

## MIT Open Access Articles

*Heterogeneity of continuous glucose monitoring features and their clinical associations in a type 2 diabetes population*

The MIT Faculty has made this article openly available. **Please share** how this access benefits you. Your story matters.

**Citation:** Healey E, Morato C, Murillo J, Kohane I. Heterogeneity of continuous glucose monitoring features and their clinical associations in a type 2 diabetes population. *Diabetes Obes Metab.* 2025; 27(7): 3957-3966.

**As Published:** <https://doi.org/10.1111/dom.16432>

**Publisher:** Wiley

**Persistent URL:** <https://hdl.handle.net/1721.1/162872>

**Version:** Final published version: final published article, as it appeared in a journal, conference proceedings, or other formally published context

**Terms of use:** Creative Commons Attribution-NonCommercial-NoDerivatives



# Heterogeneity of continuous glucose monitoring features and their clinical associations in a type 2 diabetes population

Elizabeth Healey BS<sup>1,2</sup>  | Carlos Morato PhD<sup>3</sup> | Jaime Murillo MD<sup>3</sup> | Isaac Kohane MD<sup>2</sup>

<sup>1</sup>Harvard-MIT Health Sciences and Technology, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA

<sup>2</sup>Harvard Medical School, Boston, Massachusetts, USA

<sup>3</sup>UnitedHealth Group, Minneapolis, Minnesota, USA

## Correspondence

Elizabeth Healey, Harvard-MIT Health Sciences and Technology, Massachusetts Institute of Technology, Cambridge, MA, USA.  
Email: [ehaley@mit.edu](mailto:ehaley@mit.edu)

## Funding information

National Science Foundation Graduate Research Fellowship Program, Grant/Award Number: 2141064

## Abstract

**Objective:** Data from continuous glucose monitors (CGM) enable the extraction of features descriptive of glycemic dynamics that may provide insight into underlying health status. In this work, we analyse CGM data from a large population of individuals with type 2 diabetes (T2D) and study the association of features with clinical covariates.

**Methods:** We retrospectively analysed CGM and electronic health record data from a large population of individuals with T2D. We extracted 25 daily CGM features for each individual over a 30-day period and performed statistical association tests on the features and clinical findings from medical claims data and laboratory records.

**Results:** Our final analysis was performed on 6533 individuals. When clustering the CGM features across the population of individuals with T2D, four distinct clusters of features emerged. Further, the CGM features had heterogeneous discriminatory power with clinical covariates, including laboratory values and the presence of claims for diabetic complications. Features related to glycemic variability, such as coefficient of variation, showed markedly lower *p*-values in many association tests for the presence of diabetic complications than mean glucose.

**Conclusions:** In examining the characteristics of different features extracted from CGM data in a large population of individuals with T2D, we found that the features were heterogeneously associated with different clinical comorbidities related to diabetes. This work motivates further research to investigate the relationship between CGM features and health outcomes in T2D to enable precision medicine.

## KEYWORDS

cohort study, continuous glucose monitoring, database research, diabetes complications, type 2 diabetes

## 1 | INTRODUCTION

Type 2 diabetes (T2D) is a chronic metabolic disease that affects millions of people around the world.<sup>1</sup> There has been recent interest in

better understanding disease phenotypes and heterogeneity within T2D to enable precision medicine.<sup>2</sup> Continuous glucose monitoring (CGM) technology has enabled the analysis of daily interstitial glucose readings, which gives insight into glycemic variability. The availability

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial-NoDerivs](https://creativecommons.org/licenses/by-nc-nd/4.0/) License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2025 The Author(s). *Diabetes, Obesity and Metabolism* published by John Wiley & Sons Ltd.

of these data presents an opportunity to study how CGM can be used to better understand heterogeneity in T2D and how CGM can be used to advance precision medicine.

CGM use among individuals with T2D has grown over the years, particularly due to its effectiveness in encouraging behaviour change.<sup>3</sup> It has been demonstrated that CGM is a useful tool to help patients with T2D manage their disease.<sup>4–6</sup> Further, retrospective analyses of CGM profiles of individuals without diabetes and with T2D have shown significant heterogeneity in glucose profiles.<sup>7</sup> Metrics derived from CGM offer insight into glycemic dynamics that conventional metrics, such as haemoglobin A1c (HbA1c), do not fully capture.<sup>8</sup> There has been recent interest in understanding whether these metrics provide insight into disease pathophysiology and heterogeneity, and whether they are useful for risk stratification and guiding treatment decisions.<sup>9</sup>

There has been recent interest in understanding the association of glycemic variability derived from CGM with clinical outcomes.<sup>10</sup> Many CGM features have been derived by looking at variability in short-, medium-, and long-term periods.<sup>11</sup> Short-term variability has been measured from CGM using intraday metrics such as the coefficient of variation and a measurement known as the mean average glucose excursion (MAGE).<sup>12</sup> Recent studies have identified over 30 indices of glycemic variability from CGM signals, representing both short-term and intermediate-term variability, that capture statistical properties, including information on peaks and time spent in either hypoglycemia or hyperglycemia.<sup>13,14</sup> While most studies have focused on extracting features from CGM in the time domain, there has been some interest in studying the frequency domain of the CGM signal.<sup>15,16</sup> One study used features derived from the frequency domain of the CGM signal and found that frequency domain features were useful in discriminating diabetes status.<sup>16</sup> The frequency domain is potentially useful in characterising a signal because it can give insight into the frequency of oscillations, which can be thought of as a dimension of glycemic variability.<sup>17</sup>

There is a need to better understand which CGM features are informative and clinically meaningful, particularly in individuals with T2D. There has been recent interest in understanding heterogeneity in CGM features in populations with and without diabetes. One study characterised CGM data in over 7000 individuals without diabetes and found that CGM-derived features were associated with clinical parameters.<sup>18</sup> Other studies have looked at the heterogeneity of CGM characteristics in hospitalised individuals with T2D and found subgroups of individuals based on CGM metrics with associations to clinical covariates.<sup>19,20</sup>

Despite widespread interest in both understanding disease heterogeneity in T2D and glycemic variability, there exist few large-scale studies examining CGM characteristics from T2D populations. In this retrospective observational study, we perform two primary analyses to better understand the discriminative power of features derived from CGM. First, we show the heterogeneity of common features derived from CGM in a large cohort of individuals with T2D. In this analysis, we look at how CGM features correlate with one another across different individuals, and we identify subgroups of CGM

features. We then perform statistical association tests of CGM features and diabetic complications and laboratory values that occur before, during, or after the period of collected CGM in a large claims database. We intentionally analyse many CGM features descriptive of glycemic variability in this work so that our results can serve as a reference for future work investigating CGM features in T2D populations.

