Supplementary Appendix

Imputation Methods

The clinical outcomes of interest to guide the formation of clusters include BMIz (Y1) and HbA1c (Y2) at the cohort visit. We denote other variables including modifiable and non-modifiable characteristics as X-variables; 28 patient co-variates were chosen to capture a breadth of individual characteristics available at the follow-up visit including sociodemographic, clinical, anthropometric, laboratory, psychosocial and behavioral measures (see **Supplementary Table S1**).

To perform Reinforcement Learning Trees (RLT), only complete data can be used. Missing X-variables were imputed. In short, a model for each covariate was constructed using all other covariates as predictors. Missing values were replaced with their model predicted values recursively until convergence. As opposed to basic linear models, random forest models were used to capture more complicated relationships between the predictors. This missForest algorithm[1] is described in further detail as follows:

Consider that there are p covariates, $X_1, \ldots, X_p$.

1. A Strawman imputation is performed for all variables initially: each missing value of X_1 is replaced by the median of all observed values of X_i if the variable is continuous, for i = 1, ..., p. If X_1 is categorical, it is replaced by the mode.
2. For all records where X_1 is observed, build a random forest model to predict X_1 using the remaining p-1 covariates, which are either observed or Strawman imputed.
3. Repeat step 2 for all p covariates.
4. Update all Strawman imputed values with their model predicted values from steps 2-3.
5. Repeat the model building and updating steps until desired convergence is observed.

This relatively new method imputes missing values using a random forest prediction model instead of a typical regression and has been shown to work well with highly correlated features. Simulations on this imputation algorithm was performed to show that the joint distribution of covariates is roughly preserved, using Kolmogorov-Smirnov test on the imputed missing at random (MAR) data and the original data. The performance of this imputation method was found to be comparable to that of widely used methods such as multiple imputation by chained equations.[2]

Reinforcement Learning Trees

The primary objective of the cluster analysis was to group individuals based on a precise weight-glycemia phenotype, i.e. two outcomes that exist across a continuum with variable relationships to each other. To avoid limitations of conventional supervised clustering,[3] we employed a novel, semi-supervised clustering technique to group individuals by five measures of joint distribution of BMIz and HbA1c measured at the follow-up visit: the means and variance of BMIz and HbA1c and their covariance. These measures were estimated by RLT using the X-variables described above (see **Supplementary Table S1**). Variables that were derived from multiple primary measures were not problematic for RLT as they represented nontrivial derivations of the primary measures. Although the data used for analysis are cross-sectional, we were thus able to obtain an estimate of the within-patient variance by fitting a model for the patient's deviation from their expected measurement, which is the formal definition of variance.

After imputation, five RLTs were constructed to predict the outcomes and model the variances of and the covariance between BMIz and HbA1c, based on the X-variables. RLT is a tree-based machine learning method that uses bootstrapping and reinforcement learning and exhibits significantly improved performance over traditional tree-based methods, such as random forests.[4] An advantage of RLT is that it assigns a variable importance (VI) value to each variable at each node, selects the variable with highest VI upon which to split, and mutes those with the

smallest VI. VI allows for the identification of the factors which most differentiate between subgroups. RLT was performed using primarily default settings from the RLT package in R, Version 3.4.2., with the exception of using 500 trees for stability, and permitting linear combination splits of up to two variables.

Clustering Methods:
Instead of using the entire covariate space which could be computationally expensive, we thus lowered the dimension to 5 where we believe the following conditional joint distribution of the two outcomes (assuming multivariate normal distribution of the two outcomes) is sufficient to represent the clinical situation of patient features. We define the measure $\hat{U}_i$ for each individual $i = 1, \ldots, n$ as follows.

$$\hat{U}_i = \begin{bmatrix} \hat{U}_1 \\ \hat{U}_2 \\ \hat{U}_3 \\ \hat{U}_4 \\ \hat{U}_5 \end{bmatrix}_i = \begin{bmatrix} \widehat{E}[Y_1|X]/\sigma_1 \\ \widehat{E}[Y_2|X]/\sigma_2 \\ \sqrt{\widehat{Var}[Y_1|X]}\big/\sigma_1 \\ \sqrt{\widehat{Var}[Y_2|X]}\big/\sigma_2 \\ \widehat{Corr}_Z[Y_1, Y_2|X] \end{bmatrix}_i, \qquad \text{(Eq. 1)}$$

where $\sigma_k = \frac{1}{n-1}\sum_{i=1}^{n}(\widehat{U}_{i,k} - \widehat{\overline{U}}_k)'$ for $k = 1, 2$, $\widehat{\overline{U}}_k$ is the average of $\hat{U}_{i,k}$'s for all $i = 1, \ldots, n$, and $\widehat{Corr}_Z$ is the Fisher's z-transformation on the correlation, which is calculated from the covariance estimate.

We standardized the means, variances, and covariance so that they are all comparable on the same unit scale; more specifically, we standardized by scaling over the standard deviation and then take Fischer's z-transformation on the correlation calculated from the covariance.

Because the values in $\hat{U}$ were standardized to the same scale, we used Euclidean distance measure to determine the dissimilarity between individuals i and j, for $i, j = 1, \ldots, n$:

$$D_{ij} = \sqrt{(\hat{U}_i - \hat{U}_j)' \, (\hat{U}_i - \hat{U}_j)} \qquad \text{(Eq. 2)}$$

This measure directly informs how far two individuals are based on their outcomes, denoised by the X-variables.

