A five-drug class model using routinely available clinical features to optimise prescribing in type 2 diabetes: a prediction model development and validation study

*John M Dennis, Katherine G Young, Pedro Cardoso, Laura M Güdemann, Andrew P McGovern, Andrew Farmer, Rury R Holman, Naveed Sattar, Trevelyan J McKinley, Ewan R Pearson, Angus G Jones, Beverley M Shields, Andrew T Hattersley, on behalf of the MASTERMIND Consortium**

Summary

Background Data to support individualised choice of optimal glucose-lowering therapy are scarce for people with type 2 diabetes. We aimed to establish whether routinely available clinical features can be used to predict the relative glycaemic effectiveness of five glucose-lowering drug classes.

Methods We developed and validated a five-drug class model to predict the relative glycaemic effectiveness, in terms of absolute 12-month glycated haemoglobin (HbA_{1c}), for initiating dipeptidyl peptidase-4 inhibitors, glucagon-like peptide-1 receptor agonists, sodium–glucose co-transporter-2 inhibitors, sulfonylureas, and thiazolidinediones. The model used nine routinely available clinical features of people with type 2 diabetes at drug initiation as predictive factors (age, duration of diabetes, sex, and baseline HbA_{1c}, BMI, estimated glomerular filtration rate, HDL cholesterol, total cholesterol, and alanine aminotransferase). The model was developed and validated with observational data from England (Clinical Practice Research Datalink [CPRD] Aurum), in people with type 2 diabetes aged 18–79 years initiating one of the five drug classes between Jan 1, 2004, and Oct 14, 2020, with holdback validation according to geographical region and calendar period. The model was further validated in individual-level data from three published randomised drug trials in type 2 diabetes (TriMaster three-drug crossover trial and two parallel-arm trials [NCT00622284 and NCT01167881]). For validation in CPRD, we assessed differences in observed glycaemic effectiveness between matched (1:1) concordant and discordant groups receiving therapy that was either concordant or discordant with model-predicted optimal therapy, with optimal therapy defined as the drug class with the highest predicted glycaemic effectiveness (ie, lowest predicted 12-month HbA_{1c}). Further validation involved pairwise drug class comparisons in all datasets. We also evaluated associations with long-term outcomes in model-concordant and model-discordant groups in CPRD, assessing 5-year risks of glycaemic failure (confirmed HbA_{1c} ≥69 mmol/mol), all-cause mortality, major adverse cardiovascular events or heart failure (MACE-HF) outcomes, renal progression, and microvascular complications using Cox proportional hazards regression adjusting for relevant demographic and clinical covariates.

Findings The five-drug class model was developed from 100 107 drug initiations in CPRD. In the overall CPRD cohort (combined development and validation cohorts), 32 305 (15·2%) of 212 166 drug initiations were of the model-predicted optimal therapy. In model-concordant groups, mean observed 12-month HbA_{1c} benefit was 5·3 mmol/mol (95% CI 4·9–5·7) in the CPRD geographical validation cohort (n=24 746 drug initiations, n=12 373 matched pairs) and 5·0 mmol/mol (4·3–5·6) in the CPRD temporal validation cohort (n=9682 drug initiations, n=4841 matched pairs) compared with matched model-discordant groups. Predicted HbA_{1c} differences were well calibrated with observed HbA_{1c} differences in the three clinical trials in pairwise drug class comparisons, and in pairwise comparisons of the five drug classes in CPRD. 5-year risk of glycaemic failure was lower in model-concordant versus model-discordant groups in CPRD (adjusted hazard ratio [aHR] 0·62 [95% CI 0·59–0·64]). For long-term non-glycaemic outcomes, model-concordant versus model-discordant groups had a similar 5-year risk of all-cause mortality (aHR 0·95 [0·83–1·09]) and lower risks of MACE-HF outcomes (aHR 0·85 [0·76–0·95]), renal progression (aHR 0·71 [0·64–0·79]), and microvascular complications (aHR 0·86 [0·78–0·96]).

Interpretation We have developed a five-drug class model that uses routine clinical data to identify optimal glucose-lowering therapies for people with type 2 diabetes. Individuals on model-predicted optimal therapy had lower 12-month HbA_{1c}, were less likely to need additional glucose-lowering therapy, and had a lower risk of diabetes complications than individuals on non-optimal therapy. With setting-specific optimisation, the use of routinely collected parameters means that the model is easy to introduce to clinical care in most countries worldwide.

Funding UK Medical Research Council.

