diabetesresearchandclinicalpractice215(2024)111803

Contents lists available at ScienceDirect

Diabetes Research and Clinical Practice

journal homepage: www.journals.elsevier.com/diabetes-research-and-clinical-practice

Cluster analysis of adult individuals with type 1 diabetes: Treatment pathways and complications over a five-year follow-up period

Francisco J. Somolinos-Simón [a], Gema García-Sáez [a,b,*], Jose Tapia-Galisteo [a,b], Rosa Corcoy [b,c,d], M. Elena Hernando [a,b]

[a] Centre for Biomedical Technology (CTB), ETSI de Telecomunicación, Universidad Politécnica de Madrid, Madrid, Spain
[b] CIBER-BBN, ISCIII, Madrid, Spain
[c] Departament de Medicina, Universitat Autònoma de Barcelona, Bellaterra, Barcelona, Spain
[d] Institut de Recerca, Hospital de la Santa Creu i Sant Pau, Barcelona, Spain

ARTICLE INFO

Keywords:
Diabetes
Clustering
Glycaemic control
Complications
And pharmacologic treatment

ABSTRACT

Aims: To identify subgroups of adults with type 1 diabetes and analyse their treatment pathways and risk of diabetes-related complications over a 5-year follow-up.
Methods: We performed a k-means cluster analysis using the T1DExchange Registry (n = 6,302) to identify subgroups based on demographic and clinical characteristics. Annual reassessments linked treatment trajectories with these clusters, considering drug and technology use. Complication risks were analysed using Cox regression.
Results: Five clusters were identified: 1) A favourable combination of all variables (31.67 %); 2) Longer diabetes duration (22.63 %); 3) Higher HbA1c levels (13.28 %); 4) Higher BMI (15.25 %); 5) Older age at diagnosis (17.17 %). Two-thirds of patients remained in their initial cluster annually. Technology adoption showed improved glycaemic control over time. Cox proportional hazards showed different risk patterns: Cluster 1 had low complication risk; Cluster 2 had the highest risk for retinopathy, coronary artery disease and autonomic neuropathy; Cluster 3 had the highest risk for albuminuria, depression and diabetic ketoacidosis; Cluster 4 had increased risk for multiple complications; Cluster 5 had the highest risk for hypertension and severe hypoglycaemia, with elevated coronary artery disease risk.
Conclusions: Clinical characteristics can identify subgroups of patients with T1DM showing differences in treatment and complications during follow-up.

1. Introduction

Type 1 diabetes mellitus (T1DM) is a chronic disease that results from a lack of endogenous insulin secretion, due to the loss of β-cells secondary to an autoimmune attack [1]. Even when T1DM represents less than 10 % of all diabetes diagnosis [2], the Institute for Health Metrics and Evaluation estimated that there were 22 million people living with T1DM globally in 2019 [3].

The goal of treatment in T1DM is to achieve and maintain glycaemic levels as close as possible to the physiological ranges. The classical standard in the global assessment of glycaemic control is glycated haemoglobin (HbA1c), owing to its well-established association with micro and macrovascular disease [4,5]. The majority of adults with T1DM remains above reference standards of care, currently recognized as HbA1c < 53 mmol/mol (7 %) [6–8].

T1DM has been consistently shown to be associated with reduced quality of life (QoL), and exercise has been identified as an important mediating factor [9]. Regular physical activity plays an important role in the management of T1DM [10] and is associated with benefits in lipid profiles, glycaemic control (reduction in daily insulin dose and HbA1c), and fitness [11]. Physical activity is associated with better Health-Related Quality-of-Life (HRQoL), and the lack of physical activity in adulthood leads to greater use of medical resources, overall leading to the recommendation to promote physical activity in this population [11].

The current standard for insulin delivery is to use a basal-bolus schema, either through multiple daily injections (MDI) or Continuous Subcutaneous Insulin Infusion (CSII) [12]. Randomized controlled trials

* Corresponding author at: ETSI de Telecomunicación, Photonics Technology and Bioengineering Dept., Universidad Politécnica de Madrid, Avda. Complutense, 30 28040, Madrid, Spain.
E-mail address: gema.garcia.saez@upm.es (G. García-Sáez).

https://doi.org/10.1016/j.diabres.2024.111803
Received 23 February 2024; Received in revised form 14 June 2024; Accepted 29 July 2024
Available online 30 July 2024

(RCTs) have shown that, when compared with MDI, CSII slightly decreases HbA1c level [13–15]. In turn, Continuous Glucose Monitoring (CGM), which was introduced two decades ago and has had over 30 clinical trials completed with more than 4,000 subjects enrolled, has been shown to improve HbA1c levels and/or decrease hypoglycaemia [16].

One disadvantage of RCTs is that participants may not reflect the actual clinical scenario due to the clear set of inclusion or exclusion criteria [17]. A way to increase the representativeness of the entire population is to use Real-world data (RWD), which forms the basis of real-world evidence (RWE) and can be extracted from a broad range of sources, such as patient registries and healthcare databases [18]. RWD have several differences compared to data collected from RCTs in controlled settings, including advantages and disadvantages. Firstly, RWD are observational as opposed to data gathered in a controlled setting. Secondly, many types of RWD are unstructured and at times inconsistent due to variations in data entry across providers and healthcare systems. Thirdly, RWD may be incomplete and lack key endpoints for analysis, as the original collection was not intended for this purpose. Fourthly, it is important to note that RWD can be generated in a high-frequency manner resulting in voluminous and dynamic data [19].

Recent approaches driven by RWD have been published to extract subgroups of patients with significantly different characteristics and risks of diabetic complications [20,21]. A refined classification can provide a powerful tool to identify those at greatest risk of complications and enable individualized treatment regimens in the same way as genetic diagnosis of monogenic diabetes guides clinicians to optimal treatment [20].

Ahlqvist and colleagues [20] used data from five cohorts and identified five clusters of individuals with diabetes using a data-driven clustering analysis [20]: severe autoimmune diabetes (SAID), severe insulin-deficient diabetes (SIDD), severe insulin-resistant diabetes (SIRD), mild obesity-related diabetes (MOD), mild age-related diabetes (MARD). The relevance of the identified novel diabetes subtypes is illustrated by the later initiation of insulin treatment but similar or earlier progression to diabetic retinopathy in SIDD compared to SAID and the higher risk of diabetes nephropathy in SIRD compared to other subgroups. This clustering-based sub-classification reflects the pathophysiology, predicts the future risk of diabetic complications and comorbidities and constitutes a suitable tool for optimising personalised therapeutic strategies [22]. Additionally, the evolution of QoL on a cluster-by-cluster basis has been investigated in relation to disease severity [23]. Werkman and colleagues [23] established that disease severity suggested by the clusters of T2DM is not entirely reflected in QoL. Use of the suggested cluster names in practice should be carefully considered, as the non-neutral nomenclature may affect disease perception in individuals with T2DM and their healthcare providers. Although, longitudinal data on QoL in diabetes are scarce, as well as clustering studies based on these data.

A recent systematic review [24] on precision subclassification of type 2 diabetes concluded that studies using complex stratification produced reproducible diabetes subtypes associated with outcomes. It is important to demonstrate whether these subtypes can be replicated in more diverse ancestries and last but not least, whether tailoring interventions to subtypes will improve outcomes.

It is well known that T1DM is a heterogeneous condition, but there have been few attempts to categorize patients in a clinically useful way. Kakoshka and colleagues [25] used hierarchical agglomerative clustering to derive six clusters based on body mass index (BMI) and HbA1c. However, due to the cross-sectional design of the study, it was not possible to determine the association of clusters with subsequent complications. Battaglia and colleagues [26] introduced the endotype concept defined by observations/hypothesis, response to immune interventions and driven by data. The initial categorization after observations/hypothesis would include two groups with differences in age at presentation, Human Leukocyte Antigen (HLA)-DR associations, autoantibody patterns and progression to insulin deficiency. Again, this research did not provide information on the risk of complications or related treatments for subjects with T1DM, which is of utmost importance. Ahola and colleagues [27] clustered risk behaviours in a cross-sectional study of adults with T1DM and found that they are adversely associated with cardiometabolic risk factors.

The aim of this study is to identify subgroups of patients with T1DM with cluster analysis and a limited number of clinical and demographic characteristics. Additionally, the study analyses treatment pathways and the cluster-wise association with diabetes-related complications during a 5-year follow-up based on data gathered during clinical practice.