## 2 | RESEARCH DESIGN AND METHODS

### 2.1 | Cohort

In this retrospective study, we investigated how features derived from daily CGM data were associated with clinical outcomes in a large cohort of individuals with T2D. We used a de-identified dataset consisting of 12 265 individuals with T2D who were enrolled in a commercial diabetes program where they were given a CGM. We included individuals in our analysis based on information available in their medical claims and the amount of complete CGM data that they had. We first ensured that each individual had at least one documented claim for T2D as determined by an International Classification of Diseases, Tenth Revision, Clinical Modification (ICD-10)<sup>21</sup> code starting with E.11. To best ensure that individuals in the cohort had T2D and not T1D, we added an additional exclusion criterion. We excluded individuals who had claims for a T1D diagnosis (identified by ICD-10 codes beginning with E10) on more than five separate days in their history, or whose days with T1D diagnosis claims exceeded 10% of their total days with T2D diagnosis claims. We also excluded individuals based on the amount of complete CGM data they had. We included only individuals who had a 30-day period of CGM data defined as having greater than 22 full days of data within a 30-day window. We defined a full CGM day as a day where there was at least 23 h of data. In a robustness analysis, we also studied the first 7-day period with at least six full days of CGM data of the final cohort, as defined above. Institutional Review Board (IRB) approval was obtained for the research in this paper through the Office of Human Research Affairs at UnitedHealth Group.

### 2.2 | CGM features and clinical covariates

We extracted several types of features from CGM, including basic statistics, features from diabetes literature, and frequency domain features. The full list of CGM features and descriptions is listed in Table 1. We extracted CGM features from each day of the first full 30-day period in their record, as defined previously. We define the first day of the collected CGM period as the “CGM index date”. For individuals with missing values in a day (up to 12 missing data points out of 288), we linearly interpolated the data. Once daily features were computed, we took the mean for each of the variables across all available days in the 30-day period. We extracted basic descriptive features, including mean, standard deviation, minimum, etc., as well as

**TABLE 1** CGM feature definitions are derived for each individual.

Statistics	
Mean	Mean
Std	Standard deviation
cv	Coefficient of variation
Var	Variance
Median	Median CGM value
Min	Minimum daily value
Max	Maximum daily value
Range	Daily range
Diabetes features	
TIR	Per cent time spent between 70 and 180 mg/dL
TAR180	Per cent time spent above 180 mg/dL
TAR250	Per cent time spent above 250 mg/dL
TITR	Per cent time spent in tight range between 70 mg/dL and 140
TBR54	Per cent time spent below 54 mg/dL
TBR70	Per cent time spent below 70 mg/dL
LBGI <sup>a</sup>	Low blood glucose index
HBGI <sup>a</sup>	High blood glucose index
J_index <sup>a</sup>	Measure of both the mean level and variability
MAGE <sup>a</sup>	Mean amplitude of glucose excursions
Frequency domain related features <sup>16</sup>	
max_amplitude	Maximum magnitude of FFT
dominant frequency	Frequency of maximum magnitude of FFT
psd_max_amplitude	Highest magnitude of PSD
psd_dominant frequency	Frequency of maximum value on PSD
bandwidth	PSD 3 dB bandwidth
psd75_frequency	Frequency cutoff on the PSD for 75% of the signal

<sup>a</sup>Computed using `cgmquantify` package.<sup>24</sup>

features found on the Ambulatory Glucose Report (AGP).<sup>22</sup> Several diabetes features, including Low Blood Glucose Index (LBGI), High Blood Glucose Index (HBGI), J-index, and MAGE, which are all known metrics for glycemic variability,<sup>23</sup> were extracted using the `cgmquantify` package (Version 0.5).<sup>24</sup> The frequency domain features extracted were described in Fico et al.<sup>16</sup> These frequency domain features included derived features from the Fast Fourier Transform (FFT) and the power spectrum density (PSD) signal found using Welch's method.<sup>25</sup> Feature extraction was performed in Python (Version 3.8.5) using SciPy (Version 1.10.1).<sup>26</sup>

We looked at two distinct sets of clinical features. We first looked for the presence of claims for four diabetic complications by ICD-10 codes by category. Table 2 gives the ICD-10 category and corresponding codes used, and the abbreviation used throughout this paper. We then extracted scalar laboratory values by Logical

**TABLE 2** Binary complication features derived for each individual. The table shows the abbreviated category name, the ICD-10 category, and the ICD-10 codes included.

Abbreviated name	ICD 10 category	ICD-10
Circulatory complications	Type 2 diabetes mellitus with circulatory complications	Codes starting with E11.5
Kidney complications	Type 2 diabetes mellitus with kidney complications	Codes starting with E11.2
Ophthalmic complications	Type 2 diabetes mellitus with ophthalmic complications	Codes starting with E11.3
Neurological complications	Type 2 diabetes mellitus with neurological complications	Codes starting with E11.4

Observation Identifiers Names and Codes (LOINC codes)<sup>27</sup> for each individual. We included laboratory values for various metabolic markers and markers of diabetic complications, including C-peptide, c-reactive protein (CRP), serum creatinine, HDL cholesterol, LDL cholesterol, albuminuria, and HbA1c. For consistency, we selected a single LOINC code corresponding to each laboratory measurement of interest. Table 3 shows the LOINC codes as well as the abbreviated names used throughout this article. Non-zero laboratory values with documented units of measurement, as listed in Table 3, were extracted using the LOINC codes, and HbA1c values were pre-processed to remove spurious values greater than 20%.

## 2.3 | Study design

In this retrospective study, we looked at the presence of clinical features occurring in different time windows with respect to the CGM start. Since individuals in this study were given a CGM at different points in their disease trajectory, it was important to look at CGM characteristics in the context of the clinical picture occurring before and after the data were recorded. Moreover, in this dataset, there was more data available for individuals prior to the CGM collection period than after the CGM collection period. We extracted clinical features occurring before and after the index date of CGM for each subject at time windows of 30 days, 1 year, and ever recorded. We created features corresponding to the binary presence of an ICD-10 code in a time range. For scalar variables, we took the value closest to the first CGM date for all time intervals.