Published Online February 25, 2025 https://doi.org/10.1016/S0140-6736(24)02617-5

See Online/Comment https://doi.org/10.1016/S0140-6736(24)02844-7

*Other members of the MASTERMIND Consortium are listed in the appendix (p 31)

Clinical and Biomedical Sciences, University of Exeter Medical School, Exeter, UK (J M Dennis PhD, K G Young PhD, P Cardoso PhD, L M Güdemann PhD, A P McGovern MD, T J McKinley PhD, Prof A G Jones PhD, B M Shields PhD, Prof Andrew T Hattersley DM); **Nuffield Department of Primary Care Health Sciences, University of Oxford, Oxford, UK** (Prof A Farmer DM); **Diabetes Trials Unit, Radcliffe Department of Medicine, University of Oxford, Oxford, UK** (Prof R R Holman FMedSci); **NIHR Oxford Biomedical Research Centre, Churchill Hospital, Oxford, UK** (Prof R R Holman); **Institute of Cardiovascular and Medical Sciences, University of Glasgow, Glasgow, UK** (Prof N Sattar MD); **Division of Diabetes, Endocrinology and Reproductive Medicine, Ninewells Hospital and Medical School, University of Dundee, Dundee, UK** (Prof E R Pearson PhD)

Correspondence to: Assoc Prof John M Dennis, Clinical and Biomedical Sciences, University of Exeter Medical School, Exeter, EX1 2HZ, UK **j.dennis@exeter.ac.uk**

See Online for appendix

Research in context

Evidence before this study

Several glucose-lowering drug classes with different mechanisms of action are available to lower blood glucose for people with type 2 diabetes. Current guidelines do not provide information to support targeted treatment based on differences in glucose-lowering response between the available drug classes. We searched PubMed and MEDLINE for articles published from Jan 1, 2004 to Sept 20, 2024, reporting observational, trial, or meta-analysis studies of 100 or more individuals that assessed whether patient characteristics are associated with differences in glucose-lowering response (change in glycated haemoglobin [HbA_{1c}] from baseline, with at least 3 months of follow-up) after initiating five major type 2 diabetes drug classes (dipeptidyl peptidase-4 [DPP-4] inhibitors, glucagon-like peptide-1 receptor [GLP-1R] agonists, sulfonylureas, sodium–glucose co-transporter-2 [SGLT2] inhibitors, and thiazolidinediones). We identified multiple observational and clinical trial studies reporting that glucose-lowering response was associated with specific characteristics of people with type 2 diabetes, and four of these studies reported approaches to using clinical characteristics to help decide between two drug classes. No studies had developed an approach to allow robust prediction of optimising glucose-lowering with more than two drug classes.

Added value of this study

We have developed and validated the first prediction model, to our knowledge, that informs on choice between five major drug classes (DPP-4 inhibitors, GLP-1R agonists, sulfonylureas, SGLT2 inhibitors, and thiazolidinediones) for people with type 2 diabetes based on individual-level differences in absolute 12-month HbA_{1c}, as a measure of glycaemic effectiveness. The model requires only clinical features and laboratory parameters that are routinely collected in people with type 2 diabetes. The model performed well in external validation, including in individual participant-level reanalysis in randomised clinical drug trial datasets. The model showed good calibration in predicting the relative glycaemic effectiveness of the five drug classes in unselected population-based data of people with type 2 diabetes in England, with accurate prediction of an approximate 5 mmol/mol greater reduction in 12-month HbA_{1c} between model-concordant drug initiations (the drug class initiated was the model-predicted optimal therapy, as the drug class with the highest predicted glycaemic effectiveness) and matched model-discordant initiations. Furthermore, model-optimal therapy was associated with a 38% lower risk of glycaemic failure over 5 years, and lower 5-year risks of major adverse cardiovascular events, renal progression, and microvascular complications than model-discordant therapy. Even in recent years since 2019 in England, only 17·8% of therapy initiations for type 2 diabetes were of model-predicted optimal treatment in clinical practice, highlighting great potential for model deployment to improve population-level glycaemic control.

Implications of all the available evidence

Lowering glucose in type 2 diabetes is known to decrease the risk of microvascular and macrovascular complications. This five-drug class model provides new and accurate information for an individual with type 2 diabetes on the likely relative effectiveness of five major glucose-lowering drug classes available globally for treatment after initial metformin. Associations of model-optimal therapy with notable improvements in glycaemic outcomes makes the model highly relevant to most people with type 2 diabetes. When making prescribing decisions, information on the likely optimal therapy for glycaemia at the individual patient level can now be considered alongside established non-glycaemic drug-specific cardiorenal benefits, side effects, and the availability and cost of medication. As the model is based solely on routine clinical data, low-cost deployment to inform treatment decisions in type 2 diabetes is possible in most countries.