A hierarchical clustering algorithm with Ward's D2 method and Euclidean distance was applied to the standardized $\hat{U}$. The number of clusters was chosen using the NbClust package in R,[5] which takes a vote from 30 methods for choosing number of clusters, including commonly used methods such as gap statistics and average silhouette. The algorithm was restricted to considering clusters between 4 and 9 to characterize the wide range of BMIz and HbA1c at the SEARCH cohort visit but avoid overfitting issues or obscure clinical interpretation. The smallest cluster was restricted to contain at least 50 individuals for adequate statistical power to detect differences in characteristics between groups. Multiple members of the analysis team provided judgment using all the available information.

Clustering Stability
To assess clustering stability, the analysis was repeated sequentially omitting individuals from the same cluster one cluster at a time and observing the distribution of remaining individuals into five clusters. In other words, all individuals from the first cluster are removed, and the analysis is repeated, clustering the remaining individuals into five clusters. Individuals from the first cluster are then brought back in while individuals from the second cluster are removed, and the analysis is repeated again, and so on. For each iteration, the Adjusted Rand Index (ARI) was reported as a measure of clustering stability.[6] ARI measures how similar two clustering methods are, correcting for chance. The mean ARI observed from these six analyses was 0.785±0.05. We interpret the limited variation and a high mean ARI as evidence that our identified clusters are sufficiently stable; the cluster assignments are not sensitive to the removal of other clusters.

Adjustments for Multiple Comparisons

For each characterizing variable, pairwise comparisons between each cluster and the referent cluster (Cluster 1) were carried out using t-tests and chi-squared or Fisher's exact tests, where appropriate. To control for the positive False Discovery Rate (pFDR)[9] associated with the pairwise comparisons, appropriate adjustments were made separately to continuous and categorical variables, with an additional Bonferroni correction to account for the two categories (continuous and categorical). q-values reported in **Supplementary Tables S2-S4** can be considered as "posterior Bayesian p-values," or the posterior probability that the null hypothesis is falsely rejected. q-values have been reported in place of p-values, because q-values control for the pFDR whereas p-values control for the Family Wise Error Rate (FWER), or the probability that at least one false rejection has been made. q-values were evaluated at the significance level of 0.05.

Additional Analyses

<u>Comparison to clusters based on raw, observed measures of BMIz and HbA1c</u>

RLT estimates of the outcomes were selected to capture the joint distribution of and provide smoothed outcome measures informed by the X-variables, as each individual is expected to exhibit some level of within-patient heterogeneity; the smoothed outcomes maintain the individual level signal with reduced noise.[7] Additional analyses to test the validity of clustering methodology explored clustered subgroups based on the raw, i.e. observed, outcome measures Y1 (BMIz) and Y2 (HbA1c). These clusters were denoted as "Y-Clusters" and are depicted in **Figure S1.** Compared to the weight-glycemia cluster, these clusters showed multiple density nodes for BMIz and HbA1c within clusters, as well as a representation of all outliers within a single cluster (**Figure 2B).** Although the Y-clusters based on the raw outcome measures displayed significantly different mean measures of BMIz and HbA1c, the clusters showed a larger within-group distribution of BMIz and HbA1c measures (**Supplementary Table S5**). Together, this analysis suggested that noise in the raw outcome variables may obscure the true subgroups of interest.[8]

<u>Age-stratified analyses</u>

To facilitate study across youth and young adults, BMIz for individuals ≥ 20 years was estimated assuming an age of 20 years (the maximum age represented in the growth reference); this approach has been operationalized in previous SEARCH studies[10,11] and elsewhere.[12] Given known challenges in the use of BMIz and in the context of the present analysis, further analyses were undertaken to assess whether the use of BMIz may bias the nature of the clusters. To check the validity of the imputed z-scores and assess for possible differential bias in the results by age (i.e. youth versus young adults), we stratified the sample by age at follow-up visit (<21 years, n=1,399, ≥21 years, n=418) and independently evaluated clusters in each sample. The number of clusters was chosen using the NbClust package in R[5] and restricted to considering between four and nine clusters. Clusters across age strata were compared for consistency in BMIz and HbA1c. We found six clusters in the Under 21 Years stratum and five clusters in the 21 Years and Older stratum (**Supplementary Table S6, Supplementary Figure S2**). No evidence of differential bias from BMIz was found; the resulting weight-glycemia phenotypes were largely consistent in the stratified samples, where Clusters 3 and 4 in the Under 21 stratum merged to form one aggregated cluster (Cluster 3) in the 21 and Older stratum. Cluster 3 and 4 merged among the ≥21-year-old strata to show one combined cluster (normal weight with poor-very poor glycemic control); this result likely reflects increases in HbA1c known to occur around 17 years of age and last through a mean of 30 years.[13]

Comparison to *a priori* Weight-Glycemia Classifications

An additional, exploratory study used clinical cut-points for BMIz and HbA1c to classify youth and young adults with type 1 diabetes into six weight-glycemia classes and study the proportion and correlates of each subgroup. This analysis was meant to provide context for the cluster analysis, to check the validity of clustered subgroups, and to test if clusters may be useful in gleaning additional insights into the weight-glycemia phenotype of type 1 diabetes.