## 2.4 | Statistical methods

Our statistical analysis consisted of two parts. In the first part, we analysed the correlation among CGM features themselves in the entire population. This allowed us to reduce the feature set by eliminating highly correlated variables. We leveraged a hierarchical clustering

**TABLE 3** Scalar laboratory values derived for each individual. The table shows the abbreviated name, laboratory name, and LOINC code.

Abbreviated name	Lab name	Units	LOINC-cd <sup>27</sup>
HDL cholesterol	Cholesterol in HDL [Mass/volume] in serum or plasma	mg/dL	2085-9
LDL cholesterol	Cholesterol in LDL [Mass/volume] in serum or plasma by calculation	mg/dl	13 457-7
Albumin/creatinine	Albumin/Creatinine [Mass ratio] in urine	mg/g, µg/mg	9318-7
Serum creatinine	Creatinine in serum or plasma	mg/dL	2160-0
C-peptide	C peptide in serum or plasma	ng/mL	1986-9
CRP	C reactive protein in Serum or Plasma	mg/L, µg/mL	1988-5
A1c	Haemoglobin A1c/Haemoglobin total in Blood	%	4548-4

technique, similar to the one employed by Tao et al. in their study reducing CGM features.<sup>19</sup> This allowed us to identify subgroups of CGM features that behaved similarly across the population. We then performed a statistical analysis on the reduced CGM feature set and clinical outcomes.

## 2.5 | Feature clustering and feature reduction

We first computed the correlation matrix of all CGM features across the population to observe which features were highly correlated. We then performed hierarchical clustering on the correlation matrix of the 25 different CGM features. We used Ward's method<sup>28</sup> for linkage and Euclidean distance as the clustering metric. Computations were performed in Python (Version 3.8.5) using SciPy (Version 1.10.1).<sup>26</sup> We reduced the number of CGM features based on the results of hierarchical clustering. We used a distance cutoff of 0.05 on the dendrogram to reduce the feature set for the subsequent part of the analysis.

## 2.6 | Association analysis

For all clinical features related to complications, we performed statistical association tests on each CGM feature comparing the group with the complication to the group without the complication. The data were censored, and different individuals had different lengths of data recorded before and after the CGM was collected. Because of this, for each analysis, we only considered the subpopulation of individuals with documented claims occurring before and after the dates of the window. For the time windows that examined whether a claim was present ever in the individual's history before or after the CGM start date, we ensured that the individual had documented claims more than a year before and after the CGM start date, respectively. We used the Mann–Whitney *U* test<sup>29</sup> for statistical analyses. For each complication, we performed multiple hypothesis corrections for the six different time intervals and the total number of CGM features analysed, which was  $p = 0.00044$ .<sup>30</sup>

We performed a correlation analysis to study the association of each CGM feature with scalar lab values, similar to prior work characterising CGM features in a large population.<sup>18</sup> We computed the Pearson correlation coefficient of CGM features with scalar lab values and reported the correlation coefficient and the *p*-value for the individuals

**TABLE 4** Final cohort characteristics.

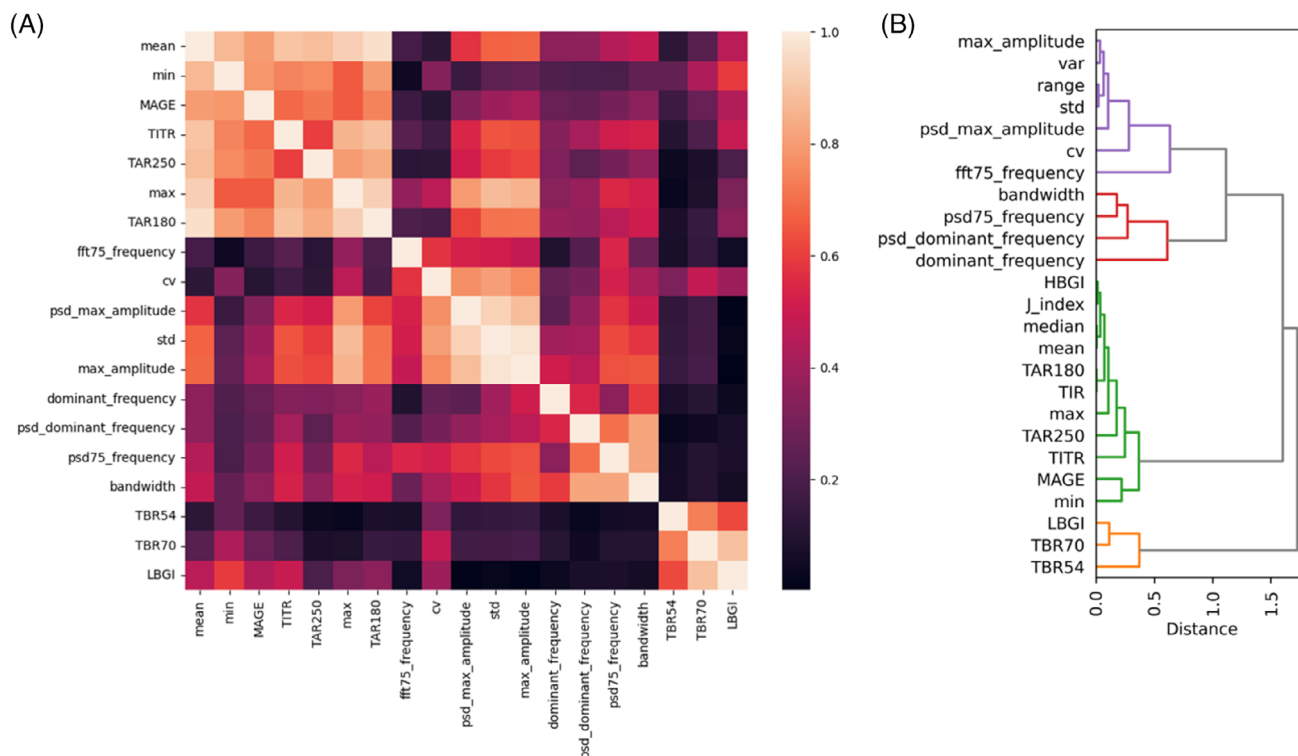
Cohort characteristics			
Cohort size ( <i>n</i> )			6533
Female (%)			45.7%
Average age (years)			53.4 (48–60)
Average number of days of claims before (days)			1228.0 (736–1733)
Average number of days of claims after (days)			243.4 (120–328)
Medications in the year prior to CGM index date			
Thiazolidinediones	7.04%	Short-acting insulins	1.32%
Rapid-acting insulins	15.22%	Amylinomimetics	<1%
Long-acting insulins	28.90%	Meglitinides	<1%
Biguanides	69.29%	Alpha-glucosidase inhibitors	<1%
Incretin mimetics	35.44%	Ergot-deriv. dopamine receptor agonists	<1%
Sulfonylureas	28.12%	Antidiabetic agents, miscellaneous	<1%
DPP-4 inhibitors	13.45%	No recorded medication	6.61%
SGLT-2 inhibitors	28.32%	Average no. of medications	2.80
Intermediate-acting insulins	1.27%		

