



The efficacy of canagliflozin in diabetes subgroups stratified by data-driven clustering or a supervised machine learning method: a post hoc analysis of canagliflozin clinical trial data

Xiantong Zou¹ · Qi Huang¹ · Yingying Luo¹ · Qian Ren¹ · Xueyao Han¹ · Xianghai Zhou¹ · Linong Ji¹

Received: 31 December 2021 / Accepted: 29 April 2022 / Published online: 8 July 2022

© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2022

Abstract

Aims/hypothesis Data-driven diabetes subgroups have shown distinct clinical characteristics and disease progression, although there is a lack of evidence that this information can guide clinical decisions. We aimed to investigate whether diabetes subgroups, identified by data-driven clustering or supervised machine learning methods, respond differently to canagliflozin.

Methods We pooled data from five randomised, double-blinded clinical trials of canagliflozin at an individual level. We applied the coordinates from the All New Diabetics in Scania (ANDIS) study to form four subgroups: mild age-related diabetes (MARD); severe insulin-deficient diabetes (SIDD); mild obesity-related diabetes (MOD) and severe insulin-resistant diabetes (SIRD). Machine learning models for HbA_{1c} lowering (ML-A1C) and albuminuria progression (ML-ACR) were developed. The primary efficacy endpoint was reduction in HbA_{1c} at 52 weeks. Concordance of a model was defined as the difference between predicted HbA_{1c} and actual HbA_{1c} decline less than 3.28 mmol/mol (0.3%).

Results The decline in HbA_{1c} resulting from treatment was different among the four diabetes clusters ($p_{\text{interaction}}=0.004$). In MOD, canagliflozin showed a robust glucose-lowering effect at week 52, compared with other drugs, with least-squares mean of HbA_{1c} decline [95% CI] being 6.6 mmol/mol (4.1, 9.2) (0.61% [0.38, 0.84]) for sitagliptin, 7.1 mmol/mol (4.7, 9.5) (0.65% [0.43, 0.87]) for glimepiride, and 9.8 mmol/mol (9.0, 10.5) (0.90% [0.83, 0.96]) for canagliflozin. This superiority persisted until 104 weeks. The proportion of individuals who achieved HbA_{1c} <53 mmol/mol (<7.0%) was highest in sitagliptin-treated individuals with MARD but was similar among drugs in individuals with MOD. The ML-A1C model and the cluster algorithm showed a similar concordance rate in predicting HbA_{1c} lowering (31.5% vs 31.4%, $p=0.996$). Individuals were divided into high-risk and low-risk groups using ML-ACR model according to their predicted progression risk for albuminuria. The effect of canagliflozin vs placebo on albuminuria progression differed significantly between the high-risk (HR 0.67 [95% CI 0.57, 0.80]) and low-risk groups (HR 0.91 [0.75, 1.11]) ($p_{\text{interaction}}=0.016$).

Conclusions/interpretation Data-driven clusters of individuals with diabetes showed different responses to canagliflozin in glucose lowering but not renal outcome prevention. Canagliflozin reduced the risk of albumin progression in high-risk individuals identified by supervised machine learning. Further studies with larger sample sizes for external replication and subtype-specific clinical trials are necessary to determine the clinical utility of these stratification strategies in sodium–glucose cotransporter 2 inhibitor treatment.

Data availability The application for the clinical trial data source is available on the YODA website (<http://yoda.yale.edu/>).

Keywords Data-driven clusters · Diabetes · Machine learning · SGLT2i

Xiantong Zou and Qi Huang contributed equally to this work

✉ Xiantong Zou
eva2172@163.com

✉ Linong Ji
Jiln@bjmu.edu.cn

¹ Department of Endocrinology and Metabolism, Peking University People's Hospital, Beijing, China

Abbreviations

ACR	Urinary albumin/creatinine ratio
ADOPT	A Diabetes Outcome Progression Trial
ANDIS	All New Diabetics in Scania
BUN	Blood urea nitrogen
CANTATA-D	Canagliflozin Treatment and Trial Analysis - DPP-4 Inhibitor Comparator Trial

Research in context

What is already known about this subject?

- Data-driven diabetes clustering divided diabetes into four subgroups: MARD, SIRD, SIDD and MOD
- The four subgroups have distinct clinical characteristics and disease trajectories, although there is a lack of evidence that these clusters can guide clinical drug choice
- Supervised machine learning models can be used to predict the response to glucose-lowering treatment

What is the key question?

- Do subgroups identified by data-driven diabetes clustering or supervised machine learning-based models respond differently to sodium–glucose cotransporter 2 inhibitors (SGLT2i)?

What are the new findings?

- The HbA_{1c}-lowering effect of canagliflozin was different among clusters
- The albuminuria progression following SGLT2i vs placebo treatment was different in high-risk and low-risk individuals stratified by a supervised machine learning model

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

- Data-driven diabetes clusters may help clinicians to decide whether to use SGLT2i for glucose lowering in certain diabetes subtypes; a supervised machine learning model, verified by external replication, may be used as a tool to stratify patients and guide SGLT2i usage for preventing albuminuria progression

CANTATA-M	Canagliflozin Treatment and Trial Analysis Monotherapy	ROC AUC	Average area under the receiver-operating characteristic curve
CANTATA-SU	Canagliflozin Treatment And Trial Analysis-Sulfonylurea	SBP	Systolic BP
CANVAS	Canagliflozin Cardiovascular Assessment Study	SCr	Serum creatine
CANVAS-R	A Study of the Effects of Canagliflozin on Renal Endpoints in Adult Participants With Type 2 Diabetes Mellitus	SGLT2i	Sodium–glucose cotransporter 2 inhibitor
		SIDD	Severe insulin-deficient diabetes
		SIRD	Severe insulin-resistant diabetes
		SU	Sulfonylurea
		SVM	Support vector machine
DBP	Diastolic BP	TC	Total cholesterol
DPP4i	Dipeptidyl peptidase-4 inhibitor	TG	Triacylglycerol
FDR	False discovery rate	TZD	Thiazolidinedione
FPG	Fasting plasma glucose	YODA	Yale University Open Data Access
GADab	GAD antibody		
HDL	HDL-cholesterol		
kNN	k-Nearest neighbours		
LDL	LDL-cholesterol		
LRT	Likelihood ratio test		
LS	Least squares		
MAE	Mean average error		
MARD	Mild age-related diabetes		
mITT	Modified intent-to-treat		
ML	Machine learning		
ML-A1C model	ML-based model for HbA _{1c} decline		
ML-ACR model	ML-based model for albuminuria progression		
MOD	Mild obesity-related diabetes		
OHD	Oral glucose-lowering drug		
RF	Random forest		

Introduction

Sodium–glucose cotransporter 2 inhibitors (SGLT2i) are novel and innovative glucose-lowering treatments, which can alter cardio-renal outcomes [1–3]. Type 2 diabetes is a highly heterogeneous disease so there is a clinical need to identify individuals who can benefit from SGLT2i treatment for better health resource allocation [4]. Artificial intelligence methods are increasingly used to stratify patients thus facilitating clinical decisions in line with the promotion of precision medicine. Among these methods, data-driven clustering was recently proposed for refining diabetes classification, with diabetes being clustered into five groups using GAD antibody

(GADab), age, BMI, HOMA2-B and HOMA2-IR [5]. Where GADab data were absent, four cluster subgroups, including mild age-related diabetes (MARD), mild obesity-related diabetes (MOD), severe insulin-deficient diabetes (SIDD) and severe insulin-resistant diabetes (SIRD), were reproduced in epidemiological studies in China and the USA [6] as well as clinical trials involving participants from different ethnicities [7]. The subgroups have distinct clinical characteristics and different trajectories towards diabetes complications [5], raising the possibility of providing ‘precision management’ [8] to diabetes patients.