The study sample from the main cluster analysis was used. Participants were excluded if they were missing a measure of BMIz (n=151) or HbA1c (n=32). A very small proportion of participants were classified as underweight (BMIz <-1.64; ~2%); these participants were excluded to prevent misclassification bias associated with combining subgroups in the analysis.

One and two cut-points were operationalized for weight status and glycemic control, respectively. Weight status was classified as normal weight (BMIz <1.04, corresponding to <85th percentile for age and sex) versus combined overweight/obesity (BMIz ≥1.04, corresponding to ≥85th percentile). Glycemic control was classified as good (HbA1c <58 mmol/mol (<7.5%), moderate (HbA1c 58 - <75 mmol/mol (7.5-<9.0%), and poor (HbA1c 75 mmol/mo] (≥9.0%).[14] Crosstabulation of the cut-points yielded six weight-glycemia classes. Descriptive statistics were used to summarize and compare BMIz, HbA1c, and a subset of sociodemographic and clinical characteristics measured at the cohort visit across subgroups. All analyses used a two-sided p-value of 0.05.

The final sample included 1785 youth and young adults with type 1 diabetes (50% female, 76.1% non-Hispanic white, mean age 17.6±4.5 years, mean diabetes duration 7.8±1.9 years.) The mean BMIz was 0.66±0.87 and the mean HbA1c was 76±21 mmol/mol (9.1±1.8%). Shown in **Supplementary Table S6,** the normal weight subgroup with poor glycemic control represented 1/3 of the sample, comprising the largest weight-glycemia class (Class 1C, 30%). Only 11% of the of SEARCH sample was classified as normal weight with good glycemic control (Class 1A). By contrast, approximately 17% of the sample was classified as overweight or obese with poor glycemic control. The smallest subgroup was overweight or obese with adequate glycemic control (5.6%). The proportion of youth classified as overweight and obese was not significantly different across strata of glycemic control (p=0.60).

There were significant differences in sociodemographic and clinical characteristics across weight-glycemic classes (**Supplementary Table S7**). Compared to subgroup with ideal weight and glycemia (Class 1A), subgroups with poor glycemic control (1C and 2C) had lower parental education, income, and private insurance use; these subgroups reported significantly lower pump use and frequency of glucose monitoring (p <0.001). The overweight/obese subgroup with poor glycemic control (Class 2C) also had the highest proportion of females (66.1% versus 44.3% in Class 1A), non-Hispanic Black youth (18.1% versus 2.5%), and Hispanic youth (16.4% versus 9.9%; all p <0.0001).

These results reinforce profound heterogeneity in the clinical presentation of type 1 diabetes; all degrees of glycemic control are represented in normal weight as well as overweight/obese youth. Second, relatively few youth show appropriate weight and glycemia. Finally, the unequal distribution of socioeconomic position and aspects of clinical care across race/ethnicity is consistent with characteristics of the weight-glycemia clusters.

References

1. Stekhoven DJ, Bühlmann P. MissForest—non-parametric missing value imputation for mixed-type data. *Bioinformatics.* 2011;28(1):112-118.
2. White IR, Royston P, Wood AM. Multiple imputation using chained equations: issues and guidance for practice. *Statistics in medicine.* 2011;30(4):377-399.
3. Gaynor S, Bair E. Identification of relevant subtypes via preweighted sparse clustering. *Computational statistics & data analysis.* 2017;116:139-154.
4. Zhu R, Zeng D, Kosorok MR. Reinforcement Learning Trees. *J Am Stat Assoc.* 2015;110(512):1770-1784.
5. Charrad M, Ghazzali N, Boiteau V, Niknafs A, Charrad MM. Package 'NbClust'. *Journal of Statistical Software.* 2014;61:1-36.
6. Steinley D. Properties of the Hubert-Arable Adjusted Rand Index. *Psychological methods.* 2004;9(3):386.
7. Zhu R, Zeng D, Kosorok MR. Reinforcement learning trees. *Journal of the American Statistical Association.* 2015;110(512):1770-1784.
8. Bair E. Semi-supervised clustering methods. *Wiley Interdisciplinary Reviews: Computational Statistics.* 2013;5(5):349-361.
9. Storey JD. The positive false discovery rate: a Bayesian interpretation and the q-value. *The Annals of Statistics.* 2003;31(6):2013-2035.
10. Dabelea D, Stafford JM, Mayer-Davis EJ, et al. Association of type 1 diabetes vs type 2 diabetes diagnosed during childhood and adolescence with complications during teenage years and young adulthood. *Jama.* 2017;317(8):825-835.
11. Sauder KA, Stafford JM, Mayer-Davis EJ, et al. Co-occurrence of early diabetes-related complications in adolescents and young adults with type 1 diabetes: an observational cohort study. *The Lancet Child & Adolescent Health.* 2019;3(1):35-43.
12. Esbenshade AJ, Simmons JH, Koyama T, Koehler E, Whitlock JA, Friedman DL. Body mass index and blood pressure changes over the course of treatment of pediatric acute lymphoblastic leukemia. *Pediatric blood & cancer.* 2011;56(3):372-378.
13. Miller KM, Foster NC, Beck RW, et al. Current state of type 1 diabetes treatment in the U.S.: updated data from the T1D Exchange clinic registry. *Diabetes Care.* 2015;38(6):971-978.
14. Petitti DB, Klingensmith GJ, Bell RA, et al. Glycemic control in youth with diabetes: the SEARCH for diabetes in Youth Study. *The Journal of pediatrics.* 2009;155(5):668-672. e663.