*Note:* Cohort characteristics show the descriptive statistics of the final cohort. The 25th percentile and 75th percentile are shown for the average age and number of claims before and after the CGM index date. The table also shows the per cent of the cohort with prescription for each diabetic medication in the year prior to the CGM index date.

with a lab value in each timeframe. For individuals with multiple lab values, we used the lab value closest to the start date of the CGM window. We only reported the results if there were 10 or more individuals with lab values in a given window.

## 2.7 | Sensitivity analysis

We performed a subgroup analysis where the association tests were performed on subsets of the cohort. We analysed the subset of the cohort with a history of any insulin or sulfonylurea medication



**FIGURE 1** Correlation of CGM Features. (A) Heatmap of the absolute value of the correlation coefficient between CGM features. (B) Dendrogram from hierarchical clustering of the correlation coefficients.

prescription prior to the CGM start date or prescription during the 30-day CGM period. We also analysed the subset with no history of any insulin or sulfonylurea prescription in that time frame.

### 3 | RESULTS

#### 3.1 | Cohort characteristics

In the total cohort of 12 265 individuals, there were 7351 that met the criteria for having sufficient CGM data. We leveraged this cohort for further analyses. Within that cohort, 6766 individuals had linked claims data with at least one claim for T2D, defined as any ICD-10 code starting with E11, and demographic information. Of this population, we identified 595 individuals with an ICD-10 code indicative of T1D (starting with E10). After applying our inclusion criteria, the final cohort consisted of 6533 individuals. Table 4 shows the descriptive statistics of the cohort, including gender ratio, age and average amount of claims history available before and after the CGM period that was analysed. Our final cohort was 45.7% female with an average age of 53.4 years (IQR: 48–60). We also show the breakdown of medications within the year before each individual's CGM index date. Most individuals (93.39%) had documented prescriptions for a T2D drug, with individuals on average having a prescription for 2.8 diabetic drugs in the year prior to CGM start. The most common prescription was biguanides, with 69.29% of the cohort having a prescription, followed by incretin mimetics (35.44%). Additionally, 28.9% of the

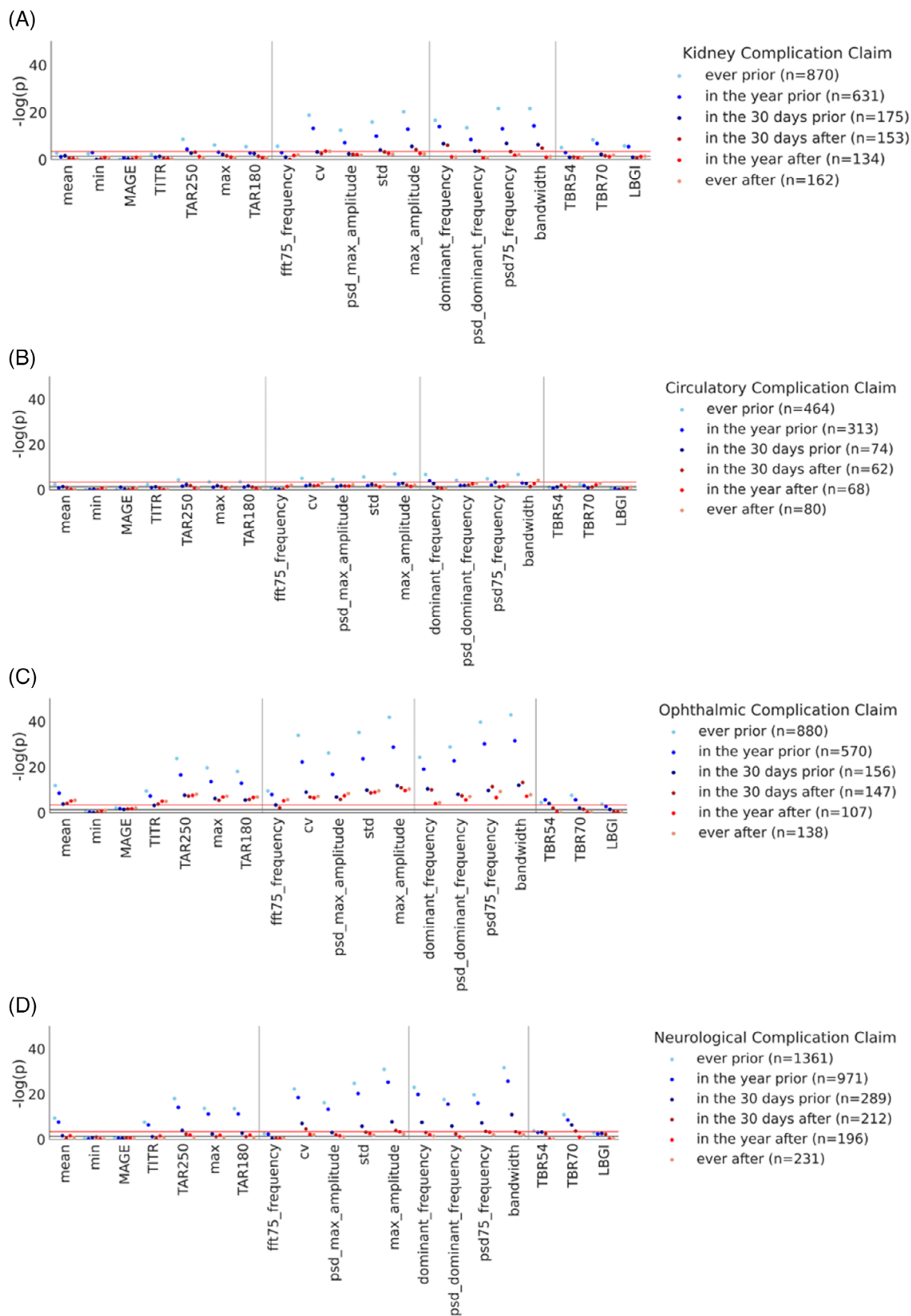
cohort had a prescription for long-acting insulins, and 15.22% of the cohort had a prescription for rapid-acting insulins. Table S1 shows the mean values of the CGM features across the cohort, and Table S2 shows statistics for the laboratory measurements found in the cohort.

