



Trajectories of clinical characteristics, complications and treatment choices in data-driven subgroups of type 2 diabetes

Xinyu Li¹ · Louise A. Donnelly² · Roderick C. Sliker^{3,4,5} · Joline W. J. Beulens^{4,5,6} · Leen M. 't Hart^{3,4,5,7} · Petra J. M. Elders^{5,8} · Ewan R. Pearson² · Anoukh van Giessen⁹ · Jose Leal¹⁰ · Talitha Feenstra^{1,9}

Received: 29 November 2023 / Accepted: 12 February 2024
© The Author(s) 2024

Abstract

Aims/hypothesis This study aimed to explore the added value of subgroups that categorise individuals with type 2 diabetes by *k*-means clustering for two primary care registries (the Netherlands and Scotland), inspired by Ahlqvist's novel diabetes subgroups and previously analysed by Sliker et al.

Methods We used two Dutch and Scottish diabetes cohorts ($N=3054$ and 6145 ; median follow-up= 11.2 and 12.3 years, respectively) and defined five subgroups by *k*-means clustering with age at baseline, BMI, HbA_{1c}, HDL-cholesterol and C-peptide. We investigated differences between subgroups by trajectories of risk factor values (random intercept models), time to diabetes-related complications (logrank tests and Cox models) and medication patterns (multinomial logistic models). We also compared directly using the clustering indicators as predictors of progression vs the *k*-means discrete subgroups. Cluster consistency over follow-up was assessed.

Results Subgroups' risk factors were significantly different, and these differences remained generally consistent over follow-up. Among all subgroups, individuals with severe insulin resistance faced a significantly higher risk of myocardial infarction both before (HR 1.65; 95% CI 1.40, 1.94) and after adjusting for age effect (HR 1.72; 95% CI 1.46, 2.02) compared with mild diabetes with high HDL-cholesterol. Individuals with severe insulin-deficient diabetes were most intensively treated, with more than 25% prescribed insulin at 10 years of diagnosis. For severe insulin-deficient diabetes relative to mild diabetes, the relative risks for using insulin relative to no common treatment would be expected to increase by a factor of 3.07 (95% CI 2.73, 3.44), holding other factors constant. Clustering indicators were better predictors of progression variation relative to subgroups, but prediction accuracy may improve after combining both. Clusters were consistent over 8 years with an accuracy ranging from 59% to 72%.

Conclusions/interpretation Data-driven subgroup allocations were generally consistent over follow-up and captured significant differences in risk factor trajectories, medication patterns and complication risks. Subgroups serve better as a complement rather than as a basis for compressing clustering indicators.

Keywords Data-driven subgroups · Longitudinal analysis · Real-world data · Routine care · Stratification of diabetes

Abbreviations

AIC	Akaike's information criterion
AMI	Acute myocardial infarction
CHF	Congestive heart failure
CKD	Chronic kidney disease
DBP	Diastolic BP
DCS	Hoorn Diabetes Care System
ESRD	End-stage renal disease

GoDARTS	Genetics of Diabetes Audit and Research in Tayside Scotland
GP	General practitioner
MARD	Mild age-related diabetes
MOD	Mild obesity-related diabetes
OAD	Oral antidiabetic drug
PVD	Peripheral vascular disease
RHAPSODY	Risk Assessment and ProgreSsiOn of Diabetes project
RHAP-MD	Mild diabetes subgroup developed by RHAPSODY
RHAP-MDH	Mild diabetes with high HDL-cholesterol subgroup developed by RHAPSODY

Jose Leal and Talitha Feenstra share last authorship.

Extended author information available on the last page of the article

Research in context

What is already known about this subject?

- A novel data-driven subgroup analysis was proposed to explore the heterogeneity of type 2 diabetes in 2018, but little is known about its clinical value
- Most studies replicated the data-driven subgroups in different cohorts with various clustering indicators, but few contained efforts to reveal subgroups' stability, long-term trajectories of risk factors and treatment patterns, and none of the studies considered the added value of subgroup information on clustering indicators

What is the key question?

- Are data-driven subgroups distinct in characteristics and what are their added values compared with clustering indicators?

What are the new findings?

- Using Dutch and Scottish diabetes cohorts, we showed significant differences among baseline subgroups in the progression of risk factors, treatment patterns and risks of complications over at least 15 years
- Combining both clustering indicators and subgroups may offer improved prediction compared with either approach alone, highlighting the additional value of subgroups in complementing clustering indicators
- Subgroups generally remained consistent over 8 years, with RHAP-SIRD (severe insulin-resistant diabetes) being most stable and RHAP-SIDD (severe insulin-deficient diabetes) least stable

How might this impact on clinical practice in the foreseeable future?

- Novel subgroups might help researchers understand the pathology of type 2 diabetes, but subgroups serve best as a complement rather than as a replacement or compression of clustering indicators

RHAP-MOD	Mild obesity-related diabetes developed by RHAPSODY
RHAP-SIDD	Severe insulin-deficient diabetes developed by RHAPSODY
RHAP-SIRD	Severe insulin-resistant diabetes developed by RHAPSODY
RL	Relative likelihood
SBP	Systolic BP
SIDD	Severe insulin-deficient diabetes
SIRD	Severe insulin-resistant diabetes

cohorts, and their risk profiles studied in both the short and medium term (5 to 15 years) [2–7]. The findings suggest distinct risks of complications and molecular profiles across the subgroups [1–6, 8]. For example, SIRD had a higher frequency of non-alcoholic fatty liver disease and higher risk of developing chronic kidney disease (CKD) [1], and subgroups may help to identify underlying molecular mechanisms related to liver [8], which may provide insights into the diverse aetiology of diabetes.

As part of the Risk Assessment and ProgreSsiOn of Diabetes project (RHAPSODY, <https://imi-rhapsody.eu>), a new set of risk subgroups clustered based on clinical parameters were defined using Dutch and Scottish diabetes registry data and the original Swedish cohort of individuals with type 2 diabetes [9]. Given that the data originated from routine care, some clinical parameters were slightly modified due to their availability [9]. Replication analyses showed good resemblance between cohorts and also compared with the original Swedish subgroups (developed by Ahlqvist et al [1]) [9, 10], except for the refinement of the original MARD cluster into two new clusters, the mild diabetes subgroup developed by RHAPSODY (RHAP-MD) and the mild diabetes with high HDL-cholesterol subgroup developed by RHAPSODY (RHAP-MDH), following the addition of HDL-cholesterol. Both RHAP-MD and RHAP-MDH

Introduction

Data-driven clustering analysis has been proposed for categorising type 2 diabetes based on six clinical parameters: age, BMI, HbA_{1c}, GAD antibodies and HOMA-2 estimates of beta cell function and insulin resistance [1]. In the study by Ahlqvist et al [1], Swedish individuals with diabetes were stratified into five subgroups, including severe autoimmune diabetes, severe insulin-deficient diabetes (SIDD), severe insulin-resistant diabetes (SIRD), mild obesity-related diabetes (MOD) and mild age-related diabetes (MARD) [1]. These subgroups were reproduced in other countries and

exhibited slow glycaemic deterioration, but they showed significantly different molecular signatures [8].

Hence, following up on prior RHAPSODY subgroup research, the current study aims to gain more insight into the clinical relevance of subgroups by studying up to 23 years of follow-up data in two of the original RHAPSODY cohorts. Using contemporary cohorts and a significantly longer follow-up than previous studies, we wanted to: (1) estimate risk factor progression, time to macrovascular complications and treatment patterns by baseline subgroup over at least 15 years; (2) explore the added value of using data-driven subgroups compared with clustering indicators in predicting the progression of risk factors, risk of complications or treatment patterns; and (3) examine the consistency of membership to the data-driven diabetes subgroups over time. Using two distinct cohorts allowed us to validate our findings.