Supplementary Table S1: 28 *X*-variables for Reinforcement Learning Trees. All measures are from the 5+ year follow-up visit unless specified

Sociodemographic measures	Age at diagnosis, diabetes duration, sex (baseline), race/ethnicity (baseline), parental education attainment, household income level, insurance type
Clinical measures	insulin dose, insulin regimen, frequency of blood glucose monitoring, severe hypoglycemic episodes in the last 6 months, emergency room visits in the last 6 months, hospitalizations in the last 6 months
Anthropometric and laboratory measures	waist circumference, waist to height ratio, systolic and diastolic blood pressure, total cholesterol, non-HDL cholesterol, HDL cholesterol, LDL cholesterol (calculated), triglycerides
Psychosocial and behavioral measures	depressive symptoms, quality of life score, physical activity, smoking status

Supplementary Table S2. q-values[†] for Pairwise Comparisons of Sociodemographic Characteristics According to Weight-Glycemia Phenotype Clusters 1-6

Characteristics	Weight-Glycemia Cluster					
	Cluster 1 vs. Cluster 2	Cluster 1 vs. Cluster 3	Cluster 1 vs. Cluster 4	Cluster 1 vs. Cluster 5	Cluster 1 vs. Cluster 6	Cluster 5 vs. Cluster 6
Weight-Glycemia						
BMIz	4.15E-71	0.144659	5.67E-29	1.76E-21	2.28E-148	1.41E-10
HbA1c (%)	4.88E-16	1.08E-181	3.35E-47	3.68E-75	5.43E-31	1.91E-63
Weight Status[‡]	7.20E-18	0.335682	1.88E-08	2.29E-39	5.89E-112	9.72E-17
Glycemic Control[§]	4.61E-11	6.17E-154	3.78E-62	4.28E-93	2.56E-22	9.65E-40
Age at follow-up (years)	0.024023	0.16736	0.002432	5.69E-04	0.228048	0.004529
Age at diagnosis (years)	0.070913	0.065987	0.006348	0.008597	0.269702	0.007673
Diabetes duration (months)	0.037604	0.075612	0.114815	0.009293	0.037353	0.146216
Female	6.62E-04	4.09E-04	0.016631	2.74E-05	0.008735	0.027848
Race/ethnicity[¶]	0.335682	1.03E-15	1.24E-12	1.20E-21	1.47E-07	2.93E-05
Parental Education	0.388102	1.30E-36	1.54E-22	1.44E-30	2.03E-18	4.09E-04
Household Income	0.249101	1.20E-21	6.11E-19	1.70E-20	1.64E-05	3.33E-06
Insurance type	0.045741	6.97E-18	1.35E-13	1.41E-23	1.90E-06	5.79E-07

Blank cells indicate where the overall test of difference was not statistically significant and no pairwise comparisons were performed.

[†]Controlled for the positive False Discovery Rate (pFDR). q-values can be considered as "posterior Bayesian p-values," or the posterior probability that the null hypothesis is falsely rejected. q-values have been reported in place of p-values, because q-values control for the pFDR. q-values were evaluated at the significance level of 0.05.

[‡]Weight status defined based on body mass index z-score (BMIz). Underweight was defined as cluster mean BMIz <-1.64 corresponding to the 5th percentile for age and sex. Normal weight was defined as cluster mean BMIz ≥-1.64 and <1.04, corresponding to ≥-the 5th and <85th percentile for age and sex. Overweight was defined as cluster mean BMIz ≥1.04 and <1.64, corresponding to ≥85th percentile and <95th percentile for age and sex. Obesity was defined as cluster mean BMIz ≥1.64 corresponding to ≥95th percentile for age and sex.

[§]Glycemic control was based on hemoglobin A1c (HbA1c) and defined as good (mean HbA1c <58 mmol/mol (<7.5%)), moderate (mean HbA1c 58 - <75 mmol/mol (7.5 - <9.0%)), poor (mean HbA1c 75 - <108 mmol/mol (9.0 - <12.0%)), and very poor (mean HbA1c ≥108 mmol/mol (≥12.0%))

[¶]Self-reported race and ethnicity were collected using 2000 U.S. Census questions.

Supplementary Table S3. q-values[†] for Pairwise Comparisons of Diabetes Care, Psychosocial, and Behavioral Factors According to Weight-Glycemia Clusters 1-6