#### 3.2 | Heterogeneity of CGM features

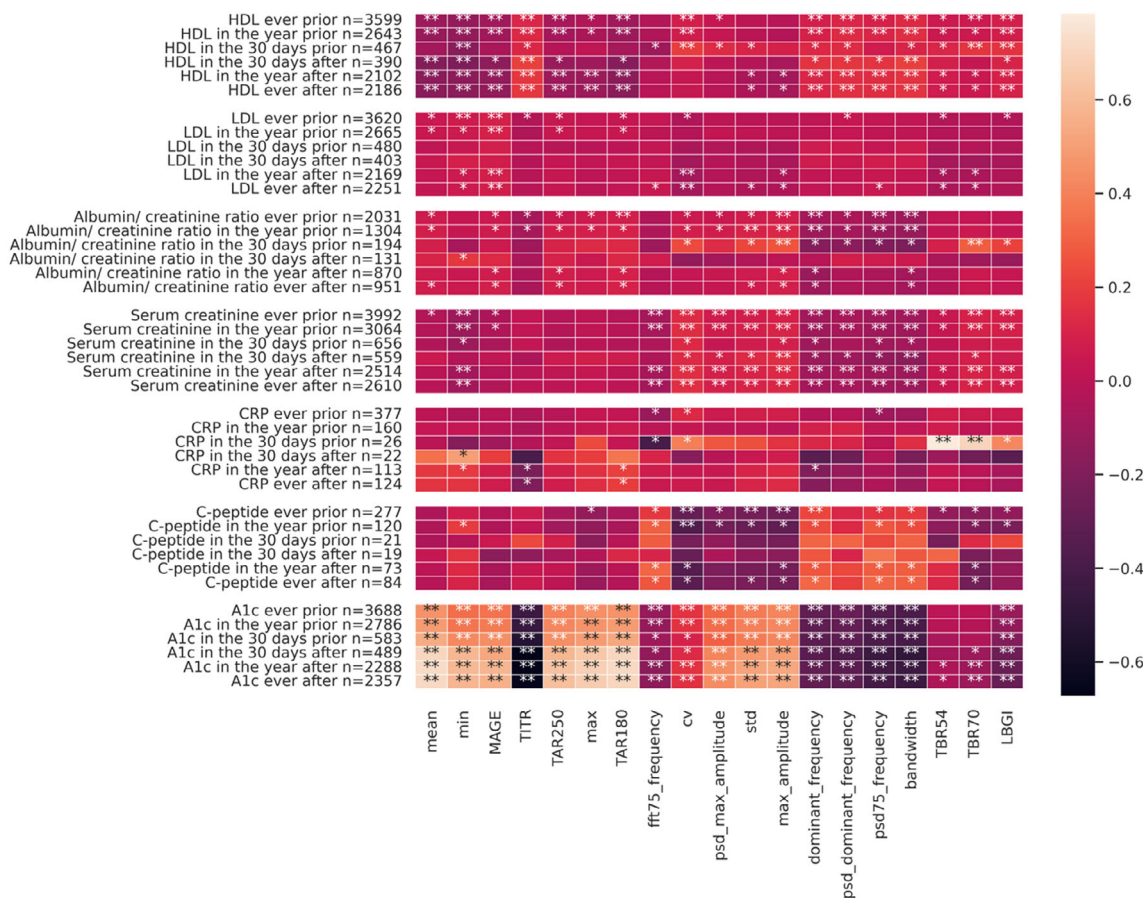
We analysed the correlation matrix of 25 features extracted from 6533 individuals. Figure 1A shows the heatmap of the absolute value of the correlation matrix between the CGM features from the reduced set of 19 features. Figure 1B shows the outcome of hierarchical clustering on the features. We identified four distinct feature clusters that emerged: a variability cluster including features such as coefficient of variation and features with magnitudes from the frequency domain, a magnitude-related cluster including mean glucose, time in range, and maximum glucose, a cluster with features from the frequency domain, and a cluster related to low blood glucose. Figure S1 shows the results of a robustness analysis analysing CGM features across a smaller time window for all individuals, which shows that the same four clusters emerge. The raw correlation values for both CGM analysis windows for all 25 features are shown in Figure S2.

#### 3.3 | Association analysis

We performed an association analysis on the reduced feature set of 19 CGM features with ICD-10 codes from four different diabetic



**FIGURE 2** Association of CGM features and clinical characteristics.  $p$ -values of association of CGM features with ICD-10 claims of diabetic complications occurring in six different time windows. The yellow line is the adjusted  $p$ -value at 0.00044 and the black line is the  $p$ -value at 0.05 for the (A) presence of diabetic kidney complications, (B) the presence of diabetic circulatory complications, (C) the presence of diabetic ophthalmic complications, and (D) the presence of neurological complications. The number of individuals with a claim in each time window is noted as  $n$ . The black line indicates a  $p$ -value  $< 0.05$  and the red line indicates a corrected  $p$ -value.



**FIGURE 3** Association of CGM features and laboratory values. Heatmap of the correlation coefficient of each CGM feature with each laboratory value by time window. The number of samples included in the analysis is indicated by  $n$ . \*Indicates a  $p$ -value  $< 0.05$ , \*\* indicates a corrected  $p$ -value.

complications. Figure 2A–D shows the  $p$ -values for each CGM feature for a diabetic complication occurring within different time windows. The plots show two different cutoffs of statistical significance. We include vertical lines separating CGM features that correspond to different subgroups identified in the previous section. The raw results with the mean values for each group are in Tables S3 and S4.