2. Materials and methods

2.1. Study population

We used data from the T1D Exchange Registry (T1DXR), a deidentified, publicly available dataset comprising 34,013 adult and paediatric participants with T1DM from 83 clinic sites in the United States [28]. The entire population selected for this study is considered to have a diagnosis of T1DM, because to be enrolled in the T1D Exchange registry, an individual must have a clinical diagnosis of presumed autoimmune T1DM and either have islet cell antibodies present, or, if antibodies were negative or unknown, then insulin must have been started at or shortly after diagnosis and used continuously thereafter [28]. The database contains longitudinal information of individuals who received routine clinical care between 2007 and 2018.[1] This information includes clinical data, demographics, associated complications, and diabetes-related treatments.

The inclusion criteria are the following: patients $>$ 18 years old at the time of enrolment; information on insulin treatment available; and information available on more than one visit. The exclusion criteria are a diagnosis of type 2 diabetes or other types of diabetes. From the total database, 9,156 patients were considered, with a total of 46,668 clinical visits registered.

Data processing has been carried out using python v3.9.

2.2. Variable selection

The variable selection for clustering analysis was carried out taking into account the variables used in previous clustering analysis [20,21,29,30] and the availability of the information in T1DXR. The variables were extracted at the time of enrolment and during follow-up visits. The variables selected for the cluster analysis are: age at diagnosis, diabetes duration, BMI and HbA1c level.

2.3. Data preparation

The dataset was cleaned, and variables converted to assure the quality of the data by reducing the noisy and irrelevant information that it could contain [31].

Patients with outliers in the selected variables were excluded, as clustering algorithms have a limited ability to deal with outliers [32]. Firstly, the Mahalanobis distance was used for outlier detection. Mahalanobis distances are based on the location and scatter of a multivariate normal distribution and can measure how distant any point in space is from the centre of this kind of distribution [33]. Then, the p-value for every Mahalanobis distance for each observation of the dataset was computed using chi squared function. Finally, the observations with a p-value less than 0.001 were removed because they are considered

[1] *The source of the data is the T1D Exchange, but the analyses, content and conclusions presented herein are solely the responsibility of the authors and have not been reviewed or approved by the T1D Exchange.*

outliers [34].

Of the 9,156 patients initially selected, 6,302 patients with complete data at the time of enrolment with respect to the variables required to build the cluster were considered. The observed differences between the included and excluded patients are minimal. The most notable differences are a slightly longer duration of diabetes in the excluded patients (mean 22.09 years [SD 14.18] vs. mean 19.68 years [SD 13.45]); a higher proportion of Black non-Hispanic subjects (5.71 % vs. 3.48 %); and a lower proportion of Hispanic or Latino subjects (3.19 % vs. 4.82 %). These differences are not considered to be clinically relevant. At each subsequent visit, all patients with complete data available at that time were re-evaluated and assigned to the previously identified clusters, including the 2,854 patients who were not considered for cluster creation because of missing data but who had complete data at subsequent visits. 917 patients were discarded due to the absence of one or more variables used to build the clusters or the presence of outliers at all follow-up visits.

Thus, 8,239 patients had complete information at one or more of the visits recorded during follow-up. The number of patients with complete data decreased to 3,552 five years after enrolment (Visit 5). At visit 6, the number of patients with complete data was drastically reduced to 1,585 patients, and to less than 270 patients at subsequent visits. Because of this reduction in the number of available cases, it was decided to analyse the follow-up of the first 5 visits after enrolment.

Before clustering analysis, we standardized age at diagnosis, diabetes duration, HbA1c, and BMI to make them comparable. The values of standardized variables are centred to a mean value of 0 and have a standard deviation (SD) of 1.

2.4. Cluster analysis

We applied K-means clustering, one of the most common and effective unsupervised algorithms for clustering analysis [35]. Given a set X of n objects, x1,…,xn, the K-means algorithm generates K disjoint clusters. The algorithm starts with a random initial partition of objects. At each iteration, K-means assigns a given object xi to cluster Sk, whose centre (i.e., centroid) ck, is the nearest to it. The cluster centres are updated at the end of each step of the algorithm. As such, the following least-squares objective function is alternatingly minimized as (1):

$$L(S,C)=\sum_{k=1}^{K}\sum_{x_i\in S_k}\sum_{v=1}^{V}(x_{iv}-c_{kv})^2 \quad (1)$$

where v = 1,…,V are the features (variables or dimensions) characterizing objects in X and ck ∈ C is the centroid of the cluster Sk ∈ S with k = 1,…,K. The sets S and C are respectively the sets of clusters and centroids. The algorithm stops when the object assignments do not change any more [32].

The K-means technique requires specification of the optimal number of clusters (K). Then, we determined the optimal number of clusters using Elbow Method and Average Silhouette Method.

2.4.1. Elbow method

By applying the k-means clustering algorithm, the elbow technique plots the variations through the number of clusters and picks the curve of elbow to get the optimal number. It depends on computing the sum of squared errors within-cluster (SSEWC) of all data points to represent the quality of aggregation between data points in the same cluster and separation between clusters [35].

2.4.2. Average silhouette method

The quality of clustering depends on the similarity of objects within clusters and the dissimilarity between clusters. That is, the larger the distance between clusters, the better the clustering effect. And the smaller the distance between objects in the same cluster, the better the clustering result. The silhouette coefficient of each object x is calculated as (2):

$$Silhouette\,(x)=\frac{b(x)-a(x)}{\max\{b(x),a(x)\}} \quad (2)$$

where, a(x) is the average distance from object x to all other objects in the same cluster, and b(x) is the minimum of the average distance from object x to all objects of other clusters. Finally, the mean silhouette coefficient over all samples is evaluated, with a possible range between [-1,1]. The closer the value is to 1, the better the efficiency of the clustering [36].

Cluster analysis was conducted at the time of initial enrolment, and each patient was assigned to a predefined cluster based on the centroids obtained at that time. Subsequently, patients were reassigned to the same cluster values during subsequent visits (from the first to the fifth-year follow-up visits) to assess cluster migration patterns using a Sankey diagram. Moreover, men and women were clustered separately to avoid accounting for sex-dependent differences in the cluster variables.

2.5. Statistical methods

Descriptive statistics are reported as mean ± SD for numerical variables. Specific traits of diabetes subgroups in the study cohort were compared using the CompareMeans function in SPSS version 28. A one-way analysis of variance (ANOVA) and Scheffe post hoc analysis were performed to assess differences between clusters. Statistical significance was defined as a p-value < 0.05.

Associations between clusters and characteristics of insulin treatment, use of diabetes technology or diabetes-related complications were calculated using crosstabs and are reported as proportions (%) of cases. The complications selected for the analysis were those with a prevalence of more than 15 % in any of the clusters during the follow-up period, while the insulin treatments selected were those which exceed 1 % in any of the clusters during the follow-up period. The risk of diabetes-related complications in the clusters was calculated using Cox regression in SPSS. For Cox regression analysis, left-censored data (if an event is known to have occurred before enrolment) and right-censored data (if the event does not occur during follow-up) [37] were taken into account. The Z-test was used to examine the statistical significance between clusters comparing the proportion of columns. Cluster 1 is used as the reference category. Significant differences are calculated at a significance level of 0.05.

All statistical analysis were done for all visits from the initial enrolment to the fifth-year of follow-up in each cluster.

3. Results

The Elbow method and Average Silhouette method led to the conclusion that the optimal number of clusters was five.

Fig. 1 shows the graphs (radar charts) where it can be observed how each variable stands out in each of the clusters obtained. The distribution of the 6,302 patients for the clustering variables at the time of enrolment was the following (see Table 1): 1,996 (31.67 %) were assigned to cluster 1, characterized by a more favourable combination of variables with a BMI in the healthy range, and the lowest value of HbA1c levels and diabetes duration; 1,426 patients (22.63 %) were assigned to cluster 2, characterized by the longest diabetes duration, a BMI in the overweight range, and HbA1c levels close to target values; 837 patients (13.28 %) were assigned to cluster 3, characterized by elevated HbA1c levels, an earlier age of diagnosis, and a relatively short diabetes duration; 961 patients (15.25 %) were assigned to cluster 4, characterized by a BMI within class II obesity ranges; and 1,082 patients (17.17 %) were assigned to cluster 5, characterized by a later age at diagnosis.