Characteristics	Weight-Glycemia Cluster					
	Cluster 1 vs. Cluster 2	Cluster 1 vs. Cluster 3	Cluster 1 vs. Cluster 4	Cluster 1 vs. Cluster 5	Cluster 1 vs. Cluster 6	Cluster 5 vs. Cluster 6
Diabetes Care Factors						
Insulin Regimen	0.042984	4.47E-16	3.01E-14	7.44E-28	9.96E-04	6.98E-13
Insulin dose (daily units/Kg)	0.185203	1.34E-05	1.14E-04	3.48E-06	0.043848	2.01E-04
Blood Glucose Monitoring Frequency	0.003068	5.60E-25	1.60E-18	5.09E-30	9.06E-04	1.35E-12
Use of CGM	0.335682	0.203666	0.28288	0.393682	0.036694	0.069077
Acute Complications (Past 6 Mo)[‡]						
Severe Hypoglycemic Episodes	--	--	--	--	--	--
Diabetic Ketoacidosis Episodes	0.159907	1.45E-10	1.51E-11	1.13E-18	0.118738	2.70E-10
Diabetes Care Provider	--	--	--	--	--	--
Psychosocial Factors						
Depressive Symptoms (CEDS Score)[§]	0.145573	1.83E-17	1.45E-04	1.40E-13	3.75E-04	1.67E-07
Quality of Life (Peds QL Score)[¶]	0.302841	5.31E-12	1.55E-04	2.79E-11	4.45E-04	2.42E-05
Lifestyle Behavioral Factors						
Adherence to DASH Diet[‡‡]	0.004454	3.19E-11	1.34E-05	3.60E-06	1.34E-05	0.142206
Total Energy Intake (kcal)	0.109875	0.130658	0.034153	0.090996	0.00527	0.002503
Total Energy from Fat (%)	--	--	--	--	--	--
Total Energy from Carbohydrate (%)	--	--	--	--	--	--
Total Energy from Protein (%)	--	--	--	--	--	--
Physically Active[§§]	0.016989	2.66E-09	3.22E-05	2.74E-05	1.82E-06	0.335682
High Screen Time[§§]	0.312454	5.29E-14	1.28E-08	1.51E-11	1.05E-05	0.001483
Smoking Status	0.109654	1.78E-04	2.22E-09	7.16E-04	0.420621	0.001878

Blank cells indicate where the overall test of difference was not statistically significant and no pairwise comparisons were performed.

[†]Controlled for the positive False Discovery Rate (pFDR). q-values can be considered as "posterior Bayesian p-values," or the posterior probability that the null hypothesis is falsely rejected. q-values have been reported in place of p-values, because q-values control for the pFDR. q-values were evaluated at the significance level of 0.05.

[‡]Self-reported, in the past 6 months

[§]Center for Epidemiologic Studies Depression Scale, total score

[¶]Peds QL, total score

[‡‡]Dietary Approach to Stop Hypertension diet, total score

[§§]Physically active defined as exercise 3-7 days per week. High screen time defined as 2+hours of screen-time per day

Supplementary Table S4. q-values[†] for Pairwise Comparisons of Clinical Characteristics According to Weight-Glycemia Phenotype Clusters 1-6

Characteristics	Weight-Glycemia Cluster					
	Cluster 1 vs. Cluster 2	Cluster 1 vs. Cluster 3	Cluster 1 vs. Cluster 4	Cluster 1 vs. Cluster 5	Cluster 1 vs. Cluster 6	Cluster 5 vs. Cluster 6
Lipids						
Total Cholesterol, mg/dL	0.164053	6.18E-31	2.58E-09	1.88E-21	6.14E-14	5.99E-12
HDL Cholesterol, mg/dL	0.205732	0.192699	0.027915	4.94E-06	1.05E-10	0.312837
LDL Cholesterol, mg/dL	0.04247	1.06E-20	4.68E-07	6.99E-18	1.27E-17	4.88E-06
VLDL Cholesterol, mg/dL	0.071374	1.45E-32	1.21E-10	1.55E-19	4.15E-18	5.13E-13
Triglycerides, mg/dL	0.075612	1.25E-32	5.12E-07	1.68E-17	2.72E-18	4.10E-12
Blood Pressure						
Systolic Blood Pressure, mmHg	0.008237	0.211486	0.243499	2.84E-10	1.98E-10	0.155705
Diastolic Blood Pressure, mmHg	0.135977	7.46E-05	0.018749	4.07E-14	5.13E-13	0.021609

Blank cells indicate where the overall test of difference was not statistically significant and no pairwise comparisons were performed.

[†]Controlled for the positive False Discovery Rate (pFDR). q-values can be considered as "posterior Bayesian p-values," or the posterior probability that the null hypothesis is falsely rejected. q-values have been reported in place of p-values, because q-values control for the pFDR. q-values were evaluated at the significance level of 0.05.

Supplementary Table S5. Body mass index z-score (BMIz) and hemoglobin A1c (HbA1c) According to Weight-Glycemia Phenotype Clusters, in the full sample and stratified by age at the follow-up visit (<vs ≥21 years).

All participants	All N=1.817	**Cluster 1** n=618 (34.0%)	**Cluster 2** n=195 (10.7%)	**Cluster 3** n=509 (28.0%)	**Cluster 4** n=79 (4.4%)	**Cluster 5** n=137 (7.5%)	**Cluster 6** n=279 (15.4%)	p-value
BMIz	0.61 (0.94)	0.59 (0.59)	-0.68 (0.65)	0.56 (0.62)	-1.05 (0.83)	1.29 (0.69)	1.74 (0.42)	<0.0001
HbA1c (mmol/mol)	76 (21)	61 (12)	68 (10)	86 (12)	113 (15)	109 (15)	70 (11)	<0.0001
HbA1c (%)	9.1 (1.9)	7.7 (1.1)	8.4 (0.9)	10.0 (1.1)	12.5 (1.4)	12.1 (1.5)	8.6 (1.0)	<0.0001