For all complications, many CGM features emerged with statistically significant associations with the complication at multiple time windows. Notably, for certain complications, such as kidney complications and circulatory complications, more CGM features from the variability subgroup emerged with very low  $p$ -values when compared to mean glucose. For example, in the year prior to the CGM index date, mean glucose in individuals with a claim for kidney complications was 156.85 mg/dL (SD = 33.77), and the mean glucose in individuals without a claim for kidney complications was 154.73 mg/dL (SD = 34.51) (no significance). The mean coefficient of variation for these groups in the same time period was 21.5% (SD = 5) and 19.9% (SD = 5) ( $p < 0.00044$ ), respectively. Interestingly, the frequency domain features often had strong associations with the presence of diabetic complication claims, particularly for ophthalmic, neurological, and kidney complications. Of note, there were fewer individuals included in the tests for the time periods occurring after the CGM index date.

The counts for the number of individuals with claims ever before the CGM index date were 870, 464, 880, and 1361 for kidney, circulatory, ophthalmic, and neurological complications, respectively, and the counts for claims ever recorded after the CGM index date were 162, 80, 138, and 231, respectively. This is partially due to the fact that the total number of individuals with available claims data was smaller after the CGM data was collected. The total number of individuals included in the association test for each time window are in Tables S3 and S4.

Figure 3 shows the results of the correlation analysis with CGM features and lab values, with raw values included in Figure S3. The heatmap shows the correlation coefficient with the laboratory value at different time intervals. In the figure, it is noted how many laboratory values were included in the statistical test. We show two different levels of statistical significance: at 0.05 and the corrected  $p$ -value for multiple hypothesis testing.

Notably, different clusters of CGM features emerged with statistical significance to different laboratory values. The laboratory values for HDL cholesterol had statistical significance at multiple time windows with features such as mean, min, and time in tight range, with correlation coefficients of  $-0.18$  ( $p < 0.00044$ ),  $-0.2$  ( $p < 0.00044$ ), and  $0.2$  ( $p < 0.00044$ ) for these variables in the time frame the CGM

data was studied (the 30-day window after the index date). Serum creatinine values did not have a statistically significant relationship with mean glucose for all time windows except the window that looked at any laboratory ever recorded previously, which had a correlation coefficient of  $-0.032$  ( $p = 0.04$ ). However, serum creatinine values had a significant relationship with features of variability such as coefficient of variation (correlation coefficient of  $0.13$ ,  $p = 0.0014$  in the 30-day window after index date) and features from the frequency domain such as the max amplitude (correlation coefficient of  $0.15$ ,  $p < 0.00044$  in the 30-day after index date window). Expectedly, mean glucose was highly correlated to HbA1c, with a correlation coefficient of  $0.71$  ( $p < 0.00044$ ) in the time window of 30 days after the CGM index date. C-peptide laboratory values had large correlation coefficient magnitudes, particularly with coefficient of variation (correlation coefficient of  $-0.34$ ,  $p = 0.003$  in the year after the CGM index date). In Figure S7, we show the associations of the CGM features with age, finding that age is negatively associated with mean glucose (correlation coefficient of  $-0.13$ ,  $p < 0.00044$ ) and positively associated with coefficient of variation (correlation coefficient of  $0.088$ ,  $p < 0.00044$ ).

The subgroup analyses revealed different associations depending on medication status. Figures S4–S6 show the results from the subgroup analyses comparing two groups: individuals with a history of insulin or sulfonylurea prescription and individuals without a history of prescription. Notably, the significance of the discriminative power of many covariates in the binary association decreased (Figure S4), though it remained statistically significant for many features for both subgroups. For some covariates, the magnitude and strength of the relationship differed depending on medication status. For example, the coefficient of variation was negatively associated with C-peptide values in the year after the index date (correlation coefficient  $-0.36$ ,  $p$ -value  $< 0.05$ ) for the group with a history of insulin or sulfonylurea use but positively associated with C-peptide values in the year after for the individuals without a history of insulin or sulfonylurea use (correlation coefficient  $0.13$ ,  $p$ -value not significant).

## 4 | DISCUSSION

In this study, we demonstrate that CGM features have heterogeneous discriminative power in a large population of individuals with T2D with several key insights. The results of this study report the associations between CGM features and clinical observations in a large population of individuals with T2D undergoing different treatment regimens. One of the primary advantages of this study was the size of the cohort and the quantity of CGM features analysed. There are limited large-scale studies analysing CGM data from individuals with T2D, and many of the existing studies have focused on a narrower set of features. Notably, our study included features that characterise the frequency domain of the CGM signal, which are largely understudied in the literature in large cohort studies. Our results highlight the potential clinical utility of analysing different features extracted from CGM as a way to understand heterogeneity in disease status. As T2D treatment becomes

increasingly more personalised to the individual,<sup>31</sup> there is a need to understand how to best leverage the data to improve precision medicine. Our characterisation of the associations of CGM features with different clinical findings paves the way for future design of risk stratification methods and phenotyping algorithms to leverage CGM to improve decision making for individuals with T2D.

In the first part of this work, we identified distinct clusters of CGM features derived from a large set of CGM features. We identified significant heterogeneity in the features themselves and significant redundancy. This is consistent with the findings from similar retrospective analyses of CGM data from individuals with T2D.<sup>9,19</sup> Of note, our study included analysis of metrics derived from the frequency domain that are not commonly used by investigators but have been shown to provide a meaningful characterisation of glucose variability.<sup>16</sup> We found that these features had strong associations with clinical covariates, suggesting these metrics should be included in analyses of glycemic variability. Additionally, we found that many CGM features had associations with clinical observations derived from claims data and laboratory values that differed in strength from one another, suggesting there are metric-specific associations. This was seen in the laboratory analysis, where HDL cholesterol and serum creatinine were associated with different clusters of CGM features. This heterogeneity is consistent with recent work showing associations between CGM features and clinical findings.<sup>18,19</sup>

Notably, we found that the association of these features with clinical covariates was different when the analysis was performed on subgroups of the population with a recent history of insulin or sulfonylurea therapy. These medications are known to be associated with increased glycemic variability,<sup>32,33</sup> and it is possible that some of the associations observed in this study were capturing medication status. Prior work has assessed how the associations of CGM metrics and clinical covariates in individuals with T2D vary based on medication status, with one study showing an inverse relationship between C-peptide and coefficient of variation in individuals on insulin therapy,<sup>34</sup> similar to our findings. Nonetheless, there has been interest in understanding the link between different metrics of glycemic variability and underlying pathophysiology.<sup>35</sup> Glycemic variability has been thought to play a role in both macrovascular and microvascular complications through endothelial damage from oxidative stress, inflammatory cytokines, and epigenetic changes.<sup>36</sup> Prior studies have investigated how metrics descriptive of glycemic variability are tied to diabetic complications, linking variability to oxidative stress,<sup>37–39</sup> autonomic neuropathy,<sup>40</sup> diabetic retinopathy<sup>41</sup> and cardiovascular disease.<sup>42,43</sup> Our findings suggest that there are many metrics of glycemic variability that can be extracted from CGM data with the potential to assist in understanding the relationship between variability and pathophysiology. The heterogeneity observed in this study should promote future prospective studies collecting CGM data from individuals with T2D with well-defined inclusion criteria to better understand the link between glycemic variability, medication status, and pathophysiology.