Introduction

Lowering blood glucose is a major aim of treatment in type 2 diabetes and is important for both symptom relief and preventing long-term microvascular and macrovascular complications. Long-term follow-up of the UK Prospective Diabetes Study has shown that the benefit of glucose lowering in terms of reducing death, myocardial infarction, and microvascular disease is maintained for up to 24 years after early intensive glycaemic control in people with type 2 diabetes.[1] The progressive deterioration in glycaemic control characteristic of type 2 diabetes[2] means that most people will require additional therapies after initial pharmacological treatment. Generally, the more effective an additional agent is, the longer it will be before further glucose-lowering medication is required. This principle provides a good rationale for optimising glucose-lowering therapies for both short-term and long-term outcomes.

Optimising glucose lowering requires identification of the drug class that is likely to be the most effective for an individual patient.[3] Current type 2 diabetes treatment guidelines provide limited information to stratify patients, predominantly based on cardiorenal risk, and do not guide on how to optimise glucose-lowering for individual patients.[4] For most individuals, several glucose-lowering therapies are considered appropriate in current guidance.[4] A pragmatic way to inform optimal treatment alongside stratification by cardiorenal risk is to use individual-level clinical features to predict the relative glycaemic effectiveness of available drug classes.[5] Prioritising features that are routinely collected in the diabetes clinic would allow a targeted approach to informing optimal treatment that could be readily

implemented at low cost. The considerable potential of this approach has been previously indicated, with robust heterogenous treatment effects for glucose lowering based on routinely available clinical features identified for five major drug classes (dipeptidyl peptidase-4 [DPP-4] inhibitors, glucagon-like peptide-1 receptor [GLP-1R] agonists, sodium–glucose co-transporter-2 [SGLT2] inhibitors, sulfonylureas, and thiazolidinediones) in observational and clinical trial data.[6–10]

To date, type 2 diabetes treatment selection models have been developed for two-drug class comparisons of glycaemic responses,[8,9] but have not considered all major non-insulin treatment options after first-line metformin in a single model. We aimed to develop and validate a model to predict relative differences in the glycaemic effectiveness of five drug classes (DPP-4 inhibitors, GLP-1R agonists, SGLT2 inhibitors, sulfonylureas, and thiazolidinediones) using routinely available clinical features, and estimate the potential effect of glycaemia-based treatment optimisation on risk of diabetes complications, including cardiovascular events and renal progression.

Methods

Data sources

For model development and initial validation we used observational data from the Clinical Practice Research Datalink (CPRD) Aurum database (October, 2021 release).[11] CPRD Aurum contains broadly population-representative anonymised longitudinal electronic health records from primary care practices in England,[11] and covered 19·3% of the UK population when data were accessed. Primary care records were linked by CPRD to secondary care records (Hospital Episode Statistics Admitted Patient Care), deprivation records (2015 English Index of Multiple Deprivation[12]), and death certificate records (Office for National Statistics). For further model validation, individual participant-level data from three randomised drug trials in type 2 diabetes were accessed: the UK-based TriMaster three-way cross-over trial (DPP-4 inhibitor [sitagliptin], SGLT2 inhibitor [canagliflozin], and thiazolidinedione [pioglitazone]),[10] and two multinational active comparator trials: NCT00622284 (DPP-4 inhibitor [linagliptin] *vs* sulfonylurea [glimepiride]) and NCT01167881 (SGLT2 inhibitor [empagliflozin] *vs* sulfonylurea [glimepiride]).[13,14]

Approval for the current study was granted by the CPRD Independent Scientific Advisory Committee (Electronic Research Applications Portal protocol number 22_002000) and Vivli (data request ID number 00005959). CPRD obtains annual research ethics approval from the UK's Health Research Authority Research Ethics Committee (East Midlands—Derby; reference number 21/EM/0265) to receive and supply patient data for public health research; no further ethical permissions were required for the analyses of these anonymised patient-level data. All trial protocols were approved by local institutional review boards, and all trial participants provided informed consent; no further ethical permissions beyond approval of the data request were required for the secondary analyses of de-identified trial data.