Currently, evidence is accumulating to address whether the clinical response of oral glucose-lowering drugs (OHDs [oral hypoglycaemic drugs]) varies across these subtypes. Dennis et al used the data from the A Diabetes Outcome Progression Trial (ADOPT) to show that sulfonylureas (SUs) may be suitable for individuals with MARD and thiazolidinediones (TZDs) may bring benefit to those with SIRD in terms of glucose lowering [7]. However, whether the effects of SGLT2i on glucose lowering and clinical outcomes are different among data-driven clusters is largely unknown. In the realm of precision medicine, supervised learning is frequently used in predicting clinical outcomes and the clinical response to the OHD [9]. This study aimed to investigate whether stratification of individuals by data-driven clustering or supervised machine learning (ML) can distinguish those deriving greater benefit from treatment with SGLT2i, and therefore guide clinical decisions, using clinical trials of canagliflozin.

Methods

Study design and participants

We used data from five randomised, double-blinded, multi-centre clinical trials of canagliflozin ([ClinicalTrials.gov](https://clinicaltrials.gov) identifiers: NCT01081834, Canagliflozin Treatment and Trial Analysis Monotherapy [CANTATA-M] [10]; NCT01106677, Canagliflozin Treatment and Trial Analysis - DPP-4 Inhibitor Comparator Trial [CANTATA-D] [11]; NCT00968812, Canagliflozin Treatment And Trial Analysis-Sulfonylurea [CANTATA-SU] [12]; NCT01032629, Canagliflozin Cardiovascular Assessment Study [CANVAS] [2]; and NCT01989754, A Study of the Effects of Canagliflozin on Renal Endpoints in Adult Participants With Type 2 Diabetes Mellitus [CANVAS-R] [2]) obtained through the Yale University Open Data Access (YODA) Project (no. 2020-4211), which has an agreement with Janssen Research & Development. The interpretation and reporting of research using this data are solely the responsibility of the authors and do not necessarily represent the official views of the YODA Project or Janssen Research & Development.

All trials complied with the International Conference on Harmonization Good Clinical Practice guidelines and the Declaration of Helsinki and the protocols were approved by local ethics committees and institutional review boards before the start of each clinical trial. No informed consent was required since all data were previously collected and all individual identities were masked.

Eligibility criteria for each trial have been published previously [2, 10–12]. Briefly, CANTATA studies were multi-centre, randomised, double-blinded, phase 3 clinical trials in individuals with type 2 diabetes who were drug-naive or who received metformin monotherapy, in which the efficacy of canagliflozin (either 100 mg or 300 mg daily) was evaluated in reference to placebo, sitagliptin 100 mg daily or glimepiride (1–8 mg daily). The CANVAS program was a randomised trial investigating the role of canagliflozin vs placebo on cardiovascular outcomes in individuals with type 2 diabetes at high cardiovascular risk [2].

For data-driven cluster assays, 6365 participants were pooled from the modified intent-to-treat (mITT) population of all CANTATA trials and CANVAS for whom there was complete information on variables for cluster allocation (Fig. 1). There were 175 participants in the CANTATA-D trial who used a placebo during the first 26 weeks and then switched to sitagliptin. These were included as participants treated using a placebo before 26 weeks for efficacy analysis and were excluded from further analysis. Only participants from CANVAS ([ClinTrials.gov](https://clinicaltrials.gov) registration no. NCT01809327) were included for outcome analysis.

As a comparison, an ML-based model for HbA_{1c} decline (ML-A1C model) and an ML-based model for albuminuria progression (ML-ACR model) were developed in CANTATA trials and CANVAS-R trial, respectively, and externally validated in CANVAS (Fig. 1).

Outcomes

Since the minimum observation period for all experimental drugs was 52 weeks among all trials, we selected 52 weeks to assess the efficacy of SGLT2i. We combined the two dosages of canagliflozin as one treatment arm. For efficacy analysis, the primary endpoint was the decline of HbA_{1c} from baseline to week 52. We also aimed to investigate the following outcomes: (1) the percentage of participants attaining HbA_{1c} <53 mmol/mol (<7.0%) after 52 weeks of treatment; and (2) changes in HbA_{1c}, fasting plasma glucose (FPG) and body weight from baseline to 104 weeks.

Renal outcome consisted of the progression of albuminuria and renal composite outcome. Progression of albuminuria was defined as a more than 30% increase in urinary albumin/creatinine ratio (ACR), or change from normoalbuminuria to

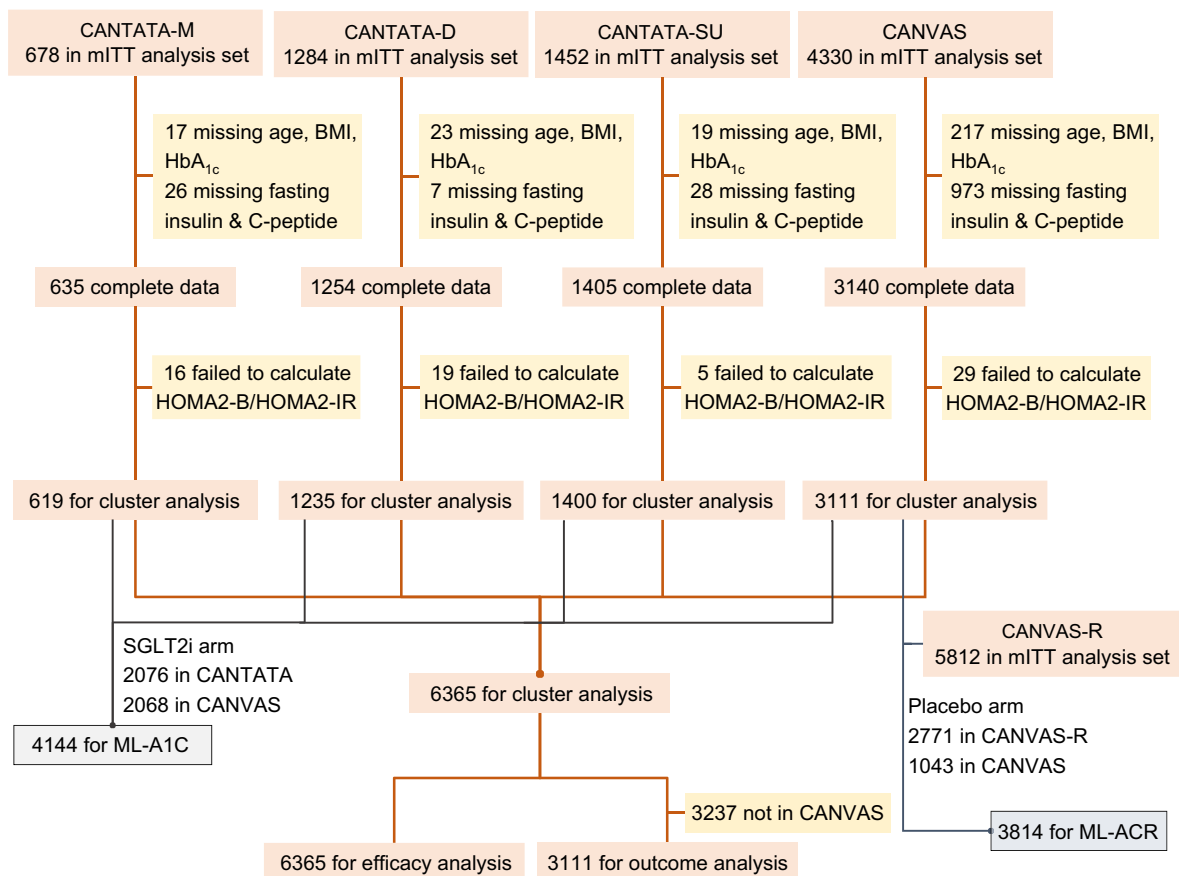


Fig. 1 Flowchart of study participant inclusion. CANTATA and CANVAS were used for data-driven analysis. CANTATA and CANVAS were used for ML-A1C development and validation ($N=4144$). CANVAS-R and CANVAS were used for ML-ACR model

development and validation ($N=3814$). Red lines represent cluster/efficacy/outcome analyses. Black lines represent ML-A1C and ML-ACR models

microalbuminuria, or from normoalbuminuria or microalbuminuria to macroalbuminuria. Renal composite outcome was defined as follows: (1) 40% reduction in eGFR from baseline sustained for at least 30 days; (2) end-stage renal disease; (3) doubling of serum creatinine; or (4) death with a proximal renal cause.