Methods

Study design and participants This retrospective study investigated 9199 individuals with type 2 diabetes in two distinct cohorts: the Hoorn Diabetes Care System (DCS, the Netherlands) and the Genetics of Diabetes Audit and Research in Tayside Scotland (GoDARTS, Scotland). The reporting of study findings followed the STrengthening the Reporting of OBservational studies in Epidemiology (STROBE) guidelines [11], as listed in the electronic supplementary material (ESM) Appendix 1.

Our study's inclusion criteria consisted of a diagnosis age ≥ 35 , GAD negativity and the availability of complete data for each of the five clustering indicators within 2 years of diagnosis. By omitting the data availability requirement for genome-wide association used in the previous RHAPSODY clustering study [9], we employed more lenient criteria, yielding a slightly larger sample size compared with Slieker et al [9].

The DCS cohort consisted of 3054 individuals (median follow-up=11.2 years) observed over the period 1998–2019 and the GoDARTS cohort consisted of 6145 individuals (median follow-up=12.3 years) over the period 2003–2018 that matched the inclusion criteria (ESM Fig. 2.1). All results were produced for both cohorts, separately.

DCS is a comprehensive dynamic prospective cohort of the natural course of type 2 diabetes from 103 general practitioners (GPs) in the West-Friesland region of the Netherlands, with over 90% of its participants being of European ancestry [12]. At baseline, 52.3% of the participants were men, with a mean age of 63 years. Educational levels varied among participants: 43.3% had a low educational level, 42.1% had a middle educational level and 14.6% had a high educational level [12]. DCS generally represents a Western European, semi-urban population [12]. GoDARTS is a longitudinal cohort that includes individuals with diabetes from the Tayside region of

Scotland, with more than 99% of its participants being white [13]. At baseline, 53.3% of the participants were men, with a mean age of 64 years [13]. GoDARTS generally represents a predominantly white population with diabetes in the East of Scotland [13]. Pseudonymised data were collected through electronic record linkage from primary and secondary care data sources [13]. Laboratory measurements of both cohorts have been described in detail in previous studies [9, 12–14] (ESM Appendix 2).

Outcomes and medications Macrovascular and microvascular outcomes, including acute myocardial infarction (AMI), congestive heart failure (CHF), peripheral vascular disease (PVD), stroke, CKD and end-stage renal disease (ESRD), were included in this study (ESM Table 2.1).

Medication use was categorised into treatment steps (ESM Table 2.2). These were defined according to the management steps described in the Dutch GP primary care guideline [15], as the relevant guidance for DCS practitioners at the time of data collection, adding information regarding the use of statins and other medication for CVD prevention.

Clustering Clustering was done on scaled clustering indicators at baseline, including age at baseline, BMI, HbA_{1c}, C-peptide (as a proxy of HOMA-2 estimates of beta cell function and insulin resistance in the absence of fasting glucose in GoDARTS [9]) and HDL-cholesterol (as a risk factor for time to insulin requirement [16]). The baseline for each individual was defined as the observation nearest to diabetes diagnosis. Therefore, it should largely reflect individuals who were either untreated or who only received first line treatment for a brief period (details in Table 1). Men and women were clustered separately and then pooled to avoid sex-dependent differences. Cluster centres were defined as the arithmetic mean of all the values belonging to the cluster. Once clusters were defined, we assigned the same cluster names as those in the original study [1, 9], based on the distribution of cluster characteristics and the lowest Euclidean distance from the previous study [9], including severe insulin-deficient diabetes developed by RHAPSODY (RHAP-SIDD; characterised by high HbA_{1c}), severe insulin-resistant diabetes developed by RHAPSODY (RHAP-SIRD; characterised by high C-peptide and age at baseline), mild obesity-related diabetes developed by RHAPSODY (RHAP-MOD; characterised by high BMI), RHAP-MD (characterised by moderate risk factors) and RHAP-MDH (characterised by high HDL-cholesterol) [9].

Statistical analysis Subgroups identified at baseline were compared with the previously published RHAPSODY subgroups [9], considering the latter as the reference. The agreement was assessed based on sensitivity, specificity, specific agreement [17], overall accuracy rate along with a 95% CI and overall κ indices of agreement [18].

Missing data (mean of 0.6% in DCS and 8.1% in GoDARTS; ESM Table 2.3) were omitted in their respective analyses to avoid excessive use of imputed data as observational evidence.

We reported baseline characteristics for each subgroup using frequencies (%) for categorical variables or mean (SD) for continuous variables. Trajectories of related clinical parameters (BMI, HbA_{1c}, HDL-cholesterol, systolic BP [SBP], diastolic BP [DBP], total cholesterol, LDL-cholesterol, blood creatinine and triglycerides) were visualised by plotting subgroup annual means, along with 1 SD boundaries based on observed variance within subgroups. The random intercept model was used to analyse longitudinal trajectory data with discrete subgroup membership, sex and diabetes duration as covariates.

Kaplan–Meier methods were applied to plot cumulative incidence for first events of each outcome since diagnosis of diabetes by subgroups. Group comparisons and pairwise comparisons were conducted by logrank tests, applying Benjamini–Hochberg correction [19] to adjust for multiple comparisons. A Cox regression model with diabetes duration as the time scale, left truncated at each individual's diagnosis of diabetes, was conducted to calculate the HR (95% CI). The Cox model was also adjusted for age at baseline and sex. Schoenfeld tests were applied to evaluate the proportional hazard assumption, and violation was indicated by $p < 0.05$ [20].

We visualised medication patterns reflecting the proportion of individuals within each subgroup in each treatment step over the follow-up period by area graphs. Multinomial logistic regression, in which treatment steps were dependent variables, with discrete subgroup membership, diabetes duration and sex as covariates, was conducted to compare the proportion in each treatment step between subgroups.

The models described above were re-estimated using clustering indicators at baseline (HbA_{1c}, C-peptide, HDL-cholesterol, age and BMI), with and without discrete subgroup membership data, to analyse the longitudinal risk factor trajectories, risk of complications and medication patterns. Akaike's information criterion (AIC) and relative likelihood (RL) were applied to compare the information loss and fitting of models [21]. Smaller AIC values indicate better goodness of fit. The p value for the comparison of AIC differences was then indicated by $RL = \exp\left(\frac{AIC_{min} - AIC}{2}\right)$. We visualised the results on a heatmap, using colours to indicate scaled AIC and text to indicate RL.

Two clustering algorithms were repeated with durations of 2–4, 4–6 and 4–8 years from diagnosis to assess the cluster consistency over time as follows: (1) de novo clustering (i.e. repeating k -means clustering); or (2) centre-based reallocation (i.e. assigning individuals to the subgroup with the lowest Euclidean distance to cluster centres identified at baseline). The agreement between estimated subgroups over time and subgroups identified at baseline was assessed, and

the cluster migration pattern was presented graphically for individuals with available clustering indicators in all four 2 year intervals (GoDARTS $n=4914$; DCS $n=2756$), along with the top ten transition trajectories. An analysis of the associated risk factors and treatment patterns was visualised in the same manner for the most representative movements. We used the Cox regression model to compare the risk of complications for those who moved between severe subgroups (including RHAP-SIRD and RHAP-SIDD) and mild subgroups (including RHAP-MD, RHAP-MOD and RHAP-MDH).

All analyses were performed in R [22] (version 4.1.0: <https://www.r-project.org/>) and R studio (version 1.4.1717: <https://www.rstudio.com/>) (ESM Appendix 3).