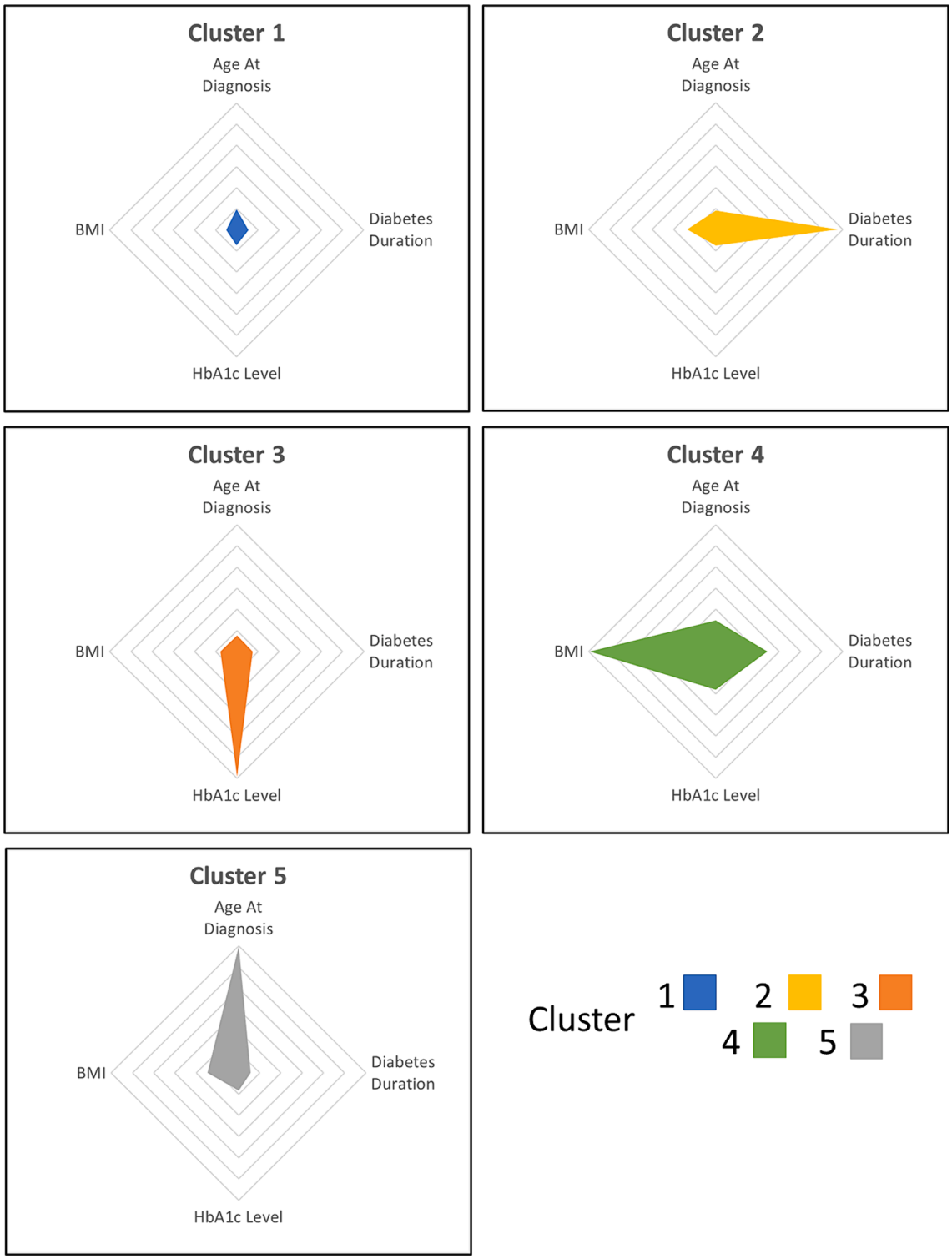

Fig. 1. Distribution of the cluster feature variables. Radar charts of each cluster with the corresponding standardized level of the feature variables. BMI, Body Mass Index; HbA1c, glycated haemoglobin.

3.1. Subgroup transitions over time

Information is available for 6,161 patients 1 year after the enrolment; for 5,489 patients after 2 years; for 5,155 patients after 3 years; for 4,751 patients after 4 years; and, for 3,552 patients after 5 years. The pattern of cluster distribution during annual visits for the next 5 years is shown in a Sankey diagram (Fig. 2). Overall, of all patients with available data at enrolment and during follow-up, one third switched cluster allocation (33.74 %). Analysing this diagram, we can see that the greatest migration of patients occurs from clusters 3 and 4 to cluster 1 (173 patients (20.67 %) and 79 patients (8.22 %) from Visit 0 to Visit 1, respectively) and from cluster 1 to cluster 3 and 4 (178 patients (8.92 %) and 40 (2.00 %) from Visit 0 to Visit 1, respectively). Supplementary tables S2, S3 and S4 present the baseline demographic and clinical characteristics, diabetes management technology and distribution of insulin prescriptions for the patients changing clusters from Visit 0 to

Table 1
Baseline demographic and clinical characteristics of the study participants stratified by clusters.

	CLUSTER 1 (N=1,996)	CLUSTER 2 (N=1,426)	CLUSTER 3 (N=837)	CLUSTER 4 (N=961)	CLUSTER 5 (N=1,082)
DEMOGRAPHICS					
Age (years)					
At diagnosis	13.57[a] (6.82)	13.47[a] (8.15)	12.68[a] (8.42)	16.45[b] (10.04)	**41.24**[c] (9.37)
Diabetes duration (years)	12.13[a] (6.86)	**38.17**[b] (8.56)	13.39[a] (8.33)	21.58[c] (10.51)	12.40[a] (8.77)
Sex					
Women	1,057[a] (52.96 %)	736[a] (51.61 %)	463[a] (55.32 %)	514[a] (53.49 %)	555[a] (51.29 %)
Men	939[a] (47.04 %)	690[a] (48.39 %)	374[a] (44.68 %)	447[a] (46.51 %)	527[a] (48.71 %)
Ethnicity					
Black non-Hispanic	34[a] (1.71 %)	38[a] (2.67 %)	69[b] (8.24 %)	48[b] (4.99 %)	30[a] (2.78 %)
Hispanic or Latino	136[a,b] (6.82 %)	33[c] (2.32 %)	80[b] (9.56 %)	42[a] (4.37 %)	13[c] (1.20 %)
White non-Hispanic	1,742[a] (87.41 %)	1,340[b] (93.71 %)	653[c] (78.02 %)	839[a] (87.30 %)	1,007[b] (93.15 %)
Other	81[a] (4.06 %)	19[b] (1.33 %)	35[a] (4.18 %)	32[a,b] (3.33 %)	31 (2.87 %)
CLINICAL CHARACTERISTICS					
Body mass index (Kg/m^2)	24.54[a] (2.93)	26.33[b] (3.42)	25.20[c] (3.75)	**35.41**[d] (3.92)	26.54[b] (3.97)
Blood pressure					
Systolic (mmHg)	118.51[a] (12.09)	124.83[b,c] (15.61)	119.91[a] (12.83)	125.80[b] (13.64)	123.52[c] (14.33)
Diastolic (mmHg)	72.08[a] (8.28)	70.75[b] (8.94)	73.64[c] (8.85)	75.22[d] (9.06)	72.91[a,c] (8.75)
LABORATORY TESTS					
HbA1c level (%)	7.31[a] (0.82)	7.33[a] (0.89)	**10.31**[b] (1.27)	7.95[c] (1.00)	7.39[a] (1.01)
Triglycerides (mg/dL)	82.77[a] (49.04)	82.32[a] (59.16)	116.06[b] (86.47)	119.16[b] (86.59)	87.18[a] (58.56)
HDL-c (mg/dL)	60.38[a] (16.45)	63.94[b] (19.18)	57.32[c] (15.93)	53.43[d] (15.71)	64.93[b] (21.32)
LDL-c (mg/dL)	88.81[a,b] (25.27)	85.25[a] (26.55)	100.80[c] (31.41)	96.37[d] (28.83)	89.11[b] (27.69)

Data are presented as mean (SD) for continuous variables or n (%) for categorical variables. Main characteristic of each cluster marked in bold black.
A one-way analysis of variance (ANOVA) and Scheffe post hoc analysis were performed to assess differences between clusters for continuous variables and a Z-test has been applied to study statistical significance between clusters for categorical variables. Statistical significance was defined as a p-value of < 0.05. Each superscript letter indicates whether or not there is a statistically significant difference at the 0.05 level between the mean values of each cluster, so that those with a different superscript letter differ significantly from each other, and those with the same letter do not differ significantly.
HbA1c, glycated haemoglobin; HDL, high density lipoprotein; LDL, low density lipoprotein cholesterol.