Participants Under 21 Years	All N=1,399	**Cluster 1** n=377 (27.0)	**Cluster 2** n=104 (7.4%)	**Cluster 3** n=360 (25.7%)	**Cluster 4** n=145 (10.4%)	**Cluster 5** n=136 (9.7%)	**Cluster 6** n=277 (19.8%)	p-value
BMIz	0.60 (0.93)	0.32 (0.59)	-1.00 (0.61)	0.51 (0.5)	-0.01 (0.96)	1.71 (0.39)	1.44 (0.49)	<0.0001
HbA1c (mmol/mol)	77 (20)	62 (10)	70 (13)	83 (12)	113 (16)	89 (13)	67 (11)	
HbA1c (%)	9.2 (1.8)	7.8 (0.9)	8.6 (1.2)	9.7 (1.1)	12.5 (1.5)	10.3 (1.2)	8.3 (1.0)	<0.0001

Participants 21 Years and Older	All N=418	**Cluster 1** n=127 (30.4%)	**Cluster 2** n=96 (23.0%)	**Cluster 3** n=59 (6.0%)		**Cluster 4** n=70 (16.8%)	**Cluster 5** n=66 (15.8%)	p-value
BMIz	0.64 (1.00)	0.90 (0.49)	-0.22 (0.64)	-0.32 (1.12)		1.03 (0.70)	1.80 (0.46)	<0.0001
HbA1c (mmol/mol)	73 (22)	56 (13)	64 (13)	104 (16)		91 (14)	67 (13)	
HbA1c (%)	8.8 (2.0)	7.3 (1.2)	8.0 (1.2)	11.7 (1.5)		10.5 (1.3)	8.3 (1.2)	<0.0001

Supplementary Table S6. Six Weight-Glycemia classifications for youth and young adults with type 1 diabetes based on 1 cut-point for weight status and 2 cut-points for glycemic control

Glycemia

Class, n (overall %) Mean BMIz Mean HbA1c, mmol/mol (%) Row % Column %		A. Adequate (HbA1c <58 mmol/mol (7.5%))	B. Fair (HbA1c ≥58 mmol/mol and <75 mmol/mol (≥7.5% and <9.0%))	C. Poor (HbA1c 75 mmol/mol (≥9.0%))	Total
	1. Normal weight (BMIz <1.04)	**Class 1A,** n=203 (11.4%) Mean BMIz: 0.18±0.58 Mean HbA1c: 51±7 mmol/mol (6.8±0.6%) 17.6% 67.2%	**Class 1B,** n=415 (23.3%) Mean BMIz: 0.18±0.62 Mean HbA1c: 68±4 mmol/mol (8.2±0.4%) 35.9% 64.0%	**Class 1C,** n=537 (30.1%) Mean BMIz: 0.14±0.63 Mean HbA1c: 93±16 mmol/mol (10.7±1.5%) 46.5% 64.3%	1155 (64.7%)
Weight	2. Overweight/Obese (BMIz ≥1.04)	**Class 2A,** n=99 (5.6) Mean BMIz: 01.58±0.38 Mean HbA1c: 49±8 mmol/mol (6.6±0.7%) 15.7%) 32.8%	**Class 2B,** n=233 (13.1%) Mean BMIz: 01.56±0.41 Mean HbA1c: 66±4 mmol/mol (8.2±0.4% 37.0%) 36.0%	**Class 2C,** n=298 (16.7%) Mean BMIz: 1.57±0.39 Mean HbA1c: 91±13 mmol/mol (10.5±1.2%) 47.3% 35.7%	391 (21.9%)
	Total	302 (16.9%)	648 (36.3%)	835 (46.8%)	1785 (100%) Mean BMIz: 0.66±0.87 Mean HbA1c: 76±20 mmol/mol (9.1±1.8)

Abbreviations: BMIz – body mass index z-score. HbA1c – Hemoglobin A1c.Class Nomenclature: 1 versus 2 signifies Normal Weight versus Overweight/Obese. A, B, and C signifies good, moderate, and poor glycemic control, respectively.

Supplementary Table S7. Selected Characteristics According to the Six Weight-Glycemia Classifications.