Cluster analysis

HOMA2-B and HOMA2-IR were calculated using the HOMA2 calculator of Oxford [13] using baseline fasting glucose and fasting C-peptide (the latter was substituted with fasting insulin due to unavailability in 59 participants). Since all participants recruited had type 2 diabetes and GADab data were missing, we assumed all participants were GADab negative. HOMA2-IR, HOMA2-B, BMI, HbA_{1c} and age, instead of age-of-onset due to unavailability, were used for cluster analysis. Using the coordinates from the All New Diabetics in Scania (ANDIS) cohort [5], we allocated the participants in each trial into the four diabetes subgroups.

Supervised ML-based predictive models

ML-A1C model We used CANTATA trials ($N=2076$) to develop the ML-A1C model to predict HbA_{1c} decline at week 52 in individuals treated with canagliflozin and validated the model in the CANVAS trial ($N=3111$). We used a set of baseline clinical variables relating to hypoglycaemic response, including age, sex, race, pulse, systolic BP (SBP), diastolic BP (DBP), height, weight, BMI, HbA_{1c}, FPG, fasting insulin, fasting C-peptide, triacylglycerol (TG), total cholesterol (TC), HDL-cholesterol (HDL), LDL-cholesterol (LDL), urinary ACR, serum creatine (SCr), eGFR, blood urea nitrogen (BUN), HOMA2-B, HOMA2-IR and background treatment (using OHD or insulin) as predictors. Weight, FPG, fasting insulin, fasting C-peptide and TC were excluded for being missing in more than 20% of samples or being in high collinearity with existing variables (Pearson coefficient >0.7). Skew variables were log₁₀-transformed. Missing data were generated by multiple imputations. Then, a spectrum of supervised ML methods, including linear regression, random forest (RF), XGBoost, k-nearest neighbours (kNN) and support vector machine (SVM) was tested (for distance-based

methods such as kNN and SVM, continuous variables were scaled before modelling). In the development dataset, tenfold cross-validation was applied for parameter tuning to obtain a minimum mean average error (MAE), calculated as the mean of the absolute difference between predictions and observations. Details on parameter calibration are displayed in electronic supplementary material (ESM) Table 1. The algorithm with the least MAE in cross-validation was called the ML-A1C model and further replicated in the external validation dataset (Figs 1, 5a).

ML-ACR model We developed a model to predict the risk of progression of albuminuria in the placebo arm of CANVAS-R ($N=2771$) and validated the model in the placebo arm of CANVAS ($N=1043$). We used a set of baseline clinical variables including age, sex, race, pulse, SBP, DBP, height, weight, BMI, HbA_{1c}, TG, TC, HDL, LDL, ACR, SCr, eGFR, BUN, background treatment, history of cardiovascular diseases and history of diabetic nephropathy to predict the progression of albuminuria (yes or no). Weight and TC were excluded after variable selection. After data processing, a spectrum of ML methods, including logistic regression, RF, XGBoost, kNN and SVM, were developed and tested by tenfold cross-validation in derivation. Parameters were tuned for the maximum average area under the receiver-operating characteristic curve (ROC AUC) in the derivation dataset (ESM Table 2). The algorithm achieving the highest ROC AUC in cross-validation was called the ML-ACR model and further replicated in the external validation dataset. We used the Youden index, calculated as the maximum of sensitivity + 1 – specificity, as the cut-off value for progression risk (Figs 1, 6a).

Statistical analysis

Participants from the mITT set from CANTATA trials and CANVAS were included. Missing data were imputed using the last observation carried forward (LOCF) approach and the last observation prior to the initiation of rescue therapy was used for participants who received glycaemic rescue therapy. For the primary efficacy outcome and other continuous endpoints, we used an ANCOVA model with different trials and corresponding baseline values as covariates to assess the least-squares (LS) mean differences and the associated two-sided 95% CIs among different drugs within each subgroup. Post hoc comparisons were performed if the overall p value was significant, and type one error was further controlled by the false discovery rate (FDR) method if applicable. The categorical efficacy endpoints (proportion of individuals achieving HbA_{1c} <53 mmol/mol [$<7.0\%$]) were analysed using a log-binomial regression model adjusted by trial and baseline HbA_{1c}.

In the clinical outcome analysis, all endpoints were tested together without a sequential conditional hypothesis, since this study was designed for an exploratory hypothesis. Annualised incidence rates per 1000 individuals were calculated. HR (95%CI) for endpoints in SGLT2i vs placebo was estimated using Cox regression models within each subgroup. In subgroup analysis, the p value for interaction ($p_{\text{interaction}}$) across subgroups was obtained through the likelihood ratio test (LRT).

For the HbA_{1c} decline endpoint, the percentage of concordance was compared using McNemar test between cluster algorithm and ML-A1C model in participants in CANVAS receiving SGLT2i treatment. For the cluster algorithm, the predicted HbA_{1c} was equal to the mean HbA_{1c} of the assigned cluster. Concordance between prediction and actual HbA_{1c} decline was defined as an absolute difference between predicted HbA_{1c} decline value and actual HbA_{1c} decline value of less than 3.28 mmol/mol (0.3%), as accepted by many clinical trials [14].

For the albuminuria progression endpoint, participants in CANVAS-R were stratified into high-risk and low-risk groups using the cut-off value of the ML-ACR model.

Sensitivity analysis was performed by repeating primary endpoints in efficacy analysis and cardio-renal outcome analysis using coordinates from ADOPT [4]. We also performed a case-complete analysis on the full cohort by using multiple imputation techniques to impute missing values in HOMA2-IR and HOMA2-B.

All statistical analyses were conducted in R version 3.6.3 (R Foundation for Statistical Computing, Vienna, Austria). Supervised ML models were constructed using the *caret* package. Multiple imputations were constructed using the *mice* package. Significance was regarded as a two-sided p value <0.05 unless otherwise specified.

Results

Cluster allocation

A total of 6365 participants were allocated to the diabetes clusters using coordinates from the ANDIS study (Fig. 1). The characteristics of the participants in the clusters were similar to those of the original ANDIS study (Fig. 2 and ESM Table 3). The median (IQR) of baseline ACR (mg/g) was 11.8 (6.0–31.6), 10.7 (5.7–29.0), 14.4 (7.5–51.5), 11.7 (6.3–34.3) ($p<0.001$) and eGFR (ml min⁻¹ [1.73 m]⁻²) was 80.0 (68.0–94.0), 88.0 (76.5–102.0), 86.0 (73.6–99.0) and 75.0 (65.0–88.0) ($p<0.001$) in the MARD, MOD, SIDD and SIRD clusters, respectively. Treatments were evenly distributed to the four clusters at randomisation (ESM Table 3).