Study population and drug initiations

In CPRD, we identified people aged 18–79 years with type 2 diabetes initiating for the first time one of five glucose-lowering medication classes: DPP-4 inhibitor (alogliptin, linagliptin, saxagliptin, sitagliptin, or vildagliptin), GLP-1R agonist (liraglutide, dulaglutide, or exenatide prolonged-release; lixisenatide and exenatide standard-release were omitted from our study due to their lower average efficacy,[15] and semaglutide was not evaluated due to the low number [n=679] of study-eligible drug initiations with complete predictor and glycated haemoglobin [HbA_{1c}] outcome data), SGLT2 inhibitor (empagliflozin, canagliflozin, or dapagliflozin), sulfonylurea (gliclazide), and thiazolidinedione (pioglitazone). We included individuals initiating these drug classes between Jan 1, 2004, and Oct 14, 2020, excluding initiations in which individuals received the medication as first-line glucose-lowering therapy (as this is not recommended in current treatment guidelines[4]), those with concurrent treatment with insulin, or those with a pre-existing diagnosis of end-stage kidney disease. Therapy initiations were identified by the date of first prescription of the drug classes of interest. We then excluded individuals with missing baseline HbA_{1c} values (to obtain stable HbA_{1c} values, baseline was defined as the closest value to drug initiation within a cutoff time window of –183 days to +7 days from baseline:[16] excluding individuals without a HbA_{1c} value in this window) or with baseline HbA_{1c} below 53 mmol/mol or above 110 mmol/mol. Further cohort details are available online. Individuals were eligible for separate inclusion in multiple drug class groups if they initiated more than one drug class over the study period and met eligibility criteria. Hereafter, reported numbers refer to the number of drug initiations rather than individual patients. Eligible clinical trial cohorts comprised all randomly assigned participants (age ranges: 30–80 years in TriMaster, 18–80 years in NCT00622284, and ≥18 years in NCT01167881) with complete baseline and HbA_{1C} outcome data for model validation.

For **full cohort details** see https://github.com/Exeter-Diabetes/CPRD-Cohort-scripts/tree/main/03-Treatment-response-(MASTERMIND)

Outcomes

The primary outcome predicted by the model was HbA_{1c} value at 12 months after drug initiation (12-month HbA_{1c}) on unchanged glucose-lowering therapy for each drug class. In CPRD, this value was defined as the closest HbA_{1c} measurement to 12 months after initiation (range 3–15 months) with no addition or cessation of other glucose-lowering medications, and continued prescription of the drug of interest. Due to its crossover design, each participant in TriMaster had a single outcome HbA_{1c} measure 3–4 months after initiation of each medication,

each of which comprised our outcome here. In the two active comparator trials, our outcome was last-observation-carried-forward HbA_{1c} at the 12-month study visit. We report all HbA_{1c} values in mmol/mol. A 1 mmol/mol difference in HbA_{1c} translates to a 0·09% change on the percentage scale.

Long-term secondary outcomes to assess potential effects of model-predicted therapy optimisation were evaluated in CPRD up to 5 years after drug initiation, and comprised time to glycaemic failure (confirmed HbA_{1c} ≥69 mmol/mol), all-cause mortality, major adverse cardiovascular events or heart failure (MACE-HF) outcomes (comprising hospital admission for myocardial infarction, stroke, or heart failure as the primary reason, or death from cardiovascular disease or heart failure as the primary cause), renal progression (>40% decrease in estimated glomerular filtration rate [eGFR] or end-stage kidney disease[17]), and microvascular complications (a composite based on whichever occurred first of progression to clinically significant albuminuria [urinary albumin-to-creatinine ratio >30 mg/g],[4] or severe retinopathy [secondary care admission for vitreous haemorrhage or laser photocoagulation, adapted from Adler et al[1]]).

Predictive and prognostic features

In this study, we distinguished between predictive features that predict differences in the relative glycaemic effectiveness of the different drug classes, thus assisting selection of treatment, and prognostic factors that predict glycaemic effectiveness independently of treatment. Nine clinical features were prespecified as predictive factors, based on previously observed heterogenous treatment effects,[6–10] plausible relations to drug mechanism of action, and availability in primary care records for most individuals. The features were: current age, duration of diabetes, sex, and baseline HbA_{1c}, BMI, eGFR (Chronic Kidney Disease Epidemiology Collaboration formula[18]), HDL cholesterol, total cholesterol, and alanine aminotransferase. Features were defined at drug initiation. All laboratory features and BMI were defined as the closest value to drug initiation (within a cutoff time window of –2 years to +7 days except for baseline HbA_{1c} [range aforementioned], as all features showed stability over time in this study). We also included the month of HbA_{1c} outcome measurement since drug initiation to account for potential time-dependent variation in response to each drug class. Prognostic features at drug initiation were included in the model to improve overall prediction accuracy,[19] and to control for differences between individuals receiving each drug class. These features were: the number of current, and previously prescribed, glucose-lowering drug classes, major UK ethnicity groups (White, Black, south Asian, Mixed background, or Other), Index of Multiple Deprivation 2015[12] quintile, and smoking status (non-smoker, active smoker, former smoker, or not recorded). Sex and ethnicity were self-reported in all datasets.