Results

Baseline characteristics and the progression of clinical parameters over time Our current subgroups identified at baseline, which were based on a larger sample size of individuals than previously published RHAPSODY subgroups [9] (2953 individuals in DCS), showed a good resemblance with an accuracy of 0.92 (95% CI 0.91, 0.93) (ESM Table 4.1), despite a slight change in clustering centroids (ESM Table 4.2).

Significant differences in baseline clustering indicators, treatment patterns and other clinical parameters were observed among subgroups identified at baseline (Table 1, ESM Figs 4.1–4.4).

Figure 1 and ESM Fig. 4.5 show that the ranking of risk factors across baseline subgroups remained relatively unchanged throughout follow-up for those risk factors used to characterise specific subgroups (e.g. the subgroup characterised by high HDL-cholesterol at baseline recorded the highest mean HDL-cholesterol during follow-up). The exception was for the trajectory of HbA_{1c} as observed in GoDARTS (Fig. 1b), where the RHAP-MOD subgroup crossed with the RHAP-SIDD subgroup after 4 years from diagnosis and became the subgroup with the highest mean HbA_{1c}. Random intercept models (ESM Table 4.3) indicated that subgroups' properties over time are not only visually distinct but also statistically significantly different. Specifically, compared with RHAP-SIDD, RHAP-SIRD had significantly higher creatinine (an average difference of 12.65 $\mu\text{mol/l}$ across the two cohorts) and RHAP-MDH had significantly lower triglyceride (an average difference of 0.48 mmol/l).

Diabetes-related complications by subgroup The risks of developing AMI, CHF, stroke, CKD and ESRD are significantly different across all subgroups (ESM Figs 5.1, 5.2). At 10 years after diagnosis, the RHAP-SIRD subgroup had the

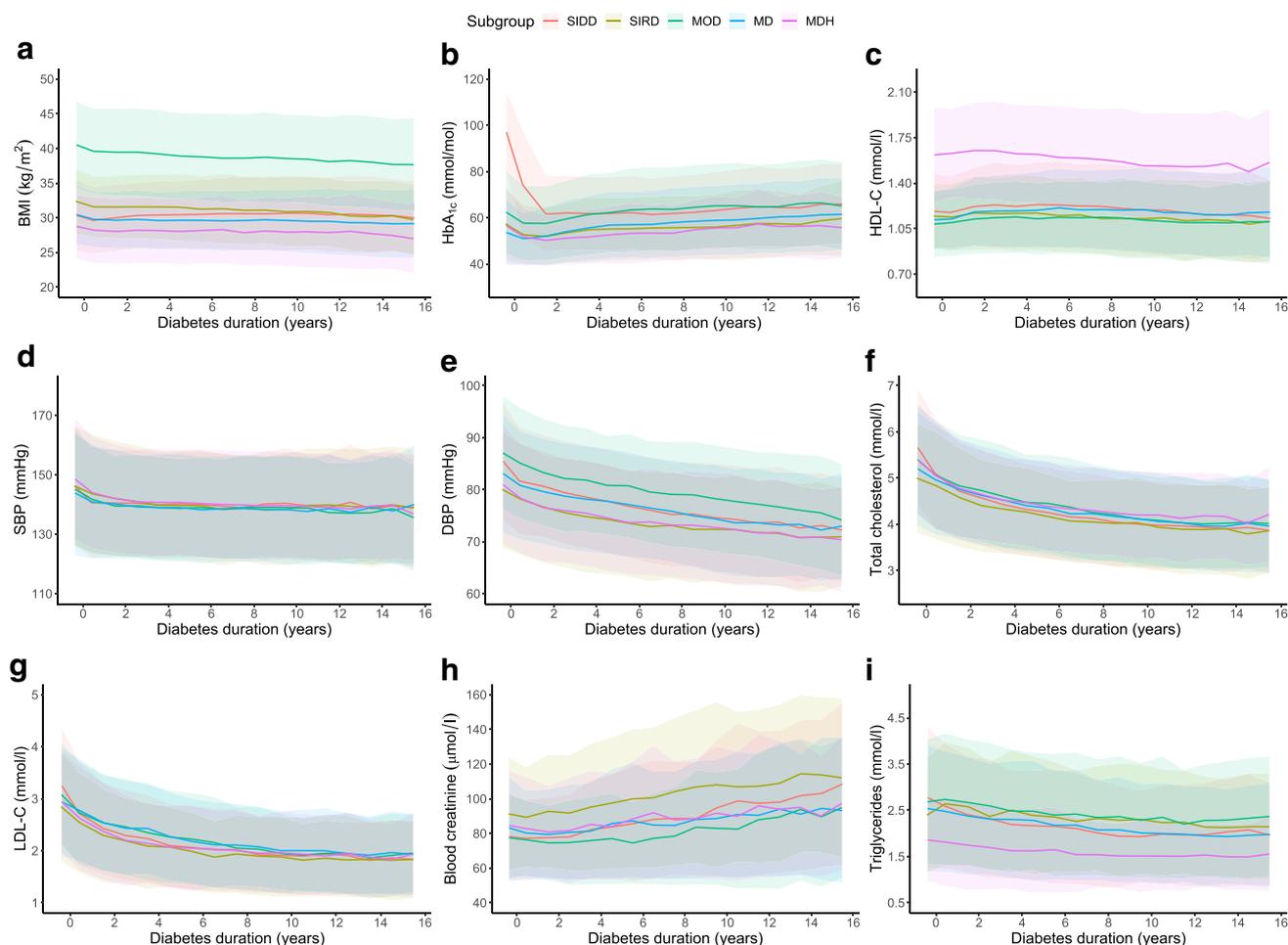


Fig. 1 Progression of clinical parameters over time based on subgroups identified at diagnosis in the GoDARTS cohort. Here, SIDD, SIRD, MOD, MD and MDH refer to RHAP-SIDD, RHAP-SIRD, RHAP-MOD, RHAP-MD and RHAP-MDH, respectively. HDL-C

and LDL-C refer to HDL-cholesterol and LDL-cholesterol, respectively. Values of selected parameters over time in each cluster are shown. The data are represented as mean values (solid line) \pm SD (shaded areas). Missing values were removed

highest incidence of AMI (28.43%; 5.12%), CHF (10.66%; 7.87%), PVD (1.78%; 1.48%) and CKD (75.32%; 57.08%) and RHAP-MDH had the highest incidence of stroke (9.45%; 5.32%) in both GoDARTS and DCS.

Proportional hazard assumptions are fulfilled except for CHF and CKD in GoDARTS (ESM Table 5.1). In both GoDARTS and DCS, Cox models (Fig. 2, ESM Fig. 5.3) indicated that compared with RHAP-MDH, HRs for RHAP-SIRD were significantly different ($p < 0.05$) and the highest among all other subgroups for AMI. Although RHAP-MDH is the so-called mild subgroup, HRs for RHAP-MDH were significantly ($p < 0.05$) higher than RHAP-MOD for stroke in both GoDARTS (HR 2.35 [95% CI 1.78, 3.1]) and DCS (HR 2.31 [1.22, 4.38]) (ESM Tables 5.2, 5.3).

Multiple comparisons of survival rate curves indicated that CHF and CKD incidence was significantly higher in the RHAP-SIRD subgroup than in RHAP-MOD and RHAP-MD (ESM Tables 5.4, 5.5) in both cohorts. Although these higher

risks of complication might be driven mainly by the higher age of the RHAP-SIRD and RHAP-MDH subgroups, Cox models adjusted for age and sex still indicated significantly higher HRs of AMI and CKD in RHAP-SIRD compared with RHAP-MDH for both GoDARTS (AMI HR 1.72 [1.46, 2.02] and CKD HR 1.67 [1.51, 1.84]; unadjusted AMI HR 1.65 [1.40, 1.94] and unadjusted CKD HR 1.38 [1.25, 1.52]) and DCS (AMI HR 2.86 [1.35, 6.04] and CKD HR 1.77 [1.51, 2.08]; unadjusted AMI HR 3.53 [1.68, 7.41] and unadjusted CKD HR 2.02 [1.72, 2.36]) (Fig. 2, ESM Fig. 5.3).