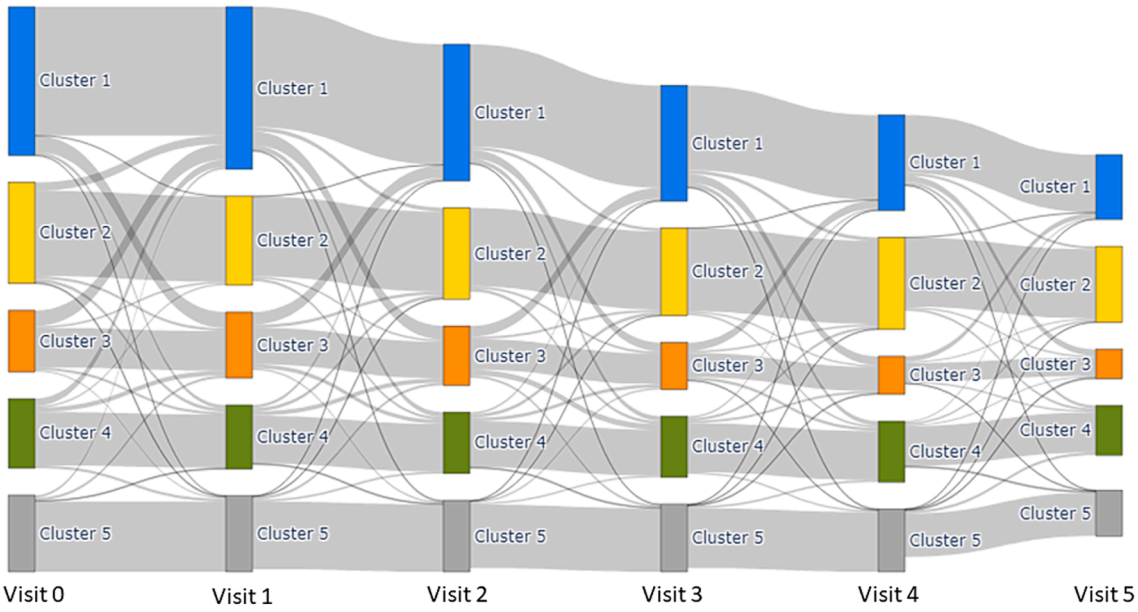

Fig. 2. Cluster redistribution at follow-up.

Visit 1. A total of 11.80 % of patients transitioning from cluster 1 to cluster 3 and 12.50 % of patients transitioning from cluster 1 to cluster 4 used CGM. In contrast, the percentage of patients who used CGM at Visit 0 and remained in the same cluster 1 was 22.89 % (Table S2). Moreover, 56.74 % of patients who transitioned from cluster 1 to cluster 3 used insulin pumps compared to 61.22 % of patients who remained in cluster 1 (Table S2). A total of 16.18 % of patients who transitioned from cluster 3 to cluster 1 used CGM devices, in comparison to 9.57 % of patients who remained in cluster 3 (Table S3). Similarly, 30.38 % of patients who transitioned from cluster 4 to cluster 1 employed CGM devices, in contrast to 24.66 % of patients who remained in cluster 4 (Table S4). Furthermore, 74.68 % of patients who transitioned from cluster 4 to cluster 1 used insulin pumps compared to 66.89 % of patients who remained in cluster 4 (Table S4). A similar pattern of patient migration between clusters 1, 3 and 4 is also observed in subsequent visits, with changes in the other clusters being less than 5 %.

3.2. *Demographic and clinical characteristics of diabetes subgroups*

Table 1 shows clinical data and metabolic characteristics of patients by cluster. Additionally, Figure S1 presents patients' characteristics at the 1-year, 2-year, 3-year, 4-year and 5-year reassessment, grouped by cluster.

At baseline, the percentage of men and women was similar in all

clusters, with slightly more women (higher than 50 %) than men in all the clusters (Table 1). In terms of evolution over time, the percentage of both men and women remains stable, with an increase in women in clusters 4 and 5 in recent years (Figure S1A). Similar results were obtained when clustering men and women separately with respect to the total population, so the results are reported for the clustering of total population.

As to ethnicity, most patients are White non-Hispanic. In cluster 3, the percentage of this ethnicity was slightly lower (78.02 %) than in other clusters, while in this same cluster, there is a higher percentage of Black non-Hispanic (8.24 %) and Hispanic or Latino (9.56 %) than in other clusters (Table 1). This distribution is maintained throughout the follow-up period.

The highest BMI at baseline was observed in patients assigned to cluster 4 and these differences were maintained over time (Figure S1B). BMI increased in all clusters through all visits, but cluster 4 is the only one that presents a mean value above 30 Kg/m2. During all visits, cluster 1 consistently displayed the lowest BMI value. Clusters 1 and 3 differ significantly from the other clusters at all visits with a p-value of < 0.05.

Patients in cluster 2, cluster 4 and cluster 5 had systolic blood pressure levels between 120 mmHg and 139 mmHg, placing them in the at risk category for hypertension [38,39]. Cluster 1 displayed values below 120 mmHg at all visits, while cluster 3 displayed values close to the risk of hypertension. With respect to diastolic blood pressure, all clusters in the study had an average below 80 mmHg over time (Figure S1C).

As for other patient characteristics, such as laboratory tests, patients assigned to cluster 3 had the highest HbA1c level, with an average value higher than 10 % throughout the entire follow-up period (Figure S1D). Patients in the other clusters had an HbA1c value between 7–8 %, with cluster 4 displaying the second-highest value, which was significantly different from the other clusters (p-value < 0.05).

Serum triglycerides were lower in patients assigned to cluster 1 (82.77 mg/dL [SD 49.04]), cluster 2 (82.32 mg/DL [SD 59.16]) and cluster 5 (87.18 mg/dL [SD 58.56]) compared to cluster 3 (116.06 mg/dL [SD 86.47]) and cluster 4 (119.16 mg/dL [SD 86.59]) (Table 1). However, the mean triglyceride levels remained below 150 mg/dL [40] throughout the follow-up period in all clusters (Figure S1E).

Mean low-density lipoprotein cholesterol (LDL-c) levels remained within the normal range over time for all clusters (Table 1). Moreover, mean high-density lipoprotein cholesterol (HDL-c) levels were in the optimal range or slightly above 60 mg/dL [41,42] in all clusters over time. However, clusters 3 and 4 exhibited lower HDL-c levels and higher LDL-c levels compared to the other three clusters. These differences between clusters decreased at the 5-year follow-up (Figure S1F).

3.3. Use of diabetes management technology in diabetes subgroups

At enrolment, MDI and CSII were used by 2,511 (39.84 %) and 3,786 (60.08 %) patients respectively (the insulin administration method is not stated in 5 patients (0.08 %)). Fig. 3A shows the use of insulin pumps during each visit, categorized by cluster allocation. It is important to note that at enrolment, CSII was the most frequently used method of insulin administration, except for cluster 3.

Throughout the visits, the use of insulin pumps increased in all clusters. At the 5-year follow-up, cluster 3 had the highest increase of patients using insulin pumps, reaching 58.82 %. Cluster 3 showed a slight improvement in glycaemic control over time (mean of 10.31 % [SD 1.27] at enrolment vs. mean of 10.04 % [SD 1.10] at 5-year follow-up) (Figure S1D).

A total of 1,427 (22.64 %) patients were using CGM at enrolment, increasing progressively with visits in all clusters, but always below 50 % (see Fig. 3B). At enrolment, cluster 3 had the lowest rate of CGM use (11.35 % compared to rates over 20 % in the other clusters) and this trend continued throughout the follow-up period.