There were several limitations of this study and considerations for the interpretations of the results. The CGM data were unblinded to individuals, and the CGM data may not be representative of the

individuals' glycemic dynamics prior to CGM use. While one of the advantages of this study was that it included a large number of individuals with T2D, the study is limited by the fact that this cohort may not be representative of the true population of individuals with T2D. This is made apparent through examination of the medication rates in this population, particularly those on insulin, which is larger than the true percentage of individuals with T2D on insulin.<sup>44</sup> Additionally, the data contained limited demographic information about the cohort, including an omission of information regarding race. Accordingly, these data may be generalisable and may not be representative of the true T2D population.

Lastly, due to the nature of observational data and medical claims, we would like to highlight several limitations of the study. We performed association tests between CGM features and the presence of ICD-10 codes for four different T2D complications. However, ICD-10 codes are imperfect, and it is possible that for many individuals, the ICD-10 codes were improperly coded or misclassified by complication type. The results should be interpreted with this limitation in mind, particularly when interpreting the different associations between the complication types. Because this study was observational, many individuals did not have laboratory records for the clinical laboratory values that this study examined. This introduced bias into the analysis, and this bias should be considered when interpreting the results.

Our results support existing literature demonstrating an association of glycemic variability and clinical covariates and suggest that CGM features hold heterogeneous discriminative value and advanced CGM analysis may offer advances in precision medicine for T2D.

#### AUTHOR CONTRIBUTIONS

EH and CM had full access to the data in the study. EH and IK conceptualised the study. EH, CM, JM, and IK contributed to the analysis. EH curated the dataset and executed the analysis and is the guarantor of this work and, as such, had full access to all the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis. All authors read and approved the final manuscript. This paper uses a proprietary dataset with individual-level information that cannot be released to the public. We summarise the dataset in the paper, and the data may be available for research collaborations.

#### ACKNOWLEDGEMENTS

E.H. is supported by the National Science Foundation Graduate Research Fellowship Program under grant number 2141064. CM and JM were employed with UnitedHealth Group during the time of this study. Open Access funding enabled and organized by MIT Hybrid 2025.

#### CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

#### DATA AVAILABILITY STATEMENT

This paper uses a proprietary dataset with individual-level information that cannot be released to the public. We summarize the dataset in the paper and the data may be available for research collaborations.

#### ORCID

Elizabeth Healey  <https://orcid.org/0000-0002-7307-8429>

#### REFERENCES

1. Khan MAB, Hashim MJ, King JK, Govender RD, Mustafa H, Al KJ. Epidemiology of type 2 diabetes—global burden of disease and forecasted trends. *J Epidemiol Glob Health*. 2020;10:107-111.
2. Ahlqvist E, Storm P, Käräjämäki A, et al. Novel subgroups of adult-onset diabetes and their association with outcomes: a data-driven cluster analysis of six variables. *Lancet Diabetes Endocrinol*. 2018;6:361-369.
3. Kompala T, Neinstein A. A new era: increasing continuous glucose monitoring use in type 2 diabetes. *Am J Manag Care*. 2019;25:SP123-6.
4. Janapala RN, Jayaraj JS, Fathima N, et al. Continuous glucose monitoring versus self-monitoring of blood glucose in type 2 diabetes mellitus: a systematic review with meta-analysis. *Cureus*. 2019;11:e5634. doi:10.7759/cureus.5634
5. Vigersky RA, Fonda SJ, Chellappa M, Walker MS, Ehrhardt NM. Short- and long-term effects of real-time continuous glucose monitoring in patients with type 2 diabetes. *Diabetes Care*. 2012;35:32-38.
6. Uhl S, Choure A, Rouse B, Loblack A, Reaven P. Effectiveness of continuous glucose monitoring on metrics of glycemic control in type 2 diabetes mellitus: a systematic review and meta-analysis of randomized controlled trials. *J Clin Endocrinol Metab*. 2024;109:1119-1131.
7. Hall H, Perelman D, Breschi A, et al. Glucotypes reveal new patterns of glucose dysregulation. *PLoS Biol*. 2018;16:e2005143.
8. Reddy N, Verma N, Dungan K. Monitoring technologies- continuous glucose monitoring, mobile technology, biomarkers of glycemic control. In: Feingold KR, Anawalt B, Blackman MR, et al., eds. *Endotext*. MDText.com, Inc.; 2023.
9. Varghese JS, Ho JC, Anjana RM, et al. Profiles of intraday glucose in type 2 diabetes and their association with complications: an analysis of continuous glucose monitoring data. *Diabetes Technol Ther*. 2021;23(8):555-564. doi:10.1089/dia.2020.0672
10. Hjort A, Iggman D, Rosqvist F. Glycemic variability assessed using continuous glucose monitoring in individuals without diabetes and associations with cardiometabolic risk markers: a systematic review and meta-analysis. *Clin Nutr*. 2024;43:915-925.
11. Sakamoto M. Type 2 diabetes and glycemic variability: various parameters in clinical practice. *J Clin Med Res*. 2018;10:737-742.
12. Zhou J, Jia W, Bao Y, et al. Glycemic variability and its responses to intensive insulin treatment in newly diagnosed type 2 diabetes. *Med Sci Monit*. 2008;14:CR552-8.
13. Longato E, Acciaroli G, Facchinetti A, et al. Glycaemic variability-based classification of impaired glucose tolerance vs. type 2 diabetes using continuous glucose monitoring data. *Comput Biol Med*. 2018;96:141-146.
14. Fabris C, Facchinetti A, Fico G, et al. Parsimonious description of glucose variability in type 2 diabetes by sparse principal component analysis. *J Diabetes Sci Technol*. 2015;10:119-124.
15. Miller M, Strange P. Use of fourier models for analysis and interpretation of continuous glucose monitoring glucose profiles. *J Diabetes Sci Technol*. 2007;1:630-638.
16. Fico G, Hernández L, Cancela J, et al. Exploring the frequency domain of continuous glucose monitoring signals to improve characterization of glucose variability and of diabetic profiles. *J Diabetes Sci Technol*. 2017;11:773-779.
17. Ajjan RA. The clinical importance of measuring glycaemic variability: Utilising new metrics to optimise glycaemic control. *Diabetes Obes Metab*. 2024;26(Suppl 7):3-16.
18. Keshet A, Shilo S, Godneva A, et al. CGMap: characterizing continuous glucose monitor data in thousands of non-diabetic individuals. *Cell Metab*. 2023;35:758-769.