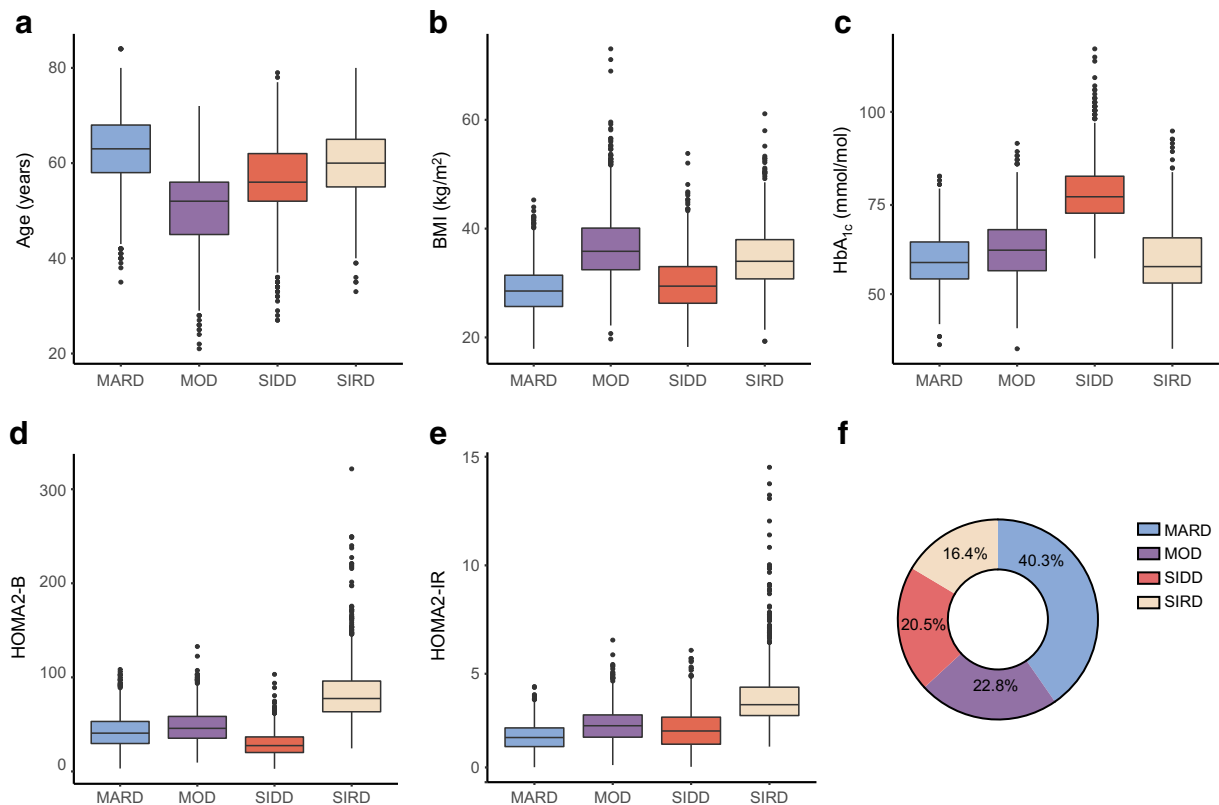


Fig. 2 Characteristics and distribution of clusters derived using coordinates from the ANDIS study ($N=6365$). (a–e) Age (a), BMI (b), HbA_{1c} (c), HOMA2-B (d) and HOMA2-IR (e) at baseline for each cluster. The line inside each box represents the median of the data, and the ends of the

box represented the first (Q1) and third (Q3) quartiles. The extreme line represents minimum and maximum values excluding outliers. Dots represented outliers (below $Q1-1.5\times IQR$ or above $Q3+1.5\times IQR$). (f) Distribution of subgroups

Glycaemic control achieved by canagliflozin in data-driven clusters

In canagliflozin-treated participants, the crude decline in HbA_{1c} at 52 weeks (95% CI) was 7.54 mmol/mol (7.1, 8.1) (0.69% [0.65, 0.74]), 9.81 mmol/mol (9.2, 10.5) (0.90% [0.84, 0.96]), 17.2 mmol/mol (16.5, 17.8) (1.57% [1.50, 1.63]) and 8.2 mmol/mol (7.4, 9.0) (0.75% [0.68, 0.82]) in MARD, MOD, SIDD and SIRD clusters, respectively ($p<0.001$). The HbA_{1c} decline was greatest in participants with MOD when baseline HbA_{1c} was adjusted for, although this difference did not reach statistical significance ($p=0.06$) (Fig. 3a and ESM Fig. 1).

There was a significant interaction between cluster and drug treatment in HbA_{1c} decline ($p_{\text{interaction}}=0.004$, Fig. 3b). In the MOD group, the LS mean (95% CI) decline of HbA_{1c} with sitagliptin (dipeptidyl peptidase-4 inhibitor [DPP4i]), glimepiride (SU) and canagliflozin was 6.6 mmol/mol (4.1, 9.2) (0.61% [0.38, 0.84]), 7.1 mmol/mol (4.7, 9.5) (0.65% [0.43, 0.87]) and 9.8 mmol/mol (9.0, 10.5) (0.90% [0.83, 0.96]), respectively, $p_{\text{SU vs SGLT2i}}=0.029$ and $p_{\text{DPP4i vs SGLT2i}}=0.029$ after FDR correction (Fig. 3b). In MARD, the highest percentage of participants reaching the HbA_{1c} control target were those treated with sitagliptin (Fig. 3c). There was no difference in either HbA_{1c} decline or the proportion of participants achieving glycaemic control

when comparing canagliflozin with the other drugs in the SIRD and SIDD groups (Fig. 3b,c).

For HbA_{1c} decline, the p value for interaction between treatment and clusters was <0.01 at all time points from week 12 to week 104. Canagliflozin caused a greater magnitude of HbA_{1c} decline than glimepiride at week 52 and 104 in the MOD group (Fig. 3e). Glimepiride yielded the largest decline in HbA_{1c} in the SIDD group compared with other OHDs at week 12 but the preponderance disappeared afterwards (Fig. 3f). For FPG decline, the p for interaction between treatment and clusters was <0.001 , 0.012 and 0.024 at week 12, 18 and 52, respectively, and >0.05 at week 52 and 104. In MOD, canagliflozin reduced FPG at the greatest magnitude compared with other OHDs, and this effect persisted until 104 weeks (Fig. 3i). The effect of body weight lowering of canagliflozin was sustained for 2 years in all subgroups (Fig. 3l-o).

Renal outcomes in data-driven clusters

In CANVAS, the progression risk of participants in the four diabetes subgroups towards albuminuria and eGFR decline was similar in the placebo arm. There was a marginal difference in the risk of developing albuminuria in canagliflozin-treated participants (ESM Fig. 2).