Treatment patterns Clear variations in treatment patterns across subgroups over diabetes duration were seen (Fig. 3, ESM Fig. 6.1). In both cohorts, RHAP-MOD had the highest proportion of prescribing other oral antidiabetic drugs (OADs; otherwise known as oral glucose-lowering drugs), including dipeptidyl peptidase-4 inhibitors, glucagon-like peptide-1 analogues, α -glucosidase inhibitors, sodium-glucose

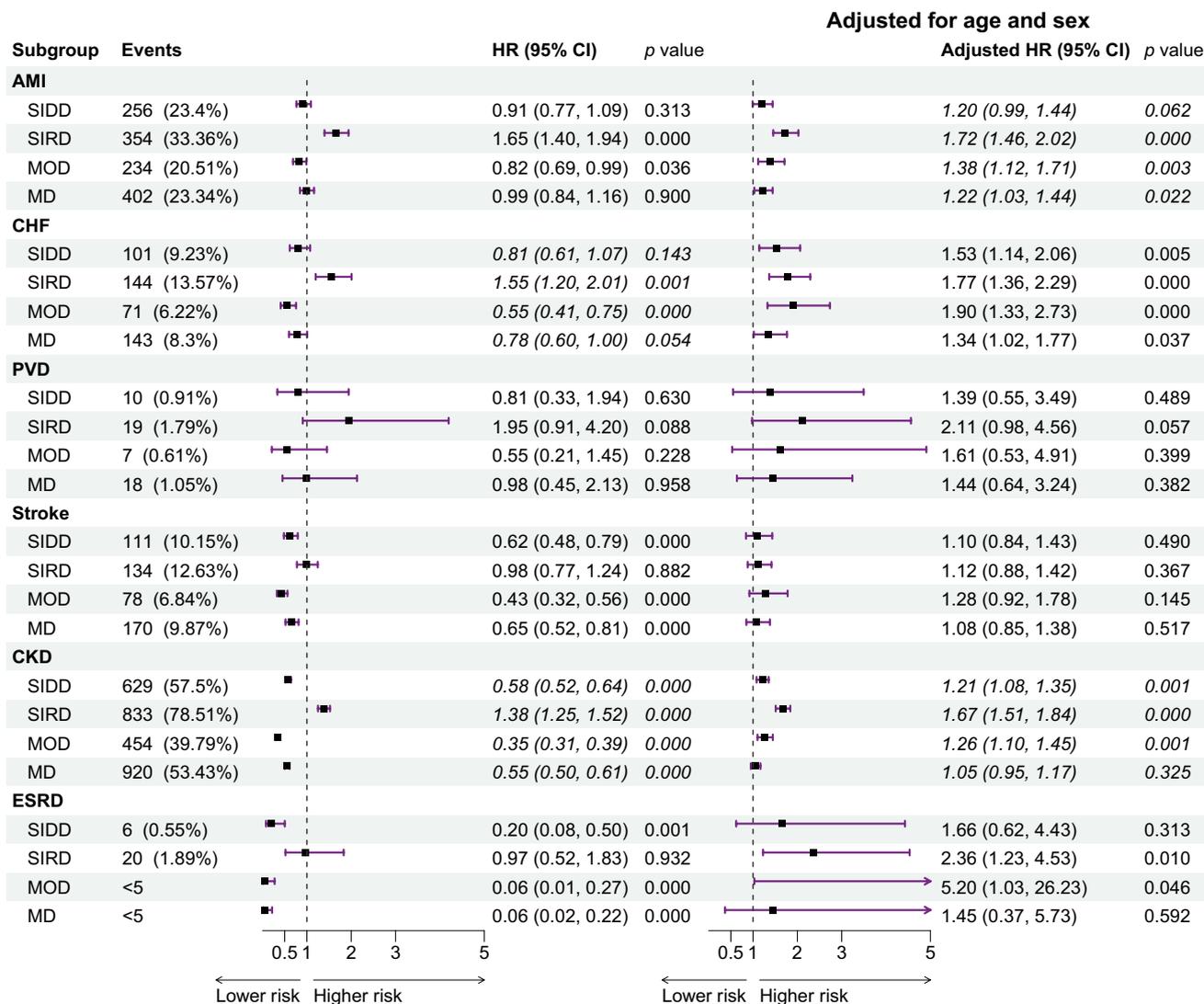


Fig. 2 The results of Cox regression analysis of GoDARTS. Here, SIDD, SIRD, MOD, MD and MDH refer to RHAP-SIDD, RHAP-SIRD, RHAP-MOD, RHAP-MD and RHAP-MDH, respectively.

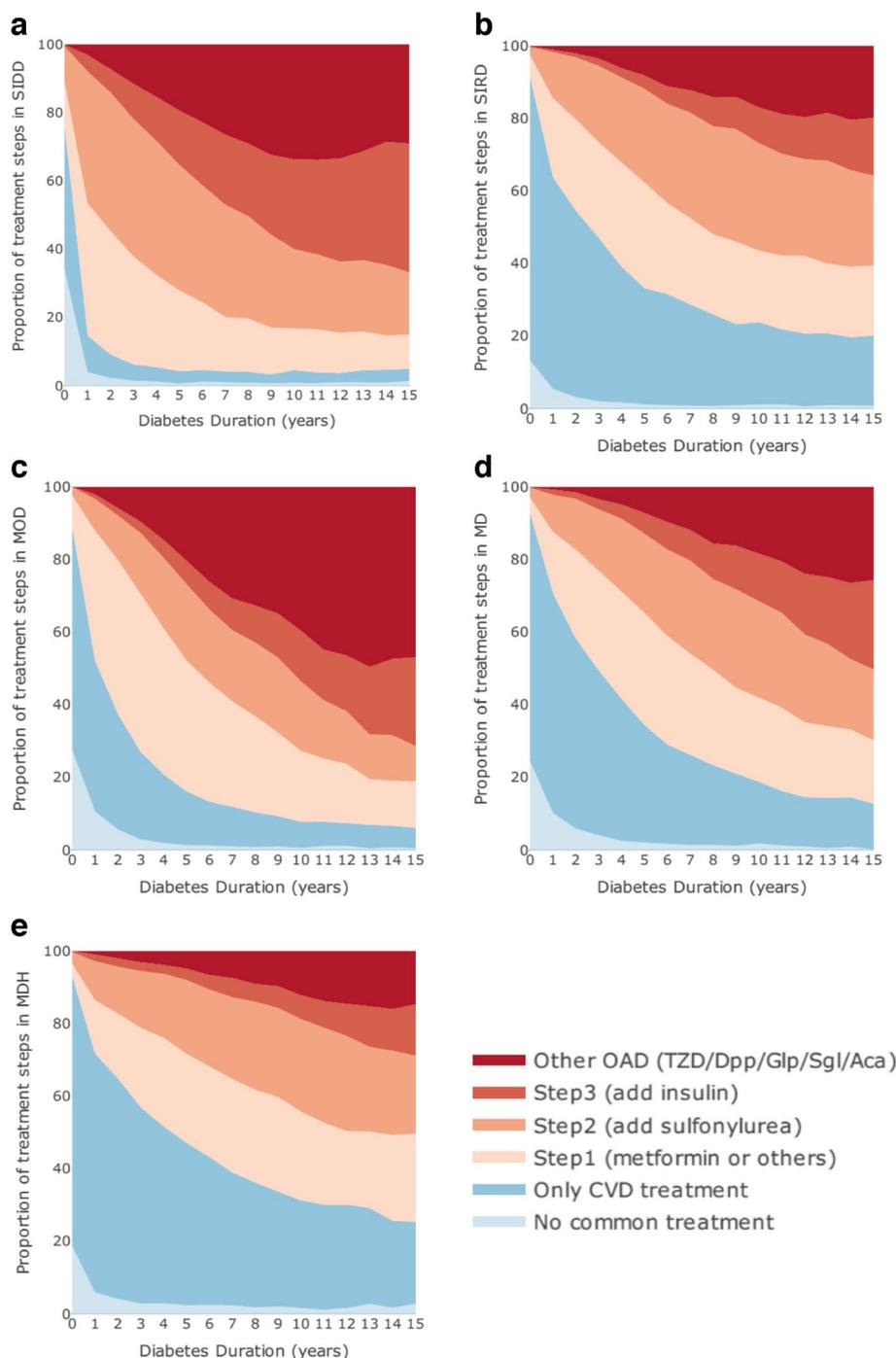
MDH is the reference group. Italics indicate that the proportional hazard assumption has not been fulfilled