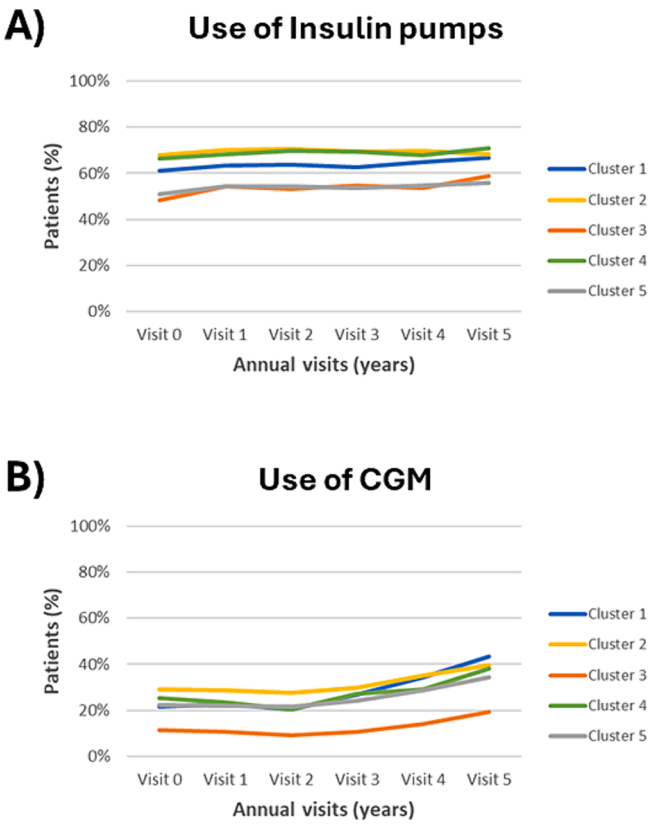

Fig. 3. Diabetes management technology at each of the visits by cluster allocation. A) Percentage of patients using insulin pumps in each of the clusters for each visit. The rest of the patients use injections/pens. B) Percentage of patients using CGM in each of the clusters for each visit. The rest of patients do not use CGM. CGM, Continuous Glucose Monitoring.

3.4. Insulin types used in diabetes subgroups

Figure S2 displays the percentage of patients using different insulin analogues at each visit for each cluster. The most commonly used analogues are aspart and lispro, both alone and in combination with glargine across all the clusters.

Figure S2A shows the use of different short acting analogues in patients treated with insulin pumps. At baseline, aspart insulin was the most commonly used analogue (40.8 %), glulisine was the least commonly used (1.6 %) and lispro was an intermediate choice (32.9 %). However, over the visits, there is a reduction in the use of insulin aspart alone in favour of insulin lispro or in combination with insulin lispro. About 10 % of patients reported the use of both aspart and lispro insulin. However, there are 17.5 % unknown values (17.8 % in cluster 1, 13.3 % in cluster 2, 25.1 % in cluster 3, 16.0 % in cluster 4 and 15.0 % in cluster 5). The percentage of each type of insulin analogue is reported in Figure S2A.

Patients in cluster 4 use less aspart insulin (45.5 % at enrolment) than patients in the other clusters (50.6 % in cluster 1, 47.4 % in cluster 2, 47.0 % in cluster 3, 56.5 % in cluster 5, at enrolment). The rate of patients in cluster 4 using aspart decreased over the years (34.5 % after 5-year follow-up), with a parallel increase in the use of lispro. Aspart insulin is the most common insulin in cluster 5 while lispro is the least common.

Figure S2B shows the type of short- and long-acting insulin analogues in patients treated with multiple daily injections. Lispro insulin is the preferred choice over aspart insulin in all clusters except cluster 5 (51.5 % vs. 48.5 % in cluster 1, 54.2 % vs. 45.8 % in cluster 2, 51.8 % vs. 48.2 % in cluster 3, 51,7% vs. 48.3 % in cluster 4, 44.9 % vs. 45.1 % in cluster 5, at enrolment). Regarding long-acting insulin analogues, in all clusters

and visits, more than 85.0 % use glargine insulin, with the highest usage in cluster 3 (92.2 % at enrolment). Less than 15.0 % of patients use detemir insulin.

3.5. Risk of diabetes complications in diabetes subgroups

The prevalence of retinopathy, albuminuria, autonomic neuropathy, coronary artery disease, hypertension and depression was higher than 15 % in any of the clusters during follow-up, so, these complications were selected for Cox regression analysis. Although, most episodes of severe hypoglycaemia and diabetic ketoacidosis (DKA) because were concentrated at the time of enrolment, with less episodes during follow-up, these complications were also included from the analysis. Cluster 2 had a higher prevalence (9.47 %) of severe hypoglycaemia while cluster 3 displayed a higher prevalence (14.58 %) of DKA at enrolment.

Fig. 4 presents the results of the Cox regression analysis comparing the risk of diabetic complications between clusters. No differences were observed between men and women. Table S1 shows the Cox proportional Hazards Ratio (HR) of diabetes-related complications among the clusters. The risk of complications varies for each outcome in each cluster. Cluster 1 (more favourable characteristics) had the lowest risk of diabetes-related complications. Cluster 2 (longer diabetes duration) had the highest risk for developing retinopathy [1.94 (1.49–2.54)], coronary artery disease [10.28 (3.05–34.59)] and autonomic neuropathy [3.58 (1.95–6.57)]. Cluster 3 (higher HbA1c) had the highest risk for developing albuminuria [4.86 (3.04–7.77)], depression [2.83 (1.60–5.01)] and DKA [1.86 (1.13–3.06)]. Cluster 4 (characterized by high BMI) had an increased risk of retinopathy [1.64 (1.22–2.21)], albuminuria [3.62 (2.27–5.78)], autonomic neuropathy [3.51 (1.84–6.69)], hypertension [3.73 (2.21–6.30)], and severe hypoglycaemia [1.89 (1.35–2.63)]. Cluster 5 (higher age at diagnosis) had an increased risk of coronary artery disease [9.79 (2.85–33.61)], and the highest risk of hypertension [4.66 (2.89–7.51)] and severe hypoglycaemia [2.11 (1.53–2.90)], while patients in cluster 5 had the lowest risk of retinopathy [0.54 (0.36–0.79)]. In terms of absolute risk, after 5 years of follow-up, over 12 % of patients in cluster 4 and cluster 2 will develop retinopathy, over 10 % will develop severe hypoglycaemia, and over 7 % of patients in these clusters will develop hypertension. Over 10 % of patients in cluster 5 will develop hypertension and severe hypoglycaemia after 5 years of follow-up, while nearly 12 % of patients in cluster 3 will develop albuminuria and 5 % of the same cluster will experience depression and DKA (see Fig. 4).

4. Discussion

The observations of this study reveal four main points. Firstly, the identification of subgroups of patients with good performance in terms of cluster analysis is possible using only four widely available characteristics. Secondly, patients within these clusters show differences in diabetes management, specifically in terms of insulin treatment and CGM use. Thirdly, cluster allocation is usually maintained over time. Finally, clustering assignment identifies different risks of complications and comorbidities.

It is important to emphasise that most clustering studies do not analyse a T1DM population independently, as they are conducted in cohorts that include both T1DM and T2DM, where all patients with T1DM are assigned to the same cluster (SAID) [20,43], they focus on cohorts with T2DM only [21,29] or have a cross-sectional design not allowing to address development of complications [25]. The current analysis has identified five new clusters in T1DM not found in previous works and has analysed over time, transition between clusters, changes in insulin treatment and diabetes technology and emergence of complications. Cluster 1 has more favourable HbA1c and overall characteristics and a low risk of complications. Patients in cluster 2 (longest duration), 3 (higher HbA1c), and 4 (higher BMI) display a higher risk of complications during follow-up.

Lipid characteristics are within the normal ranges [40,41] in all clusters. However, patients assigned to cluster 3 and cluster 4 displayed significantly higher triglyceride and LDL levels, as well as significantly lower HDL levels. These factors could contribute to the risk of complications in these clusters. The results suggest that obesity and glycaemic control are important contributors to dyslipidaemia in youth with diabetes [44].

During the follow-up period, the rate of CGM and insulin pump usage increased while HbA1c levels decreased. A potential connection between these two observations can be hypothesized, as the natural disease progression would lead to higher levels of HbA1c. Nevertheless, it is possible that other factors may have contributed to the observed reduction in HbA1c, such as improvements in mental health or healthier habitual practices.

Patients assigned to cluster 3 have the highest HbA1c levels and use CGM the least (11.35 % at enrolment) and the lowest percentages of insulin pump together with cluster 5. In this subgroup at 5 years, the number of patients using CGM increases (19.18 %) together with that of insulin pump (58.82 %) while HbA1c levels improve (mean 10.04 % [SD 1.10]). In the remaining clusters, there is also a slight increase in the use of CGM, while HbA1c levels remain lower.

Patients in cluster 3 use insulin pump less frequently, both at baseline and at follow-up. The increase in insulin pump usage during follow-up was higher in this cluster compared to the others, likely related to the parallel improvement in HbA1c, which was more pronounced than in the other clusters.