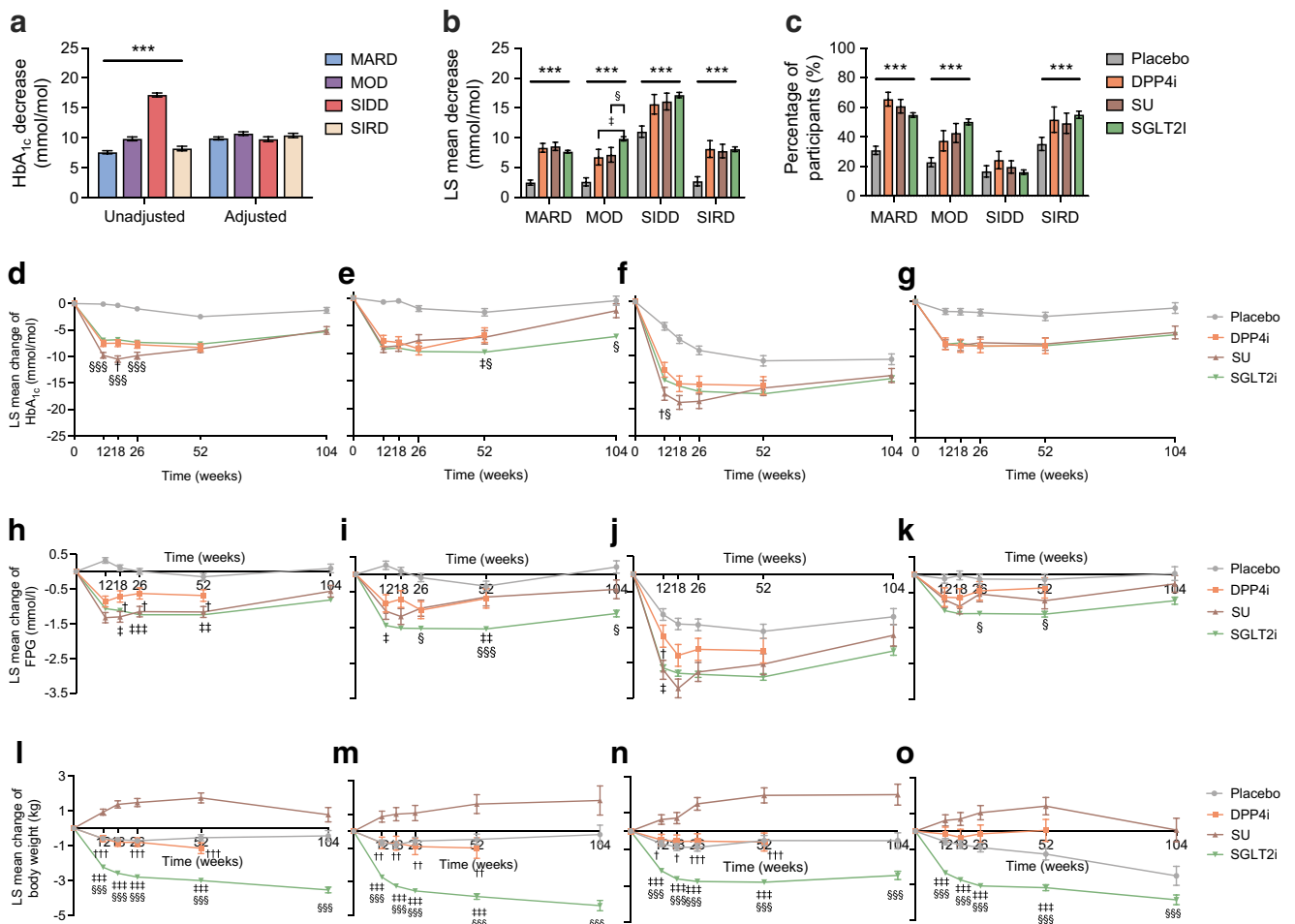


Fig. 3 Glucose-lowering efficacy of SGLT2i and DPP4i, SU and placebo in all diabetes clusters derived from the ANDIS study. **(a)** The decline of HbA_{1c} unadjusted and adjusted by baseline HbA_{1c} in each cluster treated with canagliflozin (*N*=6190). **(b)** Decline of HbA_{1c} in each cluster treated with different OHDs adjusted for baseline HbA_{1c} for each treatment (*N*=6190). *p*_{interaction}=0.004. **(c)** Percentage of participants that achieved HbA_{1c} <53 mmol/mol (<7.0%) in each subgroup (*N*=6190). *p*_{interaction} <0.001. **(d–o)** The trajectory of HbA_{1c} (**d–g**), FPG (**h–k**) and body weight (**l–o**) up to 2 years in each cluster (*N*=6365 at baseline) for participants in the MARD (**d, h, l**), MOD (**e, i, m**), SIDD (**f, j, n**) and SIRD clusters (**g, k, o**). Data are presented as mean ± SEM or % (95% CI). Continuous data

were analysed using the ANCOVA model followed by post hoc comparison and baseline values and trials were adjusted unless specified. Categorical data were analysed by log-binomial regression model adjusted by study and baseline HbA_{1c}. ****p*<0.001 for differences between OHDs within each cluster. Post hoc comparison: †*p*<0.05, ††*p*<0.01 and †††*p*<0.001 for SU vs DPP4i; ‡*p*<0.05, ‡‡*p*<0.01 and ‡‡‡*p*<0.001 for DPP4i vs SGLT2i; §*p*<0.05 and §§§*p*<0.001 for SU vs SGLT2i. The significance for multiple comparisons was adjusted by FDR if applicable. The SGLT2i treatment was canagliflozin 100 mg + 300 mg combined, the SU treatment was glimepiride and the DPP4i treatment was sitagliptin

Canagliflozin significantly reduced the risk for the renal outcome (HR 0.83 [95% CI 0.73, 0.93]), driven mainly by the effect on albuminuria progression (HR 0.78 [95% CI 0.69, 0.89]) rather than the effect on renal composite outcome (HR 0.90 [95% CI 0.65, 1.24]) (Fig. 4). The *p* for interaction of treatment by group was >0.05 for all endpoints. The HR (95% CI) of albumin progression risk with canagliflozin treatment vs placebo was 0.68 (0.52, 0.89) in MOD and 0.67 (0.51, 0.89) in SIDD.

We performed a sensitivity analysis in the case-complete dataset (ESM Figs 3–5, ESM Table 4) and the main endpoints were similar. We also used coordinates from the ADOPT study to generate four subgroups. The cluster characteristics and distribution were highly concordant with those in the ANDIS study

(Cohen κ 0.76 [95% CI 0.75, 0.78]; ESM Fig. 6, ESM Table 5). The main endpoints were also similar (ESM Figs 7, 8).

Prediction of canagliflozin-induced changes in glycaemic control by the ML-A1C model

We developed the ML-A1C model to predict the HbA_{1c} decline using canagliflozin treatment using XGBoost, which showed the best internal MAE among all algorithms (Fig. 5 and ESM Table 6). The external MAE was similar between the clustering algorithm and the ML-A1C model (*p*=0.936) (Fig. 5b and ESM Table 6). HbA_{1c}, HOMA2-B, ACR, HOMA2-IR and HDL were the top five important variables for the prediction (Fig. 5c). The concordance rate, which was

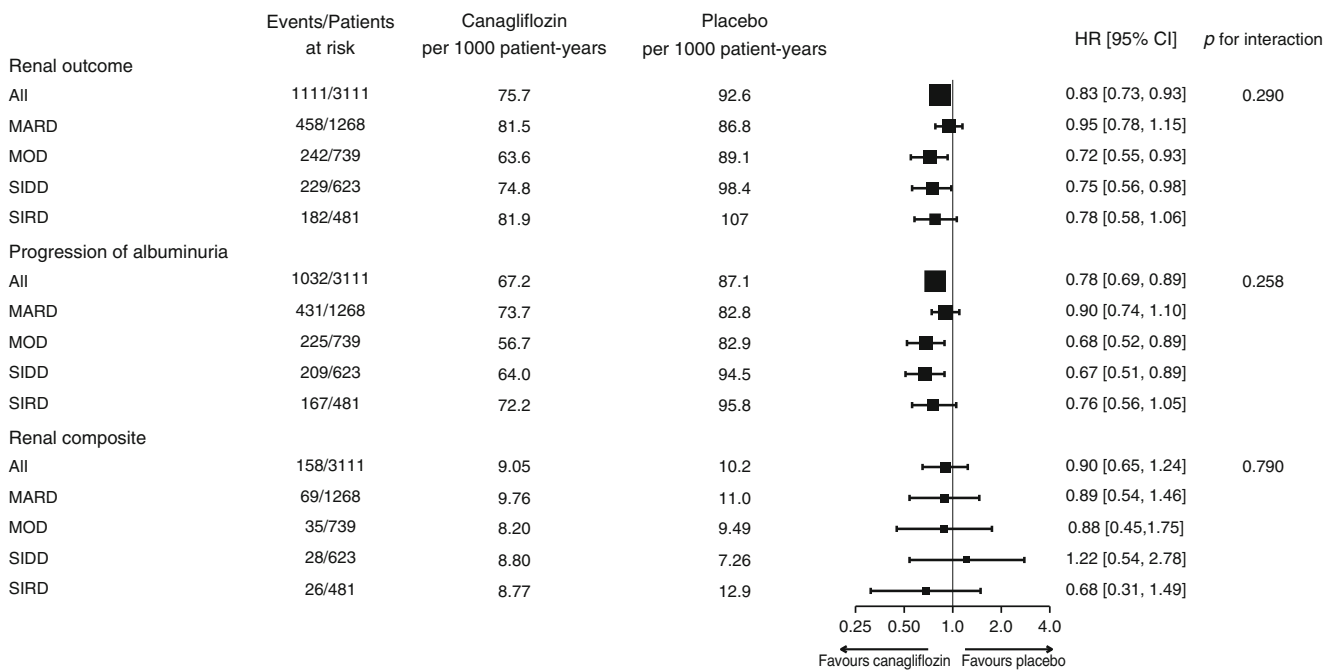


Fig. 4 Forest plot of effects of SGLT2i vs placebo on renal outcomes in CANVAS ($N=3111$) in clusters derived from the ANDIS study. Events are shown per 1000 patient-years in the canagliflozin vs placebo arms in the overall population, and in subgroups. Renal outcome included

progression of albuminuria and renal composite (eGFR decline, end-stage renal disease, doubling of SCr and renal death). HR with 95% CI was estimated using the Cox regression model, and $p_{\text{interaction}}$ (treatment by group) was tested using the LRT method

the percentage of participants with predicted HbA_{1c} similar to actual HbA_{1c} , showed no difference between the cluster strategy (31.4%) and the ML-A1C model (31.5%) ($p=0.966$).