Glycaemic effectiveness model development and validation

In the overall CPRD study population, development and validation cohorts were defined by region of England, according to National Health Service Strategic Health Authority regions, and calendar period (development cohort: drug initiations in the South Central, South East Coast, South West, London, East of England, and East Midlands regions between Jan 1, 2004, and Oct 13, 2018; validation cohort 1: drug initiations in the North East, North West, Yorkshire and The Humber, and West Midlands regions between Jan 1, 2004, and Oct 13, 2018; and validation cohort 2: drug initiations in all regions between Oct 14, 2018, and Oct 14, 2020; appendix p 2). This holdback of data allowed independent validation to be done by geography and in the most recent 2 years of study data. The final cohorts for model development and validation comprised all individuals with complete baseline features and 12-month HbA_{1c} outcome data.

In the development cohort, the model was fitted as a flexible linear regression model to predict absolute 12-month HbA_{1C} (ie, glycaemic effectiveness) for each drug class. Continuous predictor features were modelled as five-knot restricted cubic splines (with knot positions set to the 5th, 27·5th, 50th, 72·5th and 95th quantiles of the distribution of each predictor),[20] with all predictive factors interacting with received drug class (specified as a five-level categorical variable). Prognostic factors were included without interactions with received drug class (model equation provided in the appendix [p 3]). To adjust for overfitting, penalised ridge regression was used and optimised for the Akaike information criterion. Standard performance metrics (optimism-adjusted R^2, calibration slope, and calibration-in-the-large) were calculated to assess performance for predicting absolute HbA_{1c} value, although these do not provide direct insight into the utility of the model to predict differential treatment effects. The final model predicted 12-month HbA_{1c} for each drug class for an individual patient, enabling identification of their optimal therapy for glycaemic effectiveness as the drug class with the lowest absolute predicted 12-month HbA_{1c} (regardless of the magnitude of the difference between predictions for different drug classes).

The five-drug class model was first evaluated in the CPRD development and validation cohorts. Unlike standard prediction models in which outcomes are directly observed, differences in 12-month HbA_{1c} by drug class were not observable at the individual level, as for an individual patient initiating one therapy, the counterfactual outcome if they had initiated an alternative therapy could not be observed.[21] Thus, direct evaluation of observed and predicted treatment effects was not

possible. Instead, to assess model accuracy, we compared differences in 12-month HbA_{1c} between drug initiations that were concordant or discordant with model-predicted optimal therapy. Firstly, we used model predictions to identify model-concordant drug initiations, defined as those for which the drug class initiated was also the model-predicted optimal therapy. We then applied 1:1 matching to identify model-discordant (initiated drug class not model-predicted optimal therapy) comparators. Matching was applied using a combination of exact matching of baseline HbA_{1c} (grouped by 5% percentiles) and sex, and nearest neighbour matching with replacement (caliper 0·05), including all predictive and prognostic factors in the matching algorithm. For each concordant drug initiation, we then estimated predicted HbA_{1c} benefit (ie, additional HbA_{1c} lowering) as the difference in predicted 12-month HbA_{1c} between the drug class received, and the drug class received in the matched comparator. Observed HbA_{1c} benefit was estimated as the unadjusted mean (with 95% CIs) of the differences in measured 12-month HbA_{1c} between each model-concordant drug initiation and the matched comparator. An illustration of this validation approach is provided in the appendix (p 3). The accuracy of predicted benefits was assessed by comparing predicted and observed benefits in the overall matched cohort, continuously using locally estimated scatterplot smoothing, and in subgroups defined by decile of predicted benefit.

We then validated the model for three-drug class pairwise comparisons in the TriMaster trial dataset. In TriMaster, participants received three drug classes (a DPP-4 inhibitor, SGLT2 inhibitor, and thiazolidinedione) in a randomised masked crossover design, meaning observed HbA_{1c} benefit could be calculated directly at the individual participant level. Model predictions systematically underestimated the mean outcome HbA_{1c} in TriMaster (likely due to higher adherence of trial participants *vs* real-world patients),[22] and so intercept-only recalibration was done to update predictions.[23] Intercept-only recalibration modified the mean HbA_{1c} outcome predictions for each drug class, but not predictions of relative differences between the drug classes in HbA_{1c} outcome due to effects of predictive features. Individual-level pairwise comparisons (SGLT2 inhibitor *vs* DPP-4 inhibitor; SGLT2 inhibitor *vs* thiazolidinedione, and thiazolidinedione *vs* DPP-4 inhibitor) were pooled to estimate mean overall predicted and mean observed HbA_{1c} benefit. The accuracy of predicted benefit was assessed by the same approach as in CPRD.