After analysing the Sankey plot to visualize transitions between patient subgroups during follow-up, most of the patients remain in the cluster to which they were assigned at enrolment. However, cluster membership is not an immutable feature since we can observe that over time some patients change clusters, as occurs in the study of Zaharia and colleagues [43]. These changes are more notable between cluster 1 and clusters 3 and 4, that is, changes in the BMI and HbA1c levels make patients move to a healthier cluster (cluster 1) if these conditions improve and vice versa. These transitions imply respectively a large decrease in HbA1c (for initial cluster 3) or BMI (for initial cluster 4). These changes are not easily achieved in clinical practice. But they are possible, i.e. linked to initiation of technologic devices (cluster 3) or bariatric surgery (cluster 4). Those patients who changed from cluster 1 to clusters 3 or 4 used CGM less frequently than those who remained in cluster 1. Furthermore, those who changed from cluster 1 to cluster 3 used insulin pumps less frequently than those who remained in cluster 1. In contrast, patients who changed from clusters 3 or 4 to cluster 1 used CGM more frequently than those who remained in clusters 3 or 4. Those who transitioned from cluster 4 to cluster 1 used insulin pumps more frequently than those who remained in cluster 4. In the clinical setting, a tool based on cluster assignment can help the clinician to target specific subgroups of patients at risk as candidates to treatment modification or escalation in the use of diabetes management technology, so that they move to a subgroup with lower risk and better prognosis. It can also be seen from the Sankey plot that clusters 2 and 5 experience a lower proportion of changes during follow-up. However, it is also observed that they may shift to subgroups with better prognosis (cluster 1) or with worse prognosis (clusters 3 and 4), so actions can be taken that have an impact on the evolution of these subgroups. However, as it has already been pointed in T2DM, the benefit of using clusters to guide clinical practice should be tested in specific trials [24].

Cluster 1 had the lowest risk of diabetes complications. Then, younger patients with shorter diabetes duration, lower HbA1c and lower BMI present a lower risk of complications. Cluster 5 (similar diabetes duration, HbA1c level and BMI, but higher age at diagnosis) had also a lower risk of complications, with the exception of hypertension [HR=4.66 (2.89–7.51)], as observed in other studies of patients with T2DM [20,43,45] and of coronary artery disease that would be consistent with higher age. Patients in cluster 3 present a higher risk of albuminuria (as in the study of Preechasuk and colleagues [45]), and

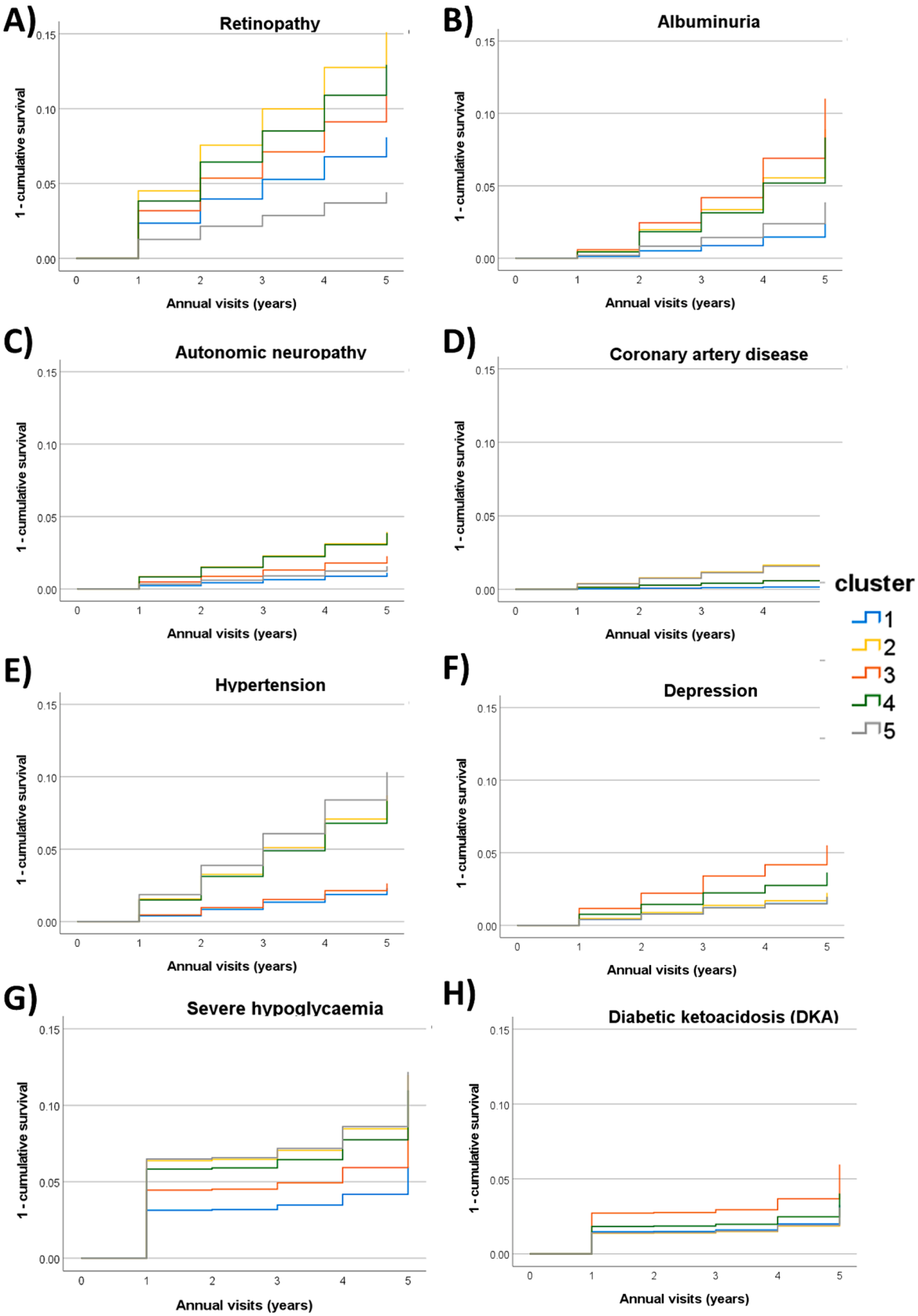

Fig. 4. Progression of complications and comorbidities over time by cluster at visit 0. Time to retinopathy (A), albuminuria (B), autonomic neuropathy (C), coronary artery disease (D), hypertension (E), depression (F), severe hypoglycaemia (G) and diabetic ketoacidosis (DKA) (H).

depression than other clusters, while patients in cluster 2 and 4 present higher risk of retinopathy and autonomic neuropathy than other clusters. These observations are in line with the known associations between diabetes duration, glycaemic control, and BMI with the risk of complications. It is important to note that current results on the risk of retinopathy in patients with T1DM (higher in cluster 2 characterized by longer diabetes duration) differ from previous studies in mixed populations where the highest hazards for retinopathy were observed in the cluster characterized by higher HbA1c [20,46]. Nevertheless, current results are consistent with risk of complications [47] associated with cumulative glycaemic exposure whereas diabetes duration was not a factor in the study of Ahlqvist [20] or Anjana [46] since they addressed new-onset or short-duration diabetes respectively. In turn, the association of depression with hyperglycaemia [48] is well established with bidirectional causality; current observation in cluster 3 would add to longitudinal studies supporting a causal role of hyperglycaemia [49]. Similarly, the risk of depression in cluster 4 would be in line with the known association between depression and obesity [50] and specifically with longitudinal studies supporting a causal role of obesity [51]. It is remarkable that the risk of complications in patients of cluster 4 (characterized by high BMI) is similar or higher than that of patients in cluster 3 (characterized by high HbA1c), even for microangiopathic ones; this underscores the relevance of factors different from hyperglycaemia. The high risk of hypertension and coronary artery disease in cluster 5 would be attributable to the higher age of this group.

Overweight and obesity are highly prevalent among youth and adults with T1DM [52,53], currently affecting about 50 % of patients with T1DM [54]. This is confirmed in this study, where average BMI of three of the clusters are in the range of overweight and one in the range of grade II obesity. The risk of developing metabolic syndrome, macrovascular and microvascular complications as a consequence of diabetes increases with the presence of obesity in individuals with T1DM [53]. This is also observed in this study, as patients in cluster 4, characterized by grade II obesity, have high risk of retinopathy, autonomic neuropathy and hypertension. For patients in this subgroup, adjustment of insulin therapy, adding other diabetes medications that positively impact body weight, or anti-obesity medications could be suggested [52–54].