Renal effect of canagliflozin in high-risk patients identified by the ML-ACR model

Since the major difference between SGLT2i vs placebo in the risk of outcomes was in albuminuria progression, we developed ML-based algorithms for the albuminuria progression endpoint (Fig. 6a and ESM Table 7). The maximum internal ROC AUC of the model was achieved using XGBoost (0.71 [95% CI 0.67, 0.74]) (Fig. 6b), which was selected for the ML-ACR model. The participants in the validation cohort were divided into high-risk and low-risk groups using the optimal cut-off value of ML-ACR. The top five most important indicators for albumin progression were ACR, BUN, eGFR, HbA_{1c} and LDL (Fig. 6c). There was a one-third decline in the risk of albuminuria progression in high-risk participants treated with canagliflozin (HR vs placebo 0.67 [95% CI 0.57, 0.80]) but not in low-risk patients (HR 0.91 [95% CI 0.75, 1.11]; $p_{\text{interaction}}$ [treatment by group]=0.016).

Discussion

In this study, we replicated the clusters derived from the ANDIS study and ADOPT. Canagliflozin treatment

outperformed other drugs in HbA_{1c} decline in individuals with MOD from 52 weeks up to 104 weeks. There was no significant difference in the effect of canagliflozin on renal endpoints (vs placebo) among the different diabetes clusters. The ML-A1C model showed a similar concordance rate with the cluster strategy in predicting the HbA_{1c} -lowering effect of canagliflozin. Canagliflozin reduced the risk of albuminuria progression in high-risk individuals identified by the ML-ACR model.

Our data highlight the possibility of using the data-driven diabetes clusters to guide clinical decisions on use of SGLT2i regarding their glucose-lowering effect. Canagliflozin showed a better and sustained glucose-lowering effect in MOD, so SGLT2i could be possibly recommended to individuals with MOD at diagnosis for long-term glucose control. Glimepiride treatment showed a good glycaemic reduction in the first 12 weeks in SIDD, so an SU could be an option for short-term glycaemic control for SIDD. Similarly, in ADOPT, the SIRD group achieved better glycaemic control with TZD and the SIDD and MARD groups showed robust HbA_{1c} decline in less than half a year with SU [7]. The MARD group consisted mainly of older participants and hypoglycaemia may be the major safety consideration in these individuals. A DDP4i could be a better option for glycaemic control in individuals with MARD. However, there was a lack of consistency in the HbA_{1c} -lowering effect and $\text{HbA}_{1c}<53$ mmol/mol (7.0%) rate in the MOD group, possibly due to the limited sample size

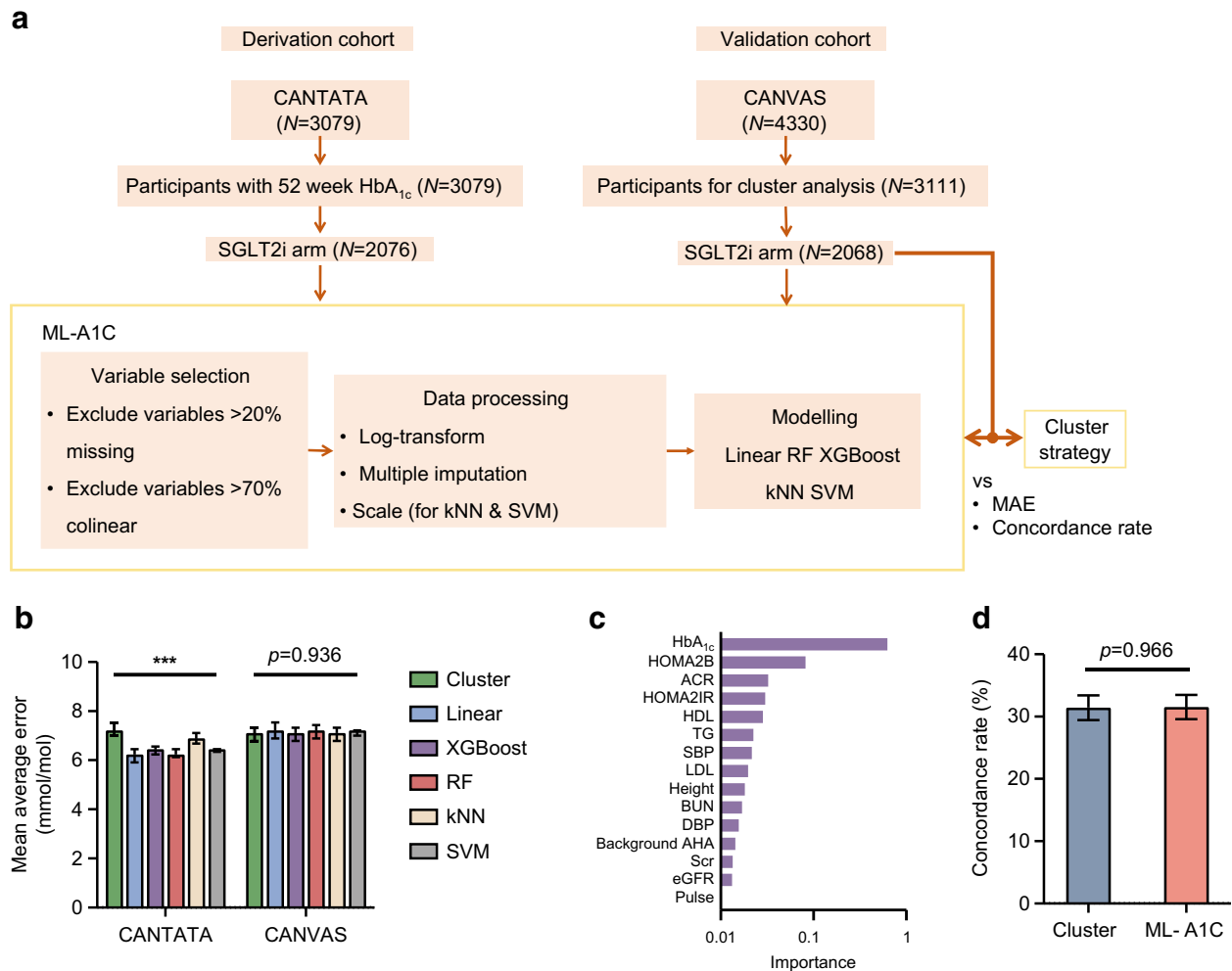


Fig. 5 Derivation, validation and application of the ML-A1C model and its comparison with data-driven clusters. **(a)** Flowchart for the ML-A1C model development. Participants treated with canagliflozin in the CANTATA and CANVAS trials were used for model development and model validation, respectively. Variables with more than 20% of data missing or 70% colinear with existing variables were excluded. Tenfold cross-validation was applied. Supervised learning methods, including linear regression, RF, XGBoost, kNN and SVM, were tested to predict the decline of HbA_{1c}. Continuous variables were log-transformed or scaled if applicable. Missing values were replaced by multiple

imputations. Parameters were calibrated to achieve the lowest MAE in CANTATA. External validation was conducted in CANVAS. XGBoost was selected as the ML-A1C model. **(b)** Internal (CANTATA) and external (CANVAS) MAE (95% CI) in cluster algorithm and ML methods. *** $p < 0.001$ tested by one-way ANOVA. **(c)** Importance scores of variables in predicting endpoint in the ML-A1C model. **(d)** Concordance rate (95% CI) was defined as predicted HbA_{1c} decline similar to actual HbA_{1c} decline, between the cluster strategy and the ML-A1C model ($p = 0.966$, tested using McNemar test)

($N = 1449$). A significant interaction of treatment by group in HbA_{1c} on-target rate was mainly driven by the fact that more participants in the MARD group achieved glycaemic control targets when treated with DPP4i than when treated with SGLT2i. Our results provided exploratory evidence that the glycaemic response to canagliflozin may differ between diabetes clusters and further studies are necessary before data-driven clusters can be used to guide clinical practice.