19. Tao R, Yu X, Lu J, et al. Multilevel clustering approach driven by continuous glucose monitoring data for further classification of type 2 diabetes. *BMJ Open Diabetes Res Care*. 2021;9(1):e001869. doi:10.1136/bmjdr-2020-001869
20. Shao X, Lu J, Tao R, et al. Clinically relevant stratification of patients with type 2 diabetes by using continuous glucose monitoring data. *Diabetes Obes Metab*. 2024;26(6):2082-2091. doi:10.1111/dom.15512
21. World Health Organization. *International Statistical Classification Diseases, Tenth Revision (ICD-10)*. World Health Organization; 1992.
22. Battelino T, Danne T, Bergenstal RM, et al. Clinical targets for continuous glucose monitoring data interpretation: recommendations from the international consensus on time in range. *Diabetes Care*. 2019;42:1593-1603.
23. Service FJ. Glucose variability. *Diabetes*. 2013;62(5):1398-1404. doi:10.2337/db12-1396
24. Bent B, Henriquez M, Dunn J. cgmquantify: Python and R packages for comprehensive analysis of interstitial glucose and glycemic variability from continuous glucose monitor data. arXiv [q-bio.QM] 2021 <http://arxiv.org/abs/2102.11212>
25. Welch P. The use of fast Fourier transform for the estimation of power spectra: a method based on time averaging over short, modified periodograms. *IEEE Trans Audio Electroacoust*. 1967;15:70-73.
26. Virtanen P, Gommers R, Oliphant TE, et al. SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat Methods*. 2020;17:261-272.
27. McDonald CJ, Huff SM, Suico JG, et al. LOINC, a universal standard for identifying laboratory observations: a 5-year update. *Clin Chem*. 2003;49(4):624-633. doi:10.1373/49.4.624
28. Ward JH Jr. Hierarchical grouping to optimize an objective function. *J Am Stat Assoc*. 1963;58:236-244.
29. Mann HB, Whitney DR. On a test of whether one of two random variables is stochastically larger than the other. *Ann Math Stat*. 1947;18:50-60.
30. Bland JM, Altman DG. Statistics notes: multiple significance tests: the Bonferroni method. *BMJ*. 1995;310:170.
31. Williams DM, Jones H, Stephens JW. Personalized type 2 diabetes management: an update on recent advances and recommendations. *Diabetes Metab Syndr Obes*. 2022;15:281-295.
32. Uemura F, Okada Y, Torimoto K, Tanaka Y. Enlarged glycemic variability in sulfonylurea-treated well-controlled type 2 diabetics identified using continuous glucose monitoring. *Sci Rep*. 2021;11:4875.
33. Mo Y, Ma X, Lu J, et al. Defining the target value of the coefficient of variation by continuous glucose monitoring in Chinese people with diabetes. *J Diabetes Investig*. 2021;12:1025-1034.
34. Jin S-M, Kim T-H, Bae JC, et al. Clinical factors associated with absolute and relative measures of glycemic variability determined by continuous glucose monitoring: an analysis of 480 subjects. *Diabetes Res Clin Pract*. 2014;104:266-272.
35. Mo Y, Lu J, Zhou J. Glycemic variability: measurement, target, impact on complications of diabetes and does it really matter? *J Diabetes Investig*. 2024;15:5-14.
36. Sun B, Luo Z, Zhou J. Comprehensive elaboration of glycemic variability in diabetic macrovascular and microvascular complications. *Cardiovasc Diabetol*. 2021;20:9.
37. Ohara M, Fukui T, Ouchi M, et al. Relationship between daily and day-to-day glycemic variability and increased oxidative stress in type 2 diabetes. *Diabetes Res Clin Pract*. 2016;122:62-70.
38. Ceriello A, Esposito K, Piconi L, et al. Oscillating glucose is more deleterious to endothelial function and oxidative stress than mean glucose in normal and type 2 diabetic patients. *Diabetes*. 2008;57:1349-1354.
39. Monnier L, Mas E, Ginet C, et al. Activation of oxidative stress by acute glucose fluctuations compared with sustained chronic hyperglycemia in patients with type 2 diabetes. *JAMA*. 2006;295:1681-1687.
40. Jun JE, Jin S-M, Baek J, et al. The association between glycemic variability and diabetic cardiovascular autonomic neuropathy in patients with type 2 diabetes. *Cardiovasc Diabetol*. 2015;14:70.
41. Hsu C-R, Chen Y-T, Sheu WH-H. Glycemic variability and diabetes retinopathy: a missing link. *J Diabetes Complications*. 2015;29:302-306.
42. Su G, Mi S, Tao H, et al. Association of glycemic variability and the presence and severity of coronary artery disease in patients with type 2 diabetes. *Cardiovasc Diabetol*. 2011;10:19.
43. Tang X, Li S, Wang Y, et al. Glycemic variability evaluated by continuous glucose monitoring system is associated with the 10-y cardiovascular risk of diabetic patients with well-controlled HbA1c. *Clin Chim Acta*. 2016;461:146-150.
44. Park J, Kim G, Kim B-S, et al. Insulin fact sheet in type 1 and 2 diabetes mellitus and trends of antidiabetic medication use in insulin users with type 2 diabetes mellitus: 2002 to 2019. *Diabetes Metab J*. 2023;47:211-219.

## SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

**How to cite this article:** Healey E, Morato C, Murillo J, Kohane I. Heterogeneity of continuous glucose monitoring features and their clinical associations in a type 2 diabetes population. *Diabetes Obes Metab*. 2025;27(7):3957-3966. doi:10.1111/dom.16432