Characteristics, Mean (SD) or n (%)	Weight-Glycemia Classifications							
	All	Class 1A: Nw, Good glycemic control	Class 1B: Nw, Moderate glycemic control	Class 1C: Nw, Poor glycemic control	Class 2A: Ow/Ob, Good glycemic control	Class 2B: Ow/Ob, Moderate glycemic control	Class 2C: Ow/Ob, Poor glycemic control	p-value[†]
	N=1785	N=203	N=415	N=537	N=99	N=233	N=298	
Weight and Glycemia Measures								
BMIz	0.66 (0.87)	0.18 (0.58)	0.18 (0.62)	0.14 (0.63)	1.58 (0.38)	1.56 (0.41)	1.57 (0.39)	<0.0001
HbA1c, %	9.1 (1.8)	6.8 (0.6)	8.2 (0.4)	10.7 (1.5)	6.6 (0.7)	8.2 (0.4)	10.5 (1.2)	<0.0001
Demographic Characteristics								
Female	892 (50.0)	90 (44.3)	175 (42.2)	263 (49.0)	42 (42.4)	125 (53.7)	197 (66.1)	<0.0001
Age at Cohort Visit, years	17.6 (4.5)	18.7 (4.9)	16.2 (4.8)	17.8 (4.2)	20.5 (4.7)	17.3 (4.4)	17.7 (3.8)	<0.0001
Age at Diagnosis, years	9.8 (4.1)	11.0 (4.4)	8.5 (4.1)	9.9 (4.0)	12.1 (4.1)	9.5 (4.1)	9.8 (3.5)	<0.0001
Diabetes Duration, months	93.2 (22.8)	92.3 (23.56)	89.9 (22.5)	94.8 (23.0)	100.6 (23.1)	92.8 (21.3)	93.6 (22.8)	0.0004
Race/ethnicity								<0.0001
Non-Hispanic White	1358 (76.1)	172 (84.7)	356 (85.8)	384 (71.5)	73 (73.7)	184 (79.0)	189 (63.4)	
Non-Hispanic Black	166 (9.3)	5 (2.5)	14 (3.4)	67 (12.5)	10 (10.0)	16 (6.9)	54 (18.1)	
Hispanic	218 (12.2)	20 (9.9)	38 (9.2)	71 (13.2)	11 (11.1)	29 (12.5)	49 (16.4)	
Socioeconomic Position								
Parental Bachelor's degree or more	932 (52.0)	154 (76.6)	255 (61.9)	229 (44.0)	54 (54.6)	122 (52.6)	118 (40.4)	<0.0001
Household Income >$75,000	668 (15.3)	92 (45.5)	199 (48.0)	149 (27.9)	33 (33.3)	98 (42.4)	97 (32.8)	<0.0001
Private Health insurance	1278 (72.0)	168 (82.8)	328 (79.6)	336 (63.2)	77 (77.8)	176 (75.9)	193 (65.2)	<0.0001
Diabetes Care Factors								
Insulin pump use (versus multiple daily injections)	1009 (57.6)	137 (70.3)	279 (67.6)	243 (45.8)	55 (63.2)	142 (61.2)	153 (52.0)	<0.0001
Blood Glucose Monitoring >4x/day	1158 (66.9)	158 (80.6)	328 (80.8)	280 (53.6)	58 (65.2)	170 (74.2)	164 (56.8)	<0.0001
1+ Severe Hypoglycemia[‡]	130 (7.3)	18 (8.9)	27 (6.5)	30 (5.6)	12 (12.1)	26 (11.2)	17 (5.7)	0.0235
1+ Recent Diabetic Ketoacidosis Episode[c]	324 (18.2)	15 (7.4)	61 (14.7)	145 (27.2)	13 (13.1)	26 (11.2)	64 (21.5)	<0.0001

Abbreviations: Nw - normal weight. Ow/Ob – overweight and obese. SD – standard deviation; BMIz – body mass index z-score; HbA1c – hemoglobin A1c

[†] P-value for overall test of difference, based on use of ANOVA, Chi-squared, of Fishers Exact test as appropriate.

[‡] Self-reported, in the past 6 months. DKA is an acute complication of hyperglycemia.

Supplementary Table S8. Measures of Weight and Glycemic Control According to Y-Clusters 1-6

| Characteristics, Mean (SD) or n (%)	All N=1817	Cluster A n=60 (3.3%)	Cluster B n=166 (9.2%)	Cluster C n=806 (44.4%)	Cluster D n=316 (4.4%)	Cluster E n=301 (17.4%)	Cluster F n=168 (9.3%)	p-value[†]
Weight-Glycemia								
BMIz	0.61 (0.94)	-0.28 (0.86)	1.33 (0.56)	0.84 (0.72)	0.00 (0.81)	1.16 (0.60)	-0.81 (0.61)	<0.001
HbA1c (%)	9.1 (1.9)	14.0 (1.1)	11.8 (0.8)	7.6 (0.9)	10.5 (0.8)	9.4 (0.4)	8.4 (0.63)	<0.001
Weight Status[‡]								<0.001
Underweight	36 (2.0)	5 (8.3)	0 (0.0)	0 (0.0)	13 (4.1)	0 (0.0)	18 (10.7)	
Normal Weight	1152 (63.4)	52 (88.7)	54 (32.5)	491 (60.9)	285 (90.2)	120 (38.9)	150 (89.3)	
Overweight	390 (21.5)	2 (3.3)	63 (38.0)	189 (23.5)	18 (5.7)	118 (39.2)	0 (0.0)	
Obese	239 (13.2)	5 (8.3)	0 (0.0)	0 (0.0)	13 (4.1)	0 (0.0)	18 (10.7)	
Glycemic Control[§]								<0.001
Good	306 (16.8)	0 (0.0)	0 (0.0)	292 (36.2)	0 (0.0)	0 (0.0)	14 (8.3)	
Moderate	656 (36.1)	0 (0.0)	0 (0.0)	514 (63.8)	0 (0.0)	27 (9.0)	115 (68.5)	
Poor	704 (38.8)	0 (0.0)	93 (56.0)	0 (0.0)	298 (94.3)	274 (91.0)	39 (23.2)	
Very Poor	151 (8.3)	60 (100.0)	73 (44.0)	0 (0.0)	18 (5.7)	0 (0.0)	0 (0.0)	

Y-clusters were generated based on the raw, observed measures of outcomes Y1 (BMIz) and Y2 (HbA1c).

[†]Bonferroni-corrected p-value for overall test of difference, based on use of ANOVA, Chi-squared or Fisher's exact test as appropriate.