Further assessment of the accuracy of model predictions was done for two-drug class comparisons in the development and validation cohorts in CPRD, and the validation cohorts from TriMaster and the parallel-arm clinical trials. In CPRD, we evaluated the accuracy of predicted HbA_{1c} differences in the ten possible two-drug class comparisons from the five-drug model, based on our previously described validation framework for a two-drug (DPP-4 inhibitor *vs* SGLT2 inhibitor) treatment selection model.[8] To maximise the validation sample size in CPRD we pooled both validation cohorts, and relied upon covariate adjustment rather than matching to control for differences between individuals receiving each drug class. For each two-drug class comparison, we identified drug initiations of either class and subdivided the initiations into deciles based on predicted differences in 12-month HbA_{1c} on each therapy. Within each decile, we contrasted observed 12-month HbA_{1c} between initiations of each drug class, using adjusted linear regression models including all predictive and prognostic features as covariates. In TriMaster, HbA_{1c} differences were calculated directly for individual participants. In the two active comparator trials, after performing intercept-only recalibration due to underestimation of mean outcome HbA_{1c}, we used the same approach as in CPRD, minimally adjusting for month of HbA_{1c} outcome measure since drug initiation, as the data were last-observation-carried-forward. Due to the lower number of participants than in CPRD, two-drug comparisons in the trial datasets were evaluated by quintile of predicted HbA_{1c} differences rather than decile.

Model impacts

In the whole CPRD cohort, we used model predictions to assess the proportion of initiations of each drug class if treatment decisions aligned with predicted optimal therapy. These proportions were assessed overall and across subgroups defined by each predictive and prognostic feature. In sensitivity analysis, we assessed the same proportions for drug initiations with a minimally clinically relevant 3 mmol/mol or greater[24] predicted HbA_{1c} benefit with model-optimal therapy compared with other drug classes. We then assessed the proportion of the whole CPRD cohort with a 3 mmol/mol or greater HbA_{1c} benefit with one or more drug classes versus the other therapies.

We also estimated the predicted population-level HbA_{1c} benefit if all drug initiations were concordant with model-optimal therapy in the whole CPRD cohort, by calculating the difference in mean predicted 12-month HbA_{1c} if all drug initiations had been of model-optimal therapy, and mean predicted 12-month HbA_{1c} based on actual treatment initiations in practice. Population-level benefit was estimated overall, and in separate sensitivity analyses including only drug initiations since Jan 1, 2019, and only second-line therapy initiations (overall, and since Jan 1, 2019), to evaluate population benefit with the most contemporary data available, and for the standard situation in which all of the five drug classes are possible treatment options after first-line metformin.

Long-term outcomes

The cohort used to evaluate associations of model concordance with long-term outcomes comprised all

drug initiations in the CPRD study population with valid baseline data for the glycaemic effectiveness model, including those with missing 12-month HbA_{1c} outcome data. We applied the same 1:1 matching of model-concordant with model-discordant drug initiations, with additional matching for major comorbidities including cardiovascular conditions, microvascular complications of diabetes, and cholesterol and blood pressure medication use (see appendix p 4 for the full covariate set). After matching, outcomes in model-concordant and model-discordant strata were contrasted up to 5 years from drug initiation with Kaplan–Meier estimators, and unadjusted and adjusted (full covariate set) Cox proportional hazards regression, with proportional hazards assumptions visually checked and confirmed. Individuals initiating therapies were followed up by an as-treated approach from drug initiation until the earliest of outcome occurrence, treatment change (date of last prescription plus 30 days, or date of switching to, or adding, another drug class used in the study), deregistration from general practice, death, or 5 years since drug initiation. This approach allowed the inclusion of multiple non-overlapping follow-up periods for drug initiations in the same individual if they had initiated more than one drug class over the study period. In sensitivity analysis, we evaluated all outcomes in the validation cohort only, and also did sensitivity analyses specific to each endpoint, including assessment of MACE-HF and renal progression primary prevention endpoints (appendix p 4). We also evaluated the continuous association between predicted HbA_{1c} benefit (modelled using a three-knot restricted cubic spline, with the three knots positioned at the 10th, 50th, and 90th quantiles of the distribution)[20] and risk of glycaemic failure using adjusted (full covariate set) Cox regression.

Analyses were done with R (version 4.3.1). We followed TRIPOD+AI guidance and our analyses were guided by the Predictive Approaches to Treatment Effect Heterogeneity statement.[25,26]

Role of the funding source

The funder of the study had no role in study design, data collection, data analysis, data interpretation, or writing of the report.

Results

100 107 drug initiations in CPRD were included in development of the five-drug class model (table 1; appendix p 2). After applying 1:1 matching in the development cohort (n=30 178 drug initiations, n=15 089 pairs), we calculated a 5·3 mmol/mol mean predicted 12-month HbA_{1c} benefit in the model-concordant group compared with the model-discordant group, similar to the mean observed 12-month HbA_{1c} benefit (5·1 mmol/mol [95% CI 4·7–5·5]). Predictions were well calibrated with good agreement between observed and predicted benefit (figure 1A; appendix p 10).