SGLT2i showed robust renal protection effects including reducing the progression of albuminuria and eGFR decline [2, 3, 15]. Our study showed that the interactions between SGLT2i and clusters were not statistically significant, despite a reduction in the risk of albuminuria progression being

observed in the MOD and SIDD groups. The main reason for this observation was that the sample size may not have been sufficient to enable detection of a treatment-by-group difference for the renal outcome (7169 participants were necessary to detect a significant difference between MARD and other groups with the current effect size). Another possibility was that data-driven clusters were not able to stratify the progression risk towards renal outcomes in our study. For example, it was suggested in previous studies that SIRD had a rapid progression towards eGFR decline compared with other subgroups [5]. Similar to previous studies, our data showed that participants in the SIRD group had the lowest baseline eGFR [7, 16]. However, the progression of

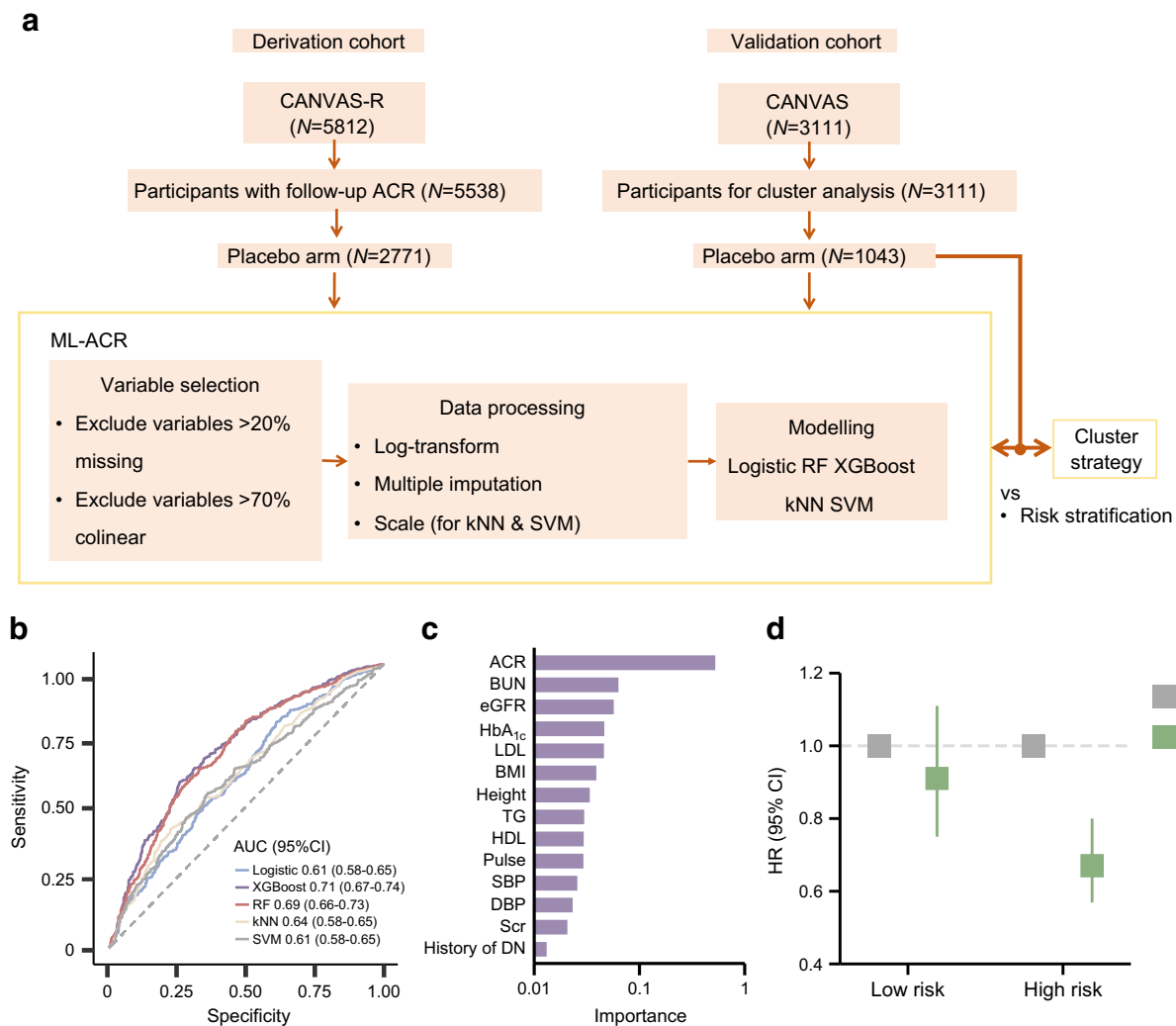


Fig. 6 Derivation, validation and application of the ML-ACR model. **(a)** Flowchart for the ML-ACR model. Participants treated with placebo in CANVAS-R and CANVAS were used for model development and model validation, respectively. Variables with more than 20% data missing or 70% colinear with existing variables were excluded. Tenfold cross-validation was applied. Supervised learning methods, including linear regression, RF, XGBoost, kNN and SVM, were tested to predict the decline of HbA_{1c}. Continuous variables were log-transformed or scaled if applicable. Missing values were replaced by multiple imputation.

Parameters were adjusted to reach the highest ROC AUC in the internal validation dataset. **(b)** External ROC AUC in all supervised learning methods. **(c)** Importance scores of variables in the ML-ACR model. The model developed using XGBoost was selected as the ML-ACR model. **(d)** HR (95% CI) for progression of albuminuria following treatment with canagliflozin and placebo in high-risk participants ($N=1874$) and low-risk participants ($N=1237$) stratified using the ML-ACR model. $p_{\text{interaction}}$ (treatment by group)=0.016, using the LRT method. DN, diabetic nephropathy

participants with SIRD towards eGFR decline was similar to that seen in other groups in the placebo-treated arm in the CANVAS trial (ESM Fig. 2). The discrepancy may be attributable to differences in disease duration (new-onset in ANDIS and 15 years in CANVAS), CVD history (60% in CANVAS) and follow-up time (15 years in ANDIS and 4.5 years in CANVAS) between our study and previous studies [5, 16, 17]. Further studies with larger sample sizes or subtype-specific studies are necessary to investigate whether there was a treatment-by-group difference in renal outcomes and even cardiovascular outcomes.

Data-driven clustering and supervised ML are two distinct methods for facilitating precision medicine in patients. In our

study, supervised learning algorithms showed a similar concordance rate in prediction of HbA_{1c} decline by canagliflozin using the data-driven clustering algorithm. Three of the five variables for cluster development were among the top five important variables for predicting the glucose-lowering response to canagliflozin. By this means, data-driven clustering is more likely to be accepted as a tool for clinical practice, because it had the advantage of stability across a spectrum of studies and ethnicities, whereas the external validity of the ML-A1C model has yet to be tested in other cohorts.

The ML-ACR model precisely predicted the progression risk for albuminuria. Although the diagnostic accuracy of the ML-ACR model was around 70%, which can hardly meet the

clinical diagnostic criteria, it was not inferior to other models developed in large cohorts to predict chronic kidney disease or albuminuria in participants with diabetes [18, 19]. In fact, data-driven clusters are not as effective as simple clinical variables in predicting drug response, as Dennis et al proposed [7]. Our study suggested that SGLT2i yielded the highest magnitude of albuminuria progression risk reduction in participants aged 55–65 years or with BMI ≤ 25 kg/m² and >30 kg/m² (ESM Fig. 9). As a replacement for data-driven clusters, the ML-ACR model successfully stratified the participants into high-risk and low-risk groups and the effect of SGLT2i vs placebo was different between the two groups. Of the top five variables for cluster development, only HbA_{1c} was ranked in the top five important variables for the prediction of albuminuria progression. This might partly explain why the clustering algorithm was not able to stratify the albuminuria development risk.