The strengths of this study include the use of clustering techniques in patients with T1DM, which allowed us to identify new subgroups that differ from cluster studies in mixed populations or in patients with T2DM; the use of the duration of diabetes, a relevant variable of glycaemic exposure, as a clustering variable; and the longitudinal design, which has allowed us to evaluate the subgroup transitions, treatment changes and emergence of complications over time.

These findings are of significant importance for the implementation of a precision medicine approach to the prognosis of individuals with T1DM. In the current study, subclassification strategies have been deployed on the basis of clinical characteristics and with differences in risk of incident complications. However, a subclassification approach at baseline may not be sufficient, and this study has shown that cluster-based assignments of T1DM may change over time. Therefore, it is of paramount importance to reassess these assignments during follow-up. Nevertheless, as previously highlighted by publications in the field [24,55], the key factor in supporting the use of clusters to guide clinical practice is the need for RCTs to assess the benefits of a cluster-based diabetes classification vs a conventional approach.

A limitation of this analysis is the reduction of patient data over time. This is common in RWD studies that include large sample sizes and long follow-up periods, but data on key variables may be missed [56]. Another limitation is the high rate of White non-Hispanic ethnicity, disturbing overall generalization. Another possible limitation is the use of the K-means clustering algorithm. This technique requires prespecifying the optimal number of clusters, has difficulty incorporating categorical variables, and is relatively sensitive to the starting conditions used to initialize the algorithm such as the choice of seed or the order of data points [32,35,57]. The lack of longitudinal data on QoL or self-management factors such as physical activity practices or dietary intakes prevented a more complete examination of lifestyle factors.

Future research is necessary to replicate the clusters and their prognostic implications in different populations. The final and more relevant step should be to investigate whether the use of patient subgroups to guide clinical practice results in improved outcomes.

5. Conclusions

Clustering techniques allow for the identification of subgroups of patients with T1DM, with different demographic, clinical and metabolic characteristics. The findings emphasize the heterogeneity within the T1DM population and the association of specific clinical characteristics with the risk of diabetes-related complications. There was a notable overall trend towards the adoption of technology solutions for diabetes management, which was associated with improved glycaemic control. The clustering of patients with T1DM could inform more targeted and effective treatment strategies, reinforcing the need for personalized care plans in managing T1DM.

CRediT authorship contribution statement

Francisco J. Somolinos-Simón: Writing – review & editing, Writing – original draft, Resources, Methodology, Formal analysis, Conceptualization. **Gema García-Sáez:** Writing – review & editing, Supervision, Methodology, Conceptualization. **Jose Tapia-Galisteo:** Writing – review & editing, Supervision, Methodology, Conceptualization. **Rosa Corcoy:** Writing – review & editing, Conceptualization. **M. Elena Hernando:** Writing – review & editing, Supervision, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work has been supported by the HTx project, which has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 825162.

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.diabres.2024.111803.

References

[1] Ilonen J, Lempainen J, Veijola R. The heterogeneous pathogenesis of type 1 diabetes mellitus. Nat Rev Endocrinol 2019;15(11):635–50.
[2] Ellis SL, Moser EG, Snell-Bergeon JK, Rodionova AS, Hazenfield RM, Garg SK. Effect of sitagliptin on glucose control in adult patients with Type 1 diabetes: A pilot, double-blind, randomized, crossover trial. Diab med 2011;28(10):1176–81.
[3] Gregory GA, Robinson TIG, Linklater SE, Wang F, Colagiuri S, De Beaufort C, et al. Global incidence, prevalence, and mortality of type 1 diabetes in 2021 with projection to 2040: A modelling study. Lancet Diab Endocrinol 2022;10(10):741.
[4] Lachin JM, Genuth S, Nathan DM, Zinman B, Rutledge BN, DCCT/EDIC Research Group. Effect of glycemic exposure on the risk of microvascular complications in the diabetes control and complications trial–revisited. Diabetes 2008;57(4):995-1001.
[5] Nathan DM, Cleary PA, Backlund JC, Genuth SM, Lachin JM, Orchard TJ, et al. Intensive diabetes treatment and cardiovascular disease in patients with type 1 diabetes. N Engl J Med 2005;353(25):2643–53.
[6] Karakus KE, Akturk HK, Alonso GT, Snell-Bergeon JK, Shah VN. Association between diabetes technology use and glycemic outcomes in adults with type 1 diabetes over a decade. Diabetes Care 2023;46(9):1646–51.
[7] ElSayed NA, Aleppo G, Aroda VR, Bannuru RR, Brown FM, Bruemmer D, et al. 6. Glycemic targets: Standards of care in diabetes-2023. Diabetes Care 2023;46(Suppl 1):S97–110.

[8] Hernar I, Cooper JG, Nilsen RM, Skinner TC, Strandberg RB, Iversen MM, et al. Diabetes distress and associations with demographic and clinical variables: A nationwide population-based registry study of 10,186 adults with type 1 diabetes in Norway. Diabetes Care 2024;47(1):126–31.

[9] Alobaid AM, Zulyniak MA, Ajjan RA, Broz J, Hopkins M, Campbell MD. Barriers to exercise in adults with type 1 diabetes and insulin resistance. Can J Diabetes 2023;47(6):503–8.

[10] Matson RIB, Leary SD, Cooper AR, Thompson C, Narendran P, Andrews RC. Objective measurement of physical activity in adults with newly diagnosed type 1 diabetes and healthy individuals. Front Public Health 2018 December;07(6):360.

[11] Dominguez-Dominguez A, Martinez-Guardado I, Dominguez-Munoz FJ, Barrios-Fernandez S, Morenas-Martin J, Garcia-Gordillo MA, et al. Association between the level of physical activity and health-related quality of life in type 1 diabetes mellitus. A preliminary study. J Clin Med 2021;10(24):5829. https://doi.org/10.3390/jcm10245829.

[12] Odegard PS, Capoccia K. Medication taking and diabetes: A systematic review of the literature. Diabetes Educ 2007;33(6):1014.

[13] Rys PM, Ludwig-Slomczynska AH, Cyganek K, Malecki MT. Continuous subcutaneous insulin infusion vs multiple daily injections in pregnant women with type 1 diabetes mellitus: a systematic review and meta-analysis of randomised controlled trials and observational studies. Eur J Endocrinol 2018;178(5):545.

[14] Benkhadra K, Alahdab F, Tamhane SU, McCoy RG, Prokop LJ, Murad MH. Continuous subcutaneous insulin infusion versus multiple daily injections in individuals with type 1 diabetes: A systematic review and meta-analysis. Endocrine 2017;55(1):77–84.

[15] Wang X, Zhao X, Chen D, Zhang M, Gu W. Comparison of continuous subcutaneous insulin infusion and multiple daily injections in pediatric type 1 diabetes: A meta-analysis and prospective cohort study. Front Endocrinol (Lausanne) 2021;12:608232.

[16] Kravarusic J, Aleppo G. Diabetes technology use in adults with type 1 and type 2 diabetes. Endocrinol Metab Clin North Am 2020;49(1):37–55.

[17] Kim H, Lee S, Kim JH. Real-world evidence versus randomized controlled trial: Clinical research based on electronic medical records. J Korean Med Sci 2018;33(34):e213.

[18] Baumfeld Andre E, Reynolds R, Caubel P, Azoulay L, Dreyer NA. Trial designs using real-world data: The changing landscape of the regulatory approval process. Pharmacoepidemiol Drug Saf 2020;29(10):1201–12.

[19] Liu F, Demosthenes P. Real-world data: A brief review of the methods, applications, challenges and opportunities. BMC Med Res Methodol 2022;22(1):287.

[20] Ahlqvist E, Storm P, Käräjämäki A, Martinell M, Dorkhan M, Carlsson A, et al. Novel subgroups of adult-onset diabetes and their association with outcomes: a data-driven cluster analysis of six variables. Lancet Diab Endocrinol 2018;6(5):361.

[21] Dennis JM, Shields BM, Henley WE, Jones AG, Hattersley AT. Disease progression and treatment response in data-driven subgroups of type 2 diabetes compared with models based on simple clinical features: an analysis using clinical trial data. Lancet Diab Endocrinol 2019;7(6):442.

[22] Tanabe H, Masuzaki H, Shimabukuro M. Novel strategies for glycaemic control and preventing diabetic complications applying the clustering-based classification of adult-onset diabetes mellitus: A perspective. Diabetes Res Clin Pract 2021;180:109067.