External calibration was good in the two CPRD validation cohorts and in TriMaster (figure 1B–D, appendix p 10). In CPRD validation, for the model-concordant groups, mean observed 12-month HbA_{1c} benefit was 5·3 (95% CI 4·9–5·7) mmol/mol in the geographical validation cohort (n=24 746 drug initiations, n=12 373 matched pairs) and 5·0 (4·3–5·6) mmol/mol in the temporal validation cohort (n=9682 drug initiations, n=4841 matched pairs) compared with the model-discordant groups (figure 1B–C).

Similarly, predicted HbA_{1c} versus observed HbA_{1c} outcome differences in all pairwise comparisons of drug classes were well calibrated in CPRD validation, and in the clinical trials (figure 2, appendix pp 15–17). For predicting absolute HbA_{1c} outcome, rather than the difference in HbA_{1c} outcome between the drug classes, overall R^2 was 0·28 (95% CI 0·27–0·28) in the CPRD development cohort and calibration slopes in all datasets were accurate (for full performance statistics, see appendix p 18).

Over the entire study period, 32 305 (15·2%) drug initiations in the overall CPRD cohort (combined development and validation cohorts, n=212 166 drug initiations) were of the predicted optimal therapy (the drug class with the lowest absolute predicted 12-month HbA_{1c}) in real-world practice. This proportion increased to 17·8% (n=3986 of 22 452 initiations) when including only drug initiations since Jan 1, 2019, and was consistent when including only second-line therapy initiations overall (15·6%; n=15 598 of 99 689) and since 2019 (16·1%; n=1422 of 8811; appendix p 19). For the entire study period, if all initiations had been of model-optimal therapy instead of the therapy initiated in clinical practice, the predicted population-level HbA_{1c} benefit would have been 4·3 mmol/mol.

In the overall CPRD cohort, GLP-1R agonists were the drug class most commonly predicted to be optimal for 12-month HbA_{1c} (n=70 895 [33·4%] of 212 166 initiations), followed by SGLT2 inhibitors (n=61 405 [28·9%]), sulfonylureas (n=58 532 [27·6%]), and thiazolidinediones (21 322 [10·0%]), with DPP-4 inhibitors indicated for only 12 initiations (<0·01%; figure 3).

Clinical characteristics were associated with the drug class that was optimal for 12-month HbA_{1c} (figure 3). This relationship was most notable by sex, whereby GLP-1R agonists were the predicted optimal therapy for 58 765 (71·9%) of 81 758 drug initiations in female patients, but only for 12 130 (9·3%) of 130 408 initiations in male patients, for whom sulfonylureas (n=55 075 [42·2%]) and SGLT2 inhibitors (n=52 464 [40·2%]) were most frequently identified. Differences in optimal therapy by clinical characteristics were similar in those with a predicted benefit of at least 3 mmol/mol on predicted optimal therapy versus all other drug classes (n=50 162 [23·6% of overall initiations in the CPRD cohort]; appendix p 20). Applying this clinically relevant threshold (≥3 mmol/mol) the model identified clear optimal glucose-lowering therapies for the majority of