cotransporter 2 inhibitors and thiazolidinediones (on average 23.83% in GoDARTS and 3.55% in DCS). Among these, the prescriptions for thiazolidinediones (6.16%) and dipeptidyl peptidase-4 inhibitors (5.15%) were the highest in GoDARTS, whereas both were less than 1% in DCS, reflecting differences in prescribing practices between the two countries. A substantially higher proportion of individuals with RHAP-SIDD received diabetes medication than other subgroups in both cohorts. More than half of individuals with RHAP-SIDD were prescribed insulin or other OADs at 10 years of diagnosis in GoDARTS (55.15%; 27.56% insulin and 27.58% other OADs) and DCS (52.48%; 49.65% insulin and 2.84% other OADs). In both GoDARTS and DCS, this proportion was the lowest in RHAP-MDH (17.34%; 14.91%), followed by RHAP-SIRD (24.58%; 17.63%), RHAP-MD (31.14%; 26.88%) and

RHAP-MOD (51.16%; 30.56%), indicating that individuals with RHAP-SIDD received the most intensive glucose control treatment, followed by RHAP-MOD. Multinomial logistic regression results (ESM Table 6.1) indicated that treatment patterns were not only visually distinct but also statistically different among all subgroups. For example, for RHAP-SIDD relative to RHAP-MD, the relative risk for using insulin (step 3) to no common treatment would be expected to increase by a factor of 3.07 (95% CI 2.73, 3.44) in GoDARTS and 11.80 (95% CI 8.98, 15.50) in DCS, given the other variables in the model are held constant.

Comparison between subgroups and clinical features to predict outcomes Using only clustering indicators compared with discrete cluster memberships resulted in better

Fig. 3 Area graph of treatment steps per individual over time for each subgroup in the GoDARTS cohort. Here, SIDD, SIRD, MOD, MD and MDH refer to RHAP-SIDD, RHAP-SIRD, RHAP-MOD, RHAP-MD and RHAP-MDH, respectively. On the x-axes, 0 indicates the period from diagnosis to less than 1 year after diagnosis; similarly, 1 indicates the period from 1 year after diagnosis to less than 2 years after diagnosis, etc. The treatment step was defined by an individual's first available observation in each diabetes duration interval. Treatment steps are defined as no common treatment (diet and exercise), only CVD treatment (Anatomical Therapeutic Chemical Classification System: C01–C10), step 1 (adding metformin [A10BA02] or repaglinide and nateglinide [A10BX]), step 2 (adding sulfonylurea [A10BB]), step 3 (adding insulin [A10A]) and other OAD (dipeptidyl peptidase-4 inhibitors [Dpp; A10BH], glucagon-like peptide-1 [Glp; A10BJ], α -glucosidase inhibitors [Aca; A10BF], sodium–glucose cotransporter 2 inhibitors [Sgl; A10BK], thiazolidinediones [TZD; A10BG], liraglutide [A10BX07], dapagliflozin [A10BX09])



fitting models in both cohorts (ESM Fig. 7.1; details in ESM Appendix 7), except for the Cox model for stroke, in which discrete subgroup membership performed slightly better than clustering indicators, though not significantly, as reflected by $RL > 0.1$. Yet adding discrete cluster memberships to clustering indicators achieved significantly lower AIC and thus a significantly better fit for the trajectories of BMI, HbA_{1c} , SBP, blood creatinine and treatment patterns in both cohorts. A detailed example can be found in ESM Appendix 8.

Consistency of subgroups classification over time In general, clusters were consistent over 8 years with an accuracy ranging from 59% to 72% (ESM Table 9.1). By de novo clustering over 8 years since diagnosis, on average, 53% of individuals migrated to other subgroups with shifted cluster centres (ESM Tables 9.1–9.3). The accuracy of allocation decreased by 4.52% from 0.70 in 2–4 years to 0.67 in 6–8 years in DCS, and by 7.11% from 0.64 in 2–4 years to 0.59 in 6–8 years in GoDARTS. The κ (0.49–0.62) indicated

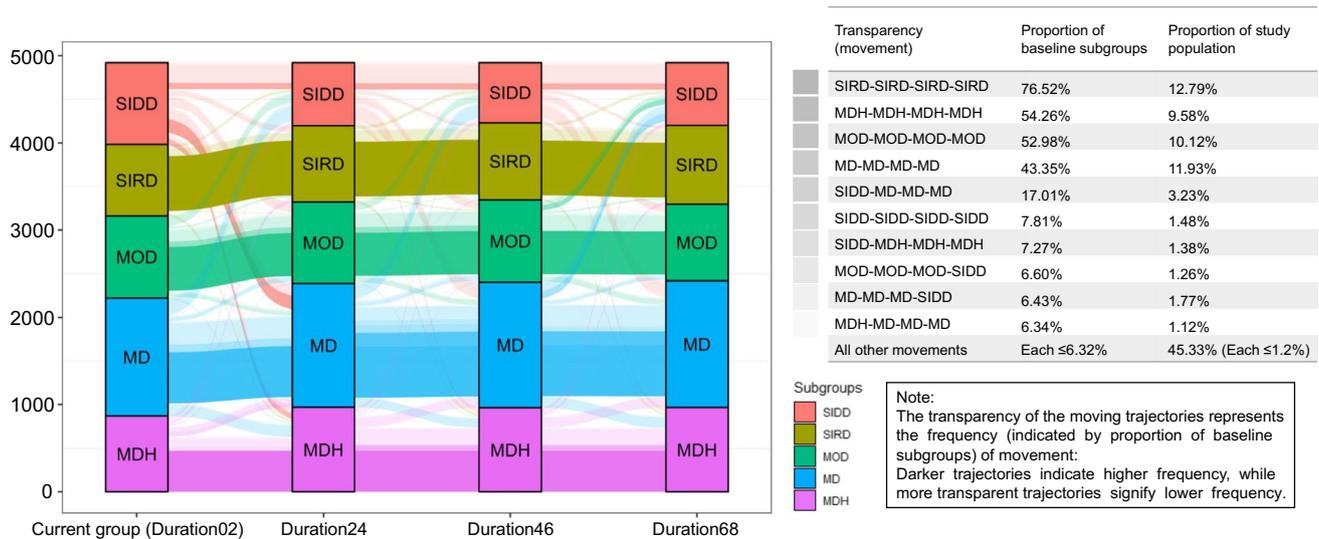


Fig. 4 Subgroups' redistribution using the centre-based reallocation method over time and characteristics of common trajectories in the GoDARTS cohort ($N=4919$). Here, SIDD, SIRD, MOD, MD and MDH refer to RHAP-SIDD, RHAP-SIRD, RHAP-MOD, RHAP-MD and RHAP-MDH, respectively. The figure shows the subgroups identified based on clinical characteristics within the first 2, 2–4, 4–6 and

6–8 years of diagnosis of type 2 diabetes, represented as Duration02, Duration24, Duration46 and Duration68, respectively, along with the top ten most frequent moving trajectories using the centre-based reallocation approach. Only individuals with information available for all four periods were included in the redistribution graph

a moderate to substantial agreement over time compared with subgroups identified at baseline. The specificity (ESM Table 9.3) was 0.91 on average, while the sensitivity and specific agreement were around 0.65 and 0.64 (lowest for RHAP-SIDD with average values of 0.25 and 0.28, respectively). By the centre-based reallocation method, accuracy (0.61–0.72), κ (0.51–0.64) and the proportion of individuals staying in the same cluster (0.46–0.72) improved by an average of 4.22%, 6.39% and 5.73% compared with the de novo clustering method.