Our study had the strength of combining canagliflozin trials to investigate not only the glucose-lowering effect of SGLT2i but also renal outcomes. This is the first study to our knowledge to explore the role of SGLT2i in data-driven diabetes clusters. This study also proposed a new model, ML-ACR, which may potentially promote precision SGLT2i usage if validated. However, there are some limitations to this study. First, we were unable to replicate our models, especially the supervised ML models, in other cohorts due to a lack of availability. Thus, the external validity of the ML-A1C and ML-ACR models was unknown. Second, since C-peptide or insulin values were required for data-driven clustering, we lost more than half of the participants of the CANVAS program when analysing the clinical outcomes. The comparison to identify treatment-by-group differences among four clusters was underpowered in the renal outcome analysis, so the question of whether data-driven clusters responded differently to SGLT2i treatment is yet to be answered. We may need other trials of SGLT2i (e.g. Empagliflozin Cardiovascular Outcome Event Trial in Type 2 Diabetes Mellitus Patients [EMPA-REG] and Dapagliflozin on the Incidence of Cardiovascular Events [DECLARE]) to enlarge the sample size and replicate our models. In addition, subgroup-specific clinical trials should be implemented to test the effect of SGLT2i among subgroups. Third, we were unable to adjust the results for the effect of diabetes duration due to a lack of information, so we replaced the age-of-onset in the original ANDIS study with age to generate the clusters. Nevertheless, clusters may vary during the disease progression [16], and the randomisation of clinical trials can be regarded as the start of an observation period. Fourth, we were unable to exclude participants who may have had severe autoimmune-related diabetes (SAID) from the analysis due to a lack of GADab information. Since all participants were diagnosed with type 2 diabetes, the proportion of participants with GADab positive may be negligible.

Conclusion

Our data provided exploratory evidence that using an SGLT2i in MOD may improve long-term glycaemic control and that using a DPP4i may achieve a higher glycaemic on-target rate in MARD. Developing supervised learning models for albuminuria progression may help to precisely identify individuals who could benefit from SGLT2i treatment. Further studies with external replication and subtype-specific clinical trials are necessary to build up evidence for guiding clinical decisions on SGLT2i use in precision stratification strategies.

Supplementary Information The online version contains peer-reviewed but unedited supplementary material available at <https://doi.org/10.1007/s00125-022-05748-9>.

Data availability The application for the clinical trial data source is available on the YODA website (<http://yoda.yale.edu/>). Models developed in this article are kept by the corresponding author and are available upon reasonable request.

Funding The main sources of funding to the authors to support their work on SGLT2i and diabetes clustering during the time of the study were as follows: Beijing Nova Program of Science and Technology (Z191100001119026) and Beijing Nova Cross program (Z211100002121169) (to XZo); the National Natural Science Foundation of China (81970708 to LJ and 81800515 to XZo); Beijing Municipal Science and Technology Commission (Z201100005520013 to LJ and Z201100005520012 to XH); and National Key R&D Program of China (2016YFC1304901 to LJ and 2016YFC1305603 to XZh).

Authors' relationships and activities The authors declare that there are no relationships or activities that might bias, or be perceived to bias, their work.

Contribution statement XZo designed the study, applied the dataset, participated in data analysis and wrote the manuscript. QH applied for the dataset, analysed the data and contributed to manuscript drafting. YL, QR, XH and XZh contributed to data analysis and manuscript revision. LJ designed the study, applied the dataset and revised the manuscript. XZo is responsible for the integrity of the work as a whole. All authors approved the final version.

References

1. Perkovic V, Jardine MJ, Neal B et al (2019) Canagliflozin and renal outcomes in type 2 diabetes and nephropathy. *N Engl J Med* 380(24):2295–2306. <https://doi.org/10.1056/NEJMoa1811744>
2. Neal B, Perkovic V, Mahaffey KW et al (2017) Canagliflozin and cardiovascular and renal events in type 2 diabetes. *N Engl J Med* 377(7):644–657. <https://doi.org/10.1056/NEJMoa1611925>
3. Zinman B, Wanner C, Lachin JM et al (2015) Empagliflozin, cardiovascular outcomes, and mortality in type 2 diabetes. *N Engl J Med* 373(22):2117–2128. <https://doi.org/10.1056/NEJMoa1504720>
4. Chung WK, Erion K, Florez JC et al (2020) Precision Medicine in Diabetes: A Consensus Report From the American Diabetes Association (ADA) and the European Association for the Study of Diabetes (EASD). *Diabetes Care* 43(7):1617–1635. <https://doi.org/10.2337/dci20-0022>

5. 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)
6. Zou X, Zhou X, Zhu Z, Ji L (2019) Novel subgroups of patients with adult-onset diabetes in Chinese and US populations. *Lancet Diabetes Endocrinol* 7(1):9–11. [https://doi.org/10.1016/S2213-8587\(18\)30316-4](https://doi.org/10.1016/S2213-8587(18)30316-4)
7. 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)
8. Ahlqvist E, Tuomi T, Groop L (2019) Clusters provide a better holistic view of type 2 diabetes than simple clinical features. *Lancet Diabetes Endocrinol* 7(9):668–669. [https://doi.org/10.1016/S2213-8587\(19\)30257-8](https://doi.org/10.1016/S2213-8587(19)30257-8)
9. Ravizza S, Huschto T, Adamov A et al (2019) Predicting the early risk of chronic kidney disease in patients with diabetes using real-world data. *Nat Med* 25(1):57–59. <https://doi.org/10.1038/s41591-018-0239-8>
10. Stenlöf K, Cefalu WT, Kim KA et al (2013) Efficacy and safety of canagliflozin monotherapy in subjects with type 2 diabetes mellitus inadequately controlled with diet and exercise. *Diabetes Obes Metab* 15(4):372–382. <https://doi.org/10.1111/dom.12054>
11. Lavalle-González FJ, Januszewicz A, Davidson J et al (2013) Efficacy and safety of canagliflozin compared with placebo and sitagliptin in patients with type 2 diabetes on background metformin monotherapy: a randomised trial. *Diabetologia* 56(12):2582–2592. <https://doi.org/10.1007/s00125-013-3039-1>
12. Cefalu WT, Leiter LA, Yoon KH et al (2013) Efficacy and safety of canagliflozin versus glimepiride in patients with type 2 diabetes inadequately controlled with metformin (CANTATA-SU): 52 week results from a randomised, double-blind, phase 3 non-inferiority trial. *Lancet* 382(9896):941–950. [https://doi.org/10.1016/s0140-6736\(13\)60683-2](https://doi.org/10.1016/s0140-6736(13)60683-2)
13. Levy JC, Matthews DR, Hermans MP (1998) Correct homeostasis model assessment (HOMA) evaluation uses the computer program. *Diabetes Care* 21(12):2191–2192. <https://doi.org/10.2337/diacare.21.12.2191>
14. European Medicines Agency (2012) Guideline on clinical investigation of medicinal products in the treatment or prevention of diabetes mellitus. Available from https://www.ema.europa.eu/en/documents/scientific-guideline/guideline-clinical-investigation-medicinal-products-treatment-prevention-diabetes-mellitus-revision_en.pdf, accessed 5 April 2022
15. Wiviott SD, Raz I, Bonaca MP et al (2019) Dapagliflozin and Cardiovascular Outcomes in Type 2 Diabetes. *N Engl J Med* 380(4):347–357. <https://doi.org/10.1056/NEJMoa1812389>
16. 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)
17. Scheen AJ (2018) Cardiovascular outcome studies in type 2 diabetes: Comparison between SGLT2 inhibitors and GLP-1 receptor agonists. *Diabetes Res Clin Pract* 143:88–100. <https://doi.org/10.1016/j.diabres.2018.06.008>
18. Khitan Z, Nath T, Santhanam P (2021) Machine learning approach to predicting albuminuria in persons with type 2 diabetes: An analysis of the LOOK AHEAD Cohort. *J Clin Hypertens (Greenwich)* 23(12):2137–2145. <https://doi.org/10.1111/jch.14397>
19. Makino M, Yoshimoto R, Ono M et al (2019) Artificial intelligence predicts the progression of diabetic kidney disease using big data machine learning. *Scientific Reports* 9(1):11862. <https://doi.org/10.1038/s41598-019-48263-5>

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