[‡]Weight status defined based on body mass index z-score (BMIz). Underweight was defined as cluster mean BMIz <-1.64 corresponding to the 5th percentile for age and sex. Normal weight was defined as cluster mean BMIz ≥-1.64 and <1.04, corresponding to ≥-the 5th and <85th percentile for age and sex. Overweight was defined as cluster mean BMIz ≥1.04 and <1.64, corresponding to ≥85th percentile and <95th percentile for age and sex. Obesity was defined as cluster mean BMIz ≥1.64 corresponding to ≥95th percentile for age and sex.

[§]Glycemic control was based on hemoglobin A1c (HbA1c) and defined as good (mean HbA1c <58 mmol/mol (<7.5%)), moderate (mean HbA1c 58 - <75 mmol/mol (7.5 - <9.0%)), poor (mean HbA1c 75 - <108 mmol/mol (9.0 - <12.0%)), and very poor (mean HbA1c ≥108 mmol/mol (≥12.0%))

Supplementary Figure S1: Box and Whisker plot for BMIz and HbA1c of the age-stratified weight-glycemia phenotypic clusters from the SEARCH for Diabetes in Youth Study. Participants were stratified by age at the cohort visit (<21 and ≥21 years, i.e. 21 years and over) and clustered based on the joint distribution of body mass index z-score (BMIz) and hemoglobin A1c (HbA1c) at the 5+ year cohort visit of the SEARCH study. For the 21 Years and Over Strata, five clusters were modeled.
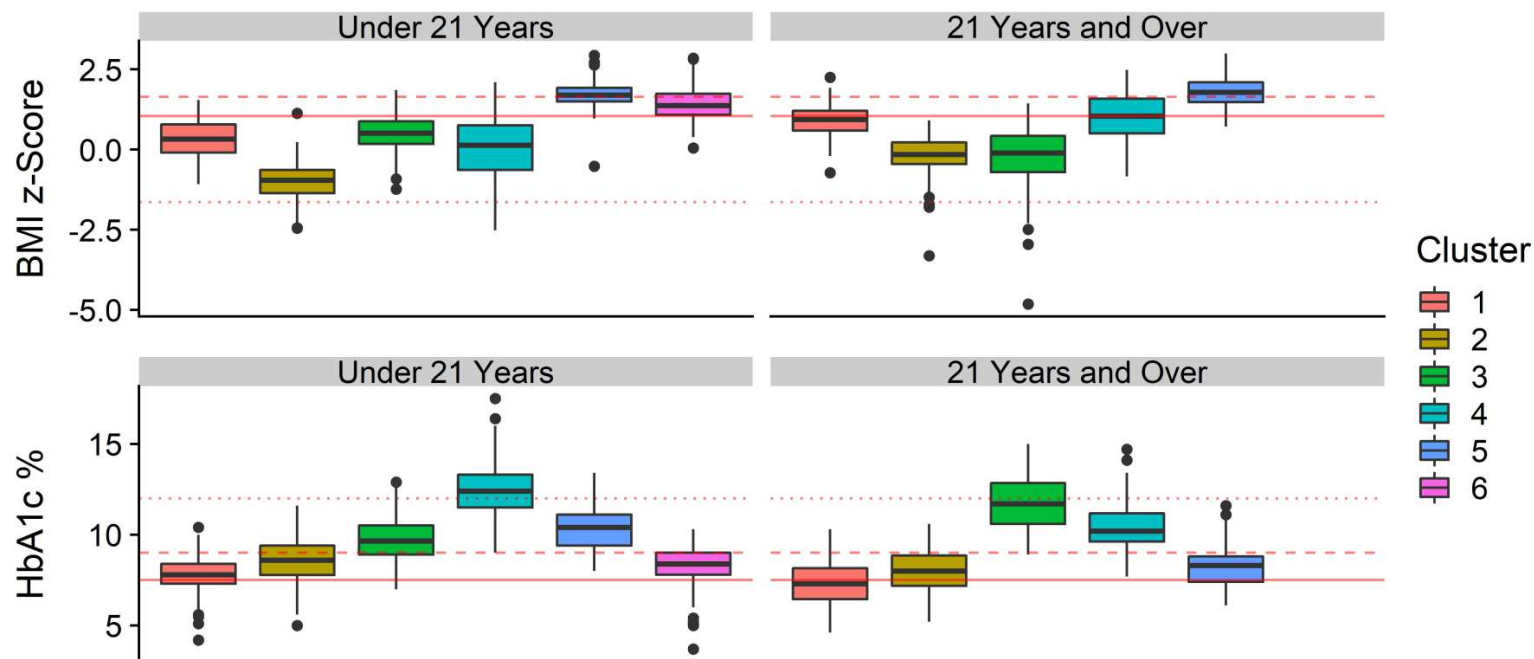
Supplementary Figure S2: Density distribution plots of Body Mass Index z-score (BMIz) and Hemoglobin A1c (HbA1c) for Y-clusters, based on observed or raw measures of BMIz and HbA1c. A: Density distribution of BMIz by Y-cluster. B: Density distribution of HbA1c by Y-cluster. 3: Density distribution plot of BMIz and HbA1c by Y-cluster. Ideal clustered subgroups should show distinct, unimodal density distributions. The area under each cluster's curve integrates to 1.