By the centre-based reallocation method, in GoDARTS, the RHAP-SIRD subgroup displayed the highest stability, with 77% of individuals remaining in the same cluster for 8 years. In contrast, the RHAP-SIDD subgroup was the least stable, with only 8% of individuals staying in the same cluster (Fig. 4). The most common transitions for RHAP-SIDD were to RHAP-MD (17%) and RHAP-MDH (7%) within the initial 2 years, with individuals maintaining their position in that subgroup for the subsequent 6 years. These individuals had a higher proportion receiving insulin-based control treatment and a greater decrease in HbA_{1c} levels than individuals who were assigned to RHAP-MD or RHAP-MDH initially and stayed for the next 8 years (ESM Fig 9.1). Similar results could be found in DCS (ESM Figs 9.2, 9.3)

No significant difference in macrovascular disease risk was found between individuals transitioning from mild to severe subgroups and those remaining in the severe subgroup over 2 years (ESM Figs 9.4, 9.5). For those in severe subgroups initially (ESM Figs 9.6, 9.7), those who stayed in severe

subgroups for 2 years had a significantly higher risk of CKD (GoDARTS: HR 1.67 [1.48, 1.89]; DCS: HR 2.4 [1.92, 3.01]) than those who moved to mild subgroups. For individuals in mild subgroups initially, those who shifted to severe subgroups in 2 years had a higher risk of AMI (GoDARTS: HR 1.5 [1.25, 1.8]; DCS: HR 2.88 [1.39, 5.95]) and CHF (GoDARTS: HR 1.38 [1.02, 1.88]; DCS: HR 3.28 [1.99, 5.42]) than those who stayed in mild subgroups. Standardisation for age and sex did not change these findings (ESM Figs 9.6, 9.7).

Discussion

Using a much longer follow-up, we confirm previous findings [2] that data-driven subgroups effectively recognised individual phenotype heterogeneity, as reflected by significant differences in risk factor progression, complication risks and treatment patterns. Integrating subgroup information with clustering indicators may offer improved prediction of progression variation compared with either approach alone, emphasising the complementary role of subgroups rather than replacing continuous indicators. While most subgroups remain generally consistent over time, the RHAP-SIDD subgroup is notably volatile, indicating the necessity to expand insights from baseline subgroups to longitudinal status.

Significant differences in clinical parameters were observed not only at baseline but also over time among the subgroups,

such as high BMI and DBP in RHAP-MOD, high HDL-cholesterol and low triglycerides in RHAP-MDH, and high blood creatinine and low total cholesterol in RHAP-SIRD. The only exception was that the trajectory of HbA_{1c} in individuals with RHAP-MOD was poorly controlled for longer diabetes duration and worse than for RHAP-SIDD individuals.

Prior research has demonstrated that the SIRD subgroup exhibited higher risks of liver disease, macroalbuminuria, nephropathy, CKD and ESRD [1, 2, 6, 23]. Our analysis also revealed that RHAP-SIRD presented a higher risk of AMI, CHF, PVD, CKD and ESRD compared with other subgroups. By definition, subgroups varied in clustering indicators, such as age at baseline, which are among the risk factors for these complications. Upon adjusting for age, RHAP-SIRD maintained a significantly higher risk of AMI and CKD compared with other subgroups.

Treatment patterns varied significantly among subgroups, with the highest proportions of other OADs and overall glucose control treatment observed in RHAP-MOD and RHAP-SIDD subgroups, respectively. This suggests that physicians' treatment decisions for individuals within these subgroups differed, likely due to variations in age and other clustering indicators, as they were unaware of the individuals' subgroup membership.

The significant differences in disease progression, complication risks and treatment patterns among subgroups highlight their utility in understanding the underlying pathways of disease progression. Slieker et al [8] demonstrated that diabetes subgroups reveal distinct molecular mechanisms in key metabolic tissues, uncovering varied causes of the disease that are not apparent when it is viewed uniformly. Beyond aiding in aetiological understanding, subgroups may also be useful for predictive purposes. However, data-driven subgroups have been criticised for their unsuitability in predicting outcomes, such as drug response or complications [4, 24]. Our study partially supports this critique, as we found that using the clustering indicators may perform better than solely using subgroups for prediction. This is due to subgroups compressing data from several individual indicators, leading to information loss. However, we found that combining subgroup membership (e.g. SIDD) with the clustering indicators (e.g. age, BMI, etc.) often enhanced the performance of the progression models, indicating a potential predictive benefit from including subgroup information.

As expected, allocating individuals with long diabetes duration based on the lowest distance to baseline centroid leads to higher consistency of baseline subgroups. Practically, using cluster centres enables easy assignment of individuals to subgroups without requiring information about other individuals. To enhance accuracy, cluster centres can be periodically updated according to the latest cohort characteristics, similar to routine updates in risk prediction models. Furthermore, our study revealed RHAP-SIRD to

be the most consistent subgroup over time, with over 70% of individuals remaining for over 8 years, signifying its distinct, partially divergent aetiology. This aligns with prior research identifying SIRD as the most genetically unique subgroup [25], exhibiting an insulin resistance molecular signature [8] and lacking associations with the type 2 diabetes locus in the *TCF7L2* gene or insulin secretion risk scores, contrary to SIDD and MOD [1, 25–27].

Ahlqvist's original study was designed to deepen the understanding of diabetes heterogeneity and enhance individualised treatment by identifying baseline phenotypes [1]. To fully benefit from the long follow-up information available, we expanded this concept to include more than just baseline subgroups, attempting to explore the dynamics of disease. As expected, we observed changes in subgroup memberships over time, reflecting the combination of treatment effects and underlying phenotypes. For example, we found that more than 28% of individuals transitioned to other subgroups after 2 years. These temporal dynamics might be shaped by interactions between disease heterogeneity, adherence to treatment and treatment efficacy. Diabetes heterogeneity, such as distinct molecular signatures and genetic characteristics [1, 8], may result in individuals consistently belonging to specific subgroups with unique phenotypes. However, the treatment meanwhile aims to shift individuals toward milder subgroups. For example, newly diagnosed individuals who subsequently meet guideline-based treatment targets (53 mmol/mol (7%) HbA_{1c} [28, 29], 0.9 mmol/l HDL-cholesterol [30], 25 kg/m² BMI [31]) will either remain or progress to the RHAP-MD subgroup over time, whereas insufficient risk factor control could result in increased progression to severe subgroups.