[23] Werkman NCC, Garcia-Saez G, Nielen JTH, Tapia-Galisteo J, Somolinos-Simon FJ, Hernando ME, et al. Disease severity-based subgrouping of type 2 diabetes does not parallel differences in quality of life: The Maastricht Study. Diabetologia 2024.

[24] Misra S, Wagner R, Ozkan B, Schon M, Sevilla-Gonzalez M, Prystupa K, et al. Precision subclassification of type 2 diabetes: A systematic review. Commun Med (Lond) 2023;3(1):138.

[25] Kahkoska AR, Nguyen CT, Jiang X, Adair LA, Agarwal S, Aiello AE, et al. Characterizing the weight-glycemia phenotypes of type 1 diabetes in youth and young adulthood. BMJ Open Diabetes Res Care 2020;8(1):e000886. https://doi.org/10.1136/bmjdrc.

[26] Battaglia M, Ahmed S, Anderson MS, Atkinson MA, Becker D, Bingley PJ, et al. Introducing the endotype concept to address the challenge of disease heterogeneity in type 1 diabetes. Diabetes Care 2020;43(1):5–12.

[27] Ahola AJ, Tikkanen-Dolenc H, Harjutsalo V, Groop P, FinnDiane Study Group. Clustering of risk behaviours and associations between risk behaviours and cardio-metabolic risk factors in adult individuals with type 1 diabetes. Diabetes Res Clin Pract 2024;208:111115.

[28] Beck RW, Tamborlane WV, Bergenstal RM, Miller KM, DuBose SN, Hall CA, et al. The T1D exchange clinic registry. J Clin Endocrinol Metab 2012;97(12):4383–9.

[29] Bello-Chavolla OY, Bahena-Lopez JP, Vargas-Vazquez A, Antonio-Villa NE, Marquez-Salinas A, Fermin-Martinez CA, et al. Clinical characterization of data-driven diabetes subgroups in Mexicans using a reproducible machine learning approach. BMJ Open Diabetes Res Care 2020;8(1):e001550. https://doi.org/10.1136/bmjdrc.

[30] Safai N, Ali A, Rossing P, Ridderstrale M. Stratification of type 2 diabetes based on routine clinical markers. Diabetes Res Clin Pract 2018;141:275–83.

[31] Béjar Alonso J. Strategies and algorithms for clustering large datasets: a review. 2013.

[32] Gondeau A, Aouabed Z, Hijri M, Peres-Neto P, Makarenkov V. Object weighting: A new clustering approach to deal with outliers and cluster overlap in computational biology. IEEE/ACM Trans Comput Biol Bioinform 2021;18(2):633–43.

[33] Etherington TR. Mahalanobis distances for ecological niche modelling and outlier detection: implications of sample size, error, and bias for selecting and parameterising a multivariate location and scatter method. PeerJ 2021;9:e11436.

[34] Etherington TR. Mahalanobis distances and ecological niche modelling: correcting a chi-squared probability error. PeerJ 2019;7:e6678.

[35] Sammouda R, El-Zaart A. An optimized approach for prostate image segmentation using K-means clustering algorithm with elbow method. Comput Intell Neurosci 2021;2021:4553832.

[36] Lin X, Zhang X. Identification of hot regions in hub protein-protein interactions by clustering and PPRA optimization. BMC Med Inform Decis Mak 2021;21(Suppl 1):143.

[37] Ihwah A. The use of Cox regression model to analyze the factors that influence consumer purchase decision on a product. Agric Agric Sci Procedia 2015;3:78–83.

[38] Giles TD, Materson BJ, Cohn JN, Kostis JB. Definition and classification of hypertension: An update. J Clin Hypertens (Greenwich) 2009;11(11):611–4.

[39] Singh JN, Nguyen T, Kerndt CC, et al. Physiology, Blood Pressure Age Related Changes. [Updated 2023 Aug 28]. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; Available from: https://www.ncbi.nlm.nih.gov/books/NBK537297/.

[40] Toth PP, Philip S, Hull M, Granowitz C. Elevated triglycerides (>/=150 mg/dL) and high triglycerides (200–499 mg/dL) are significant predictors of hospitalization for new-onset kidney disease: A real-world analysis of high-risk statin-treated patients. Cardiorenal Med 2019;9(6):400–7.

[41] Lee Y, Siddiqui WJ. Cholesterol Levels. [Updated 2023 Jul 24]. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing LLC; 2023. Available from: https://www.ncbi.nlm.nih.gov/books/NBK542294/.

[42] Executive Summary of the Third Report of the National Cholesterol Education Program (NCEP) Expert Panel on Detection, Evaluation, and Treatment of High Blood Cholesterol in Adults (Adult Treatment Panel III) Expert Panel on Detection, Evaluation, and Treatment of High Blood Cholesterol in Adults The Third Report of the Expert Panel on Detection, Evalu- ation, and Treatment of High Blood Cholesterol in Adults (Adult Treatment Panel III, or ATP III) constitutes the National. JAMA. 2001;285(19):2486-97.

[43] Zaharia OP, Strassburger K, Strom A, Bonhof GJ, Karusheva Y, Antoniou S, et al. Risk of diabetes-associated diseases in subgroups of patients with recent-onset diabetes: a 5-year follow-up study. Lancet Diabetes Endocrinol 2019;7(9):684–94.

[44] Ashour D, AlSubaihin A, Lim YY, Legault L. Overweight and obese youth with type 1 or type 2 diabetes share similar elevation in triglycerides during middle and late adolescence. Obes Res Clin Pract 2022;16(2):138–43.

[45] Preechasuk L, Khaedon N, Lapinee V, Tangjittipokin W, Srivanichakorn W, Sriwijitkamol A, et al. Cluster analysis of Thai patients with newly diagnosed type 2 diabetes mellitus to predict disease progression and treatment outcomes : A prospective cohort study. BMJ Open Diabetes Res Care 2022;10(6):e003145. https://doi.org/10.1136/bmjdrc.

[46] Anjana RM, Baskar V, Nair ATN, Jebarani S, Siddiqui MK, Pradeepa R, et al. Novel subgroups of type 2 diabetes and their association with microvascular outcomes in an Asian Indian population: A data-driven cluster analysis: the INSPIRED study. BMJ Open Diabetes Res Care 2020;8(1):e001506. https://doi.org/10.1136/bmjdrc.

[47] Miller RG, Orchard TJ. Understanding metabolic memory: A tale of two studies. Diabetes 2020;69(3):291–9.

[48] Lustman PJ, Anderson RJ, Freedland KE, de Groot M, Carney RM, Clouse RE. Depression and poor glycemic control: a meta-analytic review of the literature. Diabetes Care 2000;23(7):934–42.

[49] Rotella F, Mannucci E. Diabetes mellitus as a risk factor for depression. A meta-analysis of longitudinal studies. Diabetes Res Clin Pract 2013;99(2):98–104.

[50] Mannan M, Mamun A, Doi S, Clavarino A. Is there a bi-directional relationship between depression and obesity among adult men and women? Systematic review and bias-adjusted meta analysis. Asian J Psychiatr 2016;21:51–66.

[51] Jokela M, Laakasuo M. Obesity as a causal risk factor for depression: Systematic review and meta-analysis of Mendelian Randomization studies and implications for population mental health. J Psychiatr Res 2023;163:86–92.

[52] Corbin KD, Driscoll KA, Pratley RE, Smith SR, Maahs DM, Mayer-Davis EJ, et al. Obesity in type 1 diabetes: Pathophysiology, clinical impact, and mechanisms. Endocr Rev 2018;39(5):629–63.

[53] Polsky S, Ellis SL. Obesity, insulin resistance, and type 1 diabetes mellitus. Curr Opin Endocrinol Diabetes Obes 2015;22(4):277–82.

[54] Mottalib A, Kasetty M, Mar JY, Elseaidy T, Ashrafzadeh S, Hamdy O. Weight management in patients with type 1 diabetes and obesity. Curr Diab Rep 2017;17(10):92–8.

[55] Herder C, Roden M. A novel diabetes typology: towards precision diabetology from pathogenesis to treatment. Diabetologia 2022;65(11):1770–81.

[56] Nazha B, Yang JC, Owonikoko TK. Benefits and limitations of real-world evidence: Lessons from EGFR mutation-positive non-small-cell lung cancer. Future Oncol 2021;17(8):965–77.

[57] Celebi ME, Kingravi HA, Vela PA. A comparative study of efficient initialization methods for the k-means clustering algorithm. Expert Syst Appl 2013;40(1):200–10.