher chance of being diabetic.

In Tables S1 and S2, a similar approach was employed, but now all three measurements were included in the model. Additionally, all slopes and the intercept were found to be significant. These analyses were performed using R.

Kolmogorov-Smirnov tests and regression of general subject cohort

In Figure S1 three Kolmogorov-Smirnov tests were performed to confirm that HbA1c, FPG, and HbA1d measurements were significantly different in diabetic and non-diabetic patients (the hypothesis of both sets coming from the same distribution was rejected). The diabetic population consisted of 23,866 individuals, while the non-diabetic population consisted of 16,786 individuals. This is illustrated in subfigures A, E, and I. Additionally, three linear regressions were performed for each pair of values, as depicted in subfigures D, G, and H, alongside their r^2 values. These analyses were also performed using R.

ADDITIONAL RESOURCES

The adult cohort data collection was registered in the Biomedical Research Ethics Portal of Andalusia with an ID of A1d14.

(https://www.juntadeandalucia.es/salud/portaldeetica/). The pediatric cohort data collection was registered under code DIAMATH no. 44.23 in (https://www.juntadeandalucia.es/salud/portaldeetica/).