The longitudinal nature of our data allowed us to estimate the impact of changes in subgroup membership over time. We found that complication risks were more closely associated with individuals' current subgroups rather than the initial subgroups they were assigned at baseline. The risks of complications for individuals progressing from mild to severe subgroups were similar to those for individuals initially allocated to and remaining in severe subgroups. Also, individuals progressing from severe to mild subgroups showed complication risks lower than for those who remained in severe subgroups. Thus, an initial allocation to a mild subgroup did not necessarily translate into mild progression, and efforts should aim at achieving or maintaining mild subgroup status. This might suggest the importance of periodically re-clustering with changing risk factors as the disease progresses to capture the evolving dynamics and guide more informed decision-making.

Our study is not without limitations. First, C-peptide, one of the five clustering indicators, was assumed to be constant, due to the lack of follow-up data. This might overestimate subgroup consistency, but its impact is likely limited due to C-peptide's

stability [1]. Second, we estimated the treatment pattern from observed data and ignored censoring (ESM Figs 10.1, 10.2), which might underestimate the proportion of individuals taking the most intensive treatment steps. Third, due to the unavailability of fasting glucose data in GoDARTS, we were unable to replicate Ahlqvist's subgroups within this registry. Ahlqvist's method captures two key pathogenic mechanisms: insulin deficiency and resistance, indicated by HOMA-IR and HOMA-B. We used C-peptide instead, which may obscure the pathology link with type 2 diabetes. Nevertheless, considering the high sensitivity and specificity of RHAP-SIDD (72% and 100%) and RHAP-SIRD (67% and 89%) in relation to Ahlqvist's subgroups (ESM Fig. 11.1), our findings for RHAP-SIDD and RHAP-SIRD may offer insights for Ahlqvist's subgroups. Of note, SIDD had worse beta cell function than other subgroups described by Ahlqvist et al [1], and this was partially conveyed by the lower C-peptide of RHAP-SIDD among the RHAP-SODY subgroups. Since C-peptide is generally stable over time [9], but beta cell function progressively declines [32], we might expect even worse stability for SIDD in Ahlqvist's subgroups. Fourth, DCS registered events were based on self-report, which could lead to an underestimation of events. However, a validation study found events to be well reported, with 86% sensitivity and 90% specificity [12]. Finally, our cohorts, predominantly consisting of white individuals, may limit the generalisability of findings to other settings.

In conclusion, the significant differences observed in subgroups' trajectories raise the possibility of identifying and understanding different phenotypes of type 2 diabetes. Also, subgroup information may improve prediction when added as a predictor. This lays the foundation for considering diabetes subgroups as complementary to, rather than replacements for, individual indicators.

Supplementary Information The online version of this article (<https://doi.org/10.1007/s00125-024-06147-y>) contains peer-reviewed but unedited supplementary material.

Acknowledgements The authors acknowledge all the DCS and GoDARTS participants and all personnel at the Amsterdam University Medical Center and University of Dundee who have helped with the cohort creation and maintenance. The authors also acknowledge the Health Informatics Centre for the provision of anonymised clinical data for GoDARTS from NHS Tayside, the original data owner. The authors thank A. A. Van Der Heijden, Amsterdam UMC, for providing DCS data and helping out on various data-related questions; S. Emamipour, University of Groningen, for helping with understanding DCS data; J. Wang, Utrecht University, S. R. A. Konings, University of Groningen, and F. Li, University of Groningen, for their scientific advice; and the audience from the 56th Annual Meeting of the European Diabetes Epidemiology Group, especially J. Dennis, University of Exeter, for the scientific advice.

Data availability The data that support the findings of this study are available from Amsterdam University Medical Center and the University of Dundee, and were accessed by the authors via a formal data request procedure and as a part of the RHAPSODY project. Therefore, the data are not publicly available. Steering committees of the

individual cohorts will consider reasonable requests for sharing of de-identified individual-level data.

Funding This project has received funding from the Innovative Medicines Initiative 2 Joint Undertaking under grant agreement no. 115881 (RHAPSODY). This Joint Undertaking receives support from the European Union's Horizon 2020 research and innovation programme and EFPIA. This work is supported by the Swiss State Secretariat for Education, Research and Innovation (SERI) under contract no. 16.0097. The Hoorn DCS cohort was supported by grants from the Netherlands Organisation for Health Research and Development (113102006, 459001015). The opinions expressed and arguments employed herein do not necessarily reflect the official views of these funding bodies. The funding sources played no roles in study design; in the collection, analysis and interpretation of data; in the writing of the report; and in the decision to submit the paper for publication.

Authors' relationships and activities ERP received payment or honoraria for lectures, presentations, speakers' bureaus, manuscript writing or educational events from Lilly, Illumina and Sanofi, and support for attending meetings and/or travel from Lilly. TF is a member of the Mount Hood Diabetes modelling challenge and a member of the ISPOR SIG open-source modelling group, and journal club organiser. The authors declare that there are no other relationships or activities that might bias, or be perceived to bias, their work.

Contribution statement XL did the literature search, designed the study, performed statistical analysis, conceptualised the paper and drafted the manuscript. AvG, JL and TF were involved in conceptualisation, formal analysis, investigation, methodology, supervision, visualisation and writing. LAD, RCS, JWJB, LMH, PJME and ERP were involved in data curation and interpretation of data. LAD, RCS, JWJB, LMH, PJME, ERP, AvG, JL and TF reviewed and edited the manuscript. XL and TF had access to all the data and verified the underlying data reported in the manuscript. XL and TF are the guarantors of this work and, as such, take responsibility for the integrity of the data and the accuracy of the data analysis. All authors contributed to the critical revision of the manuscript and approved the final manuscript.

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

References

1. Ahlqvist E, Storm P, Karajamaki A et al (2018) Novel subgroups of adult-onset diabetes and their association with outcomes: a data-driven cluster analysis of six variables. *Lancet Diabetes Endocrinol* 6(5):361–369. [https://doi.org/10.1016/S2213-8587\(18\)30051-2](https://doi.org/10.1016/S2213-8587(18)30051-2)
2. Zaharia OP, Strassburger K, Strom A et al (2019) Risk of diabetes-associated diseases in subgroups of patients with recent-onset

- diabetes: a 5-year follow-up study. *Lancet Diabetes Endocrinol* 7(9):684–694. [https://doi.org/10.1016/S2213-8587\(19\)30187-1](https://doi.org/10.1016/S2213-8587(19)30187-1)
3. Kahkoska AR, Geybels MS, Klein KR et al (2020) Validation of distinct type 2 diabetes clusters and their association with diabetes complications in the DEVOTE, LEADER and SUSTAIN-6 cardiovascular outcomes trials. *Diabetes Obes Metab* 22(9):1537–1547. <https://doi.org/10.1111/dom.14063>
 4. Dennis JM, Shields BM, Henley WE, Jones AG, Hattersley AT (2019) 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 Diabetes Endocrinol* 7(6):442–451. [https://doi.org/10.1016/S2213-8587\(19\)30087-7](https://doi.org/10.1016/S2213-8587(19)30087-7)
 5. Anjana RM, Baskar V, Nair ATN et al (2020) 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* 8(1):e001506. <https://doi.org/10.1136/bmjdr-2020-001506>
 6. Tanabe H, Saito H, Kudo A et al (2020) Factors associated with risk of diabetic complications in novel cluster-based diabetes subgroups: a Japanese retrospective cohort study. *J Clin Med* 9(7):2083. <https://doi.org/10.3390/jcm9072083>
 7. Herder C, Roden M (2022) A novel diabetes typology: towards precision diabetology from pathogenesis to treatment. *Diabetologia* 65(11):1770–1781. <https://doi.org/10.1007/s00125-021-05625-x>
 8. Sliker RC, Donnelly LA, Fitipaldi H et al (2021) Distinct molecular signatures of clinical clusters in people with type 2 diabetes: an IMI-RHAPSODY study. *Diabetes* 70(11):2683–2693. <https://doi.org/10.2337/db20-1281>
 9. Sliker RC, Donnelly LA, Fitipaldi H et al (2021) Replication and cross-validation of type 2 diabetes subtypes based on clinical variables: an IMI-RHAPSODY study. *Diabetologia* 64(9):1982–1989. <https://doi.org/10.1007/s00125-021-05490-8>
 10. Li X, van Giessen A, Altunkaya J et al (2023) Potential value of identifying type 2 diabetes subgroups for guiding intensive treatment: a comparison of novel data-driven clustering with risk-driven subgroups. *Diabetes Care* 46(7):1395–1403. <https://doi.org/10.2337/dc22-2170>
 11. Cuschieri S (2019) The STROBE guidelines. *Saudi J Anaesth* 13:31–34. https://doi.org/10.4103/sja.SJA_543_18
 12. van der Heijden AA, Rauh SP, Dekker JM et al (2017) The Hoorn Diabetes Care System (DCS) cohort. A prospective cohort of persons with type 2 diabetes treated in primary care in the Netherlands. *BMJ Open* 7(5):e015599. <https://doi.org/10.1136/bmjopen-2016-015599>
 13. Hebert HL, Shepherd B, Milburn K et al (2018) Cohort profile: Genetics of Diabetes Audit and Research in Tayside Scotland (GoDARTS). *Int J Epidemiol* 47(2):380–381j. <https://doi.org/10.1093/ije/dyx140>
 14. Jones AG, Lonergan M, Rodgers LR et al (2015) Studies of diabetes treatment stratification should correct for baseline HbA1c: a MASTERMIND study. *Diabetic Med* 32:94–94
 15. NHG-werkgroep (2021) NHG-Standaard Diabetes mellitus type 2 (M01). Available from <https://richtlijnen.nhg.org/standaarden/diabetes-mellitus-type-2>. Accessed 12 March 2024
 16. Zhou KX, Donnelly LA, Morris AD et al (2014) Clinical and genetic determinants of progression of type 2 diabetes: a DIRECT study. *Diabetes Care* 37(3):718–724. <https://doi.org/10.2337/dc13-1995>
 17. de Vet HCW, Mokkink LB, Terwee CB, Hoekstra OS, Knol DL (2013) Clinicians are right not to like Cohen's kappa. *BMJ Br Med J* 346:f2125. <https://doi.org/10.1136/bmj.f2125>
 18. Viera AJ, Garrett JM (2005) Understanding interobserver agreement: the kappa statistic. *Fam Med* 37(5):360–363
 19. Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate - a practical and powerful approach to multiple testing. *J R Stat Soc B* 57(1):289–300. <https://doi.org/10.1111/j.2517-6161.1995.tb02031.x>
 20. Lagakos SW, Schoenfeld DA (1984) Properties of proportional-hazards score tests under misspecified regression-models. *Biometrics* 40(4):1037–1048. <https://doi.org/10.2307/2531154>
 21. Wagenmakers EJ, Farrell S (2004) AIC model selection using Akaike weights. *Psychon Bull Rev* 11(1):192–196. <https://doi.org/10.3758/Bf03206482>
 22. R Core Team (2021) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria
 23. Bello-Chavolla OY, Bahena-Lopez JP, Vargas-Vazquez A et al (2020) Clinical characterization of data-driven diabetes subgroups in Mexicans using a reproducible machine learning approach. *BMJ Open Diabetes Res Care* 8(1):e001550. <https://doi.org/10.1136/bmjdr-2020-001550>
 24. Lugner M, Gudbjornsdottir S, Sattar N et al (2021) Comparison between data-driven clusters and models based on clinical features to predict outcomes in type 2 diabetes: nationwide observational study. *Diabetologia* 64(9):1973–1981. <https://doi.org/10.1007/s00125-021-05485-5>
 25. Ahlqvist E, Prasad RB, Groop L (2020) Subtypes of type 2 diabetes determined from clinical parameters. *Diabetes* 69(10):2086–2093. <https://doi.org/10.2337/dbi20-0001>
 26. Saxena R, Gianniny L, Burt NP et al (2006) Common single nucleotide polymorphisms in TCF7L2 are reproducibly associated with type 2 diabetes and reduce the insulin response to glucose in nondiabetic individuals. *Diabetes* 55(10):2890–2895. <https://doi.org/10.2337/db06-0381>
 27. Grant SFA, Thorleifsson G, Reynisdottir I et al (2006) Variant of transcription factor 7-like 2 (TCF7L2) gene confers risk of type 2 diabetes. *Nat Genet* 38(3):320–323. <https://doi.org/10.1038/ng1732>
 28. ElSayed NA, Aleppo G, Aroda VR et al (2022) 6. Glycemic targets: standards of care in diabetes—2023. *Diabetes Care* 46(Suppl 1):S97–S110. <https://doi.org/10.2337/dc23-S006>
 29. Cosentino F, Grant PJ, Aboyans V et al (2020) 2019 ESC guidelines on diabetes, pre-diabetes, and cardiovascular diseases developed in collaboration with the EASD. *Eur Heart J* 41(2):255–323. <https://doi.org/10.1093/eurheartj/ehz486>
 30. ElSayed NA, Aleppo G, Aroda VR et al (2022) 10. Cardiovascular disease and risk management: standards of care in diabetes—2023. *Diabetes Care* 46(Suppl 1):S158–S190. <https://doi.org/10.2337/dc23-S010>
 31. ElSayed NA, Aleppo G, Aroda VR et al (2022) 8. Obesity and weight management for the prevention and treatment of type 2 diabetes: standards of care in diabetes—2023. *Diabetes Care* 46(Suppl 1):S128–S139. <https://doi.org/10.2337/dc23-S008>
 32. Kahn SE, Cooper ME, Del Prato S (2014) Pathophysiology and treatment of type 2 diabetes: perspectives on the past, present, and future. *Lancet* 383(9922):1068–1083. [https://doi.org/10.1016/S0140-6736\(13\)62154-6](https://doi.org/10.1016/S0140-6736(13)62154-6)

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

Authors and Affiliations

Xinyu Li¹  · Louise A. Donnelly² · Roderick C. Sliker^{3,4,5}  · Joline W. J. Beulens^{4,5,6}  · Leen M. 't Hart^{3,4,5,7}  · Petra J. M. Elders^{5,8}  · Ewan R. Pearson²  · Anoukh van Giessen⁹  · Jose Leal¹⁰  · Talitha Feenstra^{1,9} 

✉ Xinyu Li
li.xinyu@rug.nl

¹ Groningen Research Institute of Pharmacy, Faculty of Science and Engineering, University of Groningen, Groningen, the Netherlands

² Division of Population Health and Genomics, Ninewells Hospital and School of Medicine, University of Dundee, Dundee, UK

³ Department of Cell and Chemical Biology, Leiden University Medical Center, Leiden, the Netherlands

⁴ Department of Epidemiology and Data Science, Amsterdam UMC, Location Vrije Universiteit Amsterdam, Amsterdam, the Netherlands

⁵ Amsterdam Public Health, Amsterdam Cardiovascular Sciences, Amsterdam, the Netherlands

⁶ Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht University, Utrecht, Netherlands

⁷ Department of Biomedical Data Sciences, Section Molecular Epidemiology, Leiden University Medical Center, Leiden, the Netherlands

⁸ Department of General Practice, Amsterdam University Medical Center, Location Vrije Universiteit Amsterdam, Amsterdam, the Netherlands

⁹ National Institute of Public Health and the Environment, Bilthoven, the Netherlands

¹⁰ Health Economics Research Centre, Nuffield Department of Population Health, University of Oxford, Oxford, UK