	DPP-4 inhibitors (n=39 011)	GLP-1R agonists (n=4275)	SGLT2 inhibitors (n=13 454)	Sulfonylureas (n=35 047)	Thiazolidinediones (n=8320)
Demographic features					
Age, years	61·2 (10·8)	57·4 (10·3)	58·0 (9·8)	59·8 (11·1)	60·1 (10·8)
Sex					
Male	24 116 (61·8%)	2374 (55·5%)	8381 (62·3%)	21 330 (60·9%)	5261 (63·2%)
Female	14 895 (38·2%)	1901 (44·5%)	5073 (37·7%)	13 717 (39·1%)	3059 (36·8%)
Duration of diabetes, years	8·5 (5·9)	8·8 (5·7)	9·0 (5·7)	5·8 (5·0)	7·6 (5·7)
Ethnicity					
White	28 370 (72·7%)	3670 (85·8%)	10 087 (75·0%)	26 735 (76·3%)	6202 (74·5%)
South Asian	6300 (16·1%)	273 (6·4%)	1964 (14·6%)	4873 (13·9%)	1327 (15·9%)
Black	2597 (6·7%)	176 (4·1%)	718 (5·3%)	1939 (5·5%)	455 (5·5%)
Mixed background	384 (1·0%)	41 (1·0%)	158 (1·2%)	334 (1·0%)	87 (1·0%)
Other	690 (1·8%)	39 (0·9%)	227 (1·7%)	535 (1·5%)	103 (1·2%)
Missing	670 (1·7%)	76 (1·8%)	300 (2·2%)	631 (1·8%)	146 (1·8%)
Index of multiple deprivation quintile					
1 (least deprived)	7729 (19·8%)	837 (19·6%)	2746 (20·4%)	7441 (21·2%)	1736 (20·9%)
2	7411 (19·0%)	786 (18·4%)	2508 (18·6%)	7067 (20·2%)	1649 (19·8%)
3	7859 (20·1%)	904 (21·1%)	2728 (20·3%)	7176 (20·5%)	1744 (21·0%)
4	9690 (24·8%)	1003 (23·5%)	3249 (24·1%)	8158 (23·3%)	2009 (24·1%)
5 (most deprived)	6322 (16·2%)	745 (17·4%)	2223 (16·5%)	5205 (14·9%)	1182 (14·2%)
Smoking status					
Non-smoker	10 516 (27·0%)	1038 (24·3%)	3605 (26·8%)	9862 (28·1%)	2360 (28·4%)
Active smoker	5981 (15·3%)	733 (17·1%)	2137 (15·9%)	5952 (17·0%)	1394 (16·8%)
Former smoker	20 389 (52·3%)	2313 (54·1%)	7267 (54·0%)	16 917 (48·3%)	3949 (47·5%)
Not recorded	2125 (5·4%)	191 (4·5%)	445 (3·3%)	2316 (6·6%)	617 (7·4%)
Baseline laboratory and vital sign measurements					
BMI (kg/m^2)	32·0 (6·3)	37·8 (7·0)	33·6 (6·6)	31·9 (6·3)	31·2 (6·2)
HbA_{1c} (mmol/mol)	72·2 (12·6)	77·9 (13·5)	75·8 (13·3)	73·0 (13·4)	73·5 (12·6)
eGFR (mL/min/1·73m^2)	87·6 (19·3)	92·8 (18·1)	95·3 (14·4)	88·1 (18·8)	85·7 (19·2)
Total cholesterol (mmol/L)	4·2 (1·0)	4·3 (1·0)	4·2 (1·0)	4·4 (1·1)	4·2 (1·0)
HDL cholesterol (mmol/L)	1·2 (0·3)	1·1 (0·3)	1·1 (0·3)	1·2 (0·3)	1·1 (0·3)
Alanine aminotransferase (IU/L)	33·2 (19·5)	37·2 (21·2)	35·9 (20·5)	35·7 (21·6)	35·6 (19·9)
Antihyperglycaemic treatment					
Number of previous antihyperglycaemic drug classes					
1	14 392 (36·9%)	514 (12·0%)	2929 (21·8%)	27 604 (78·8%)	2451 (29·5%)
2	17 826 (45·7%)	1167 (27·3%)	4162 (30·9%)	6118 (17·5%)	4758 (57·2%)
3	5844 (15·0%)	1600 (37·4%)	4079 (30·3%)	1125 (3·2%)	928 (11·2%)
≥4	949 (2·4%)	994 (23·3%)	2284 (17·0%)	200 (0·6%)	183 (2·2%)
Number of current antihyperglycaemic drug classes including the index drug initiation					
1	1648 (4·2%)	142 (3·3%)	391 (2·9%)	2861 (8·2%)	212 (2·5%)
2	20 291 (52·0%)	1528 (35·7%)	5066 (37·7%)	27 792 (79·3%)	3668 (44·1%)
3	16 489 (42·3%)	2156 (50·4%)	6331 (47·1%)	4302 (12·3%)	4244 (51·0%)
≥4	583 (1·5%)	449 (10·5%)	1666 (12·4%)	92 (0·3%)	196 (2·4%)
Year of drug start	2014 (2·7)	2015 (2·5)	2016 (1·4)	2011 (3·7)	2010 (3·5)
Time of HbA_{1c} outcome measure since drug initiation (months)	9·8 (3·4)	9·3 (3·5)	9·7 (3·4)	10·0 (3·3)	9·8 (3·4)
HbA_{1c} outcome (mmol/mol)	65·1 (16·0)	66·8 (17·4)	64·0 (14·1)	60·6 (14·9)	62·9 (15·6)
HbA_{1c} response (change from baseline; mmol/mol)	−7·1 (14·8)	−11·2 (16·7)	−11·8 (14·2)	−12·4 (15·1)	−10·7 (14·7)

Data are mean (SD) or n (%), where n represents the number of individual patients initiating each drug class over the study period. Baseline laboratory and vital sign measurements were defined as the closest recorded value to drug initiation. Clinical features of the validation cohorts are reported in the appendix (pp 5–9). DPP-4=dipeptidyl peptidase-4. eGFR=estimated glomerular filtration rate. GLP-1R=glucagon-like peptide-1 receptor. HbA_{1c}=glycated haemoglobin. IU=international units. SGLT2=sodium–glucose co-transporter-2.

Table 1: **Demographic and clinical features of the CPRD model development cohort at drug initiation, by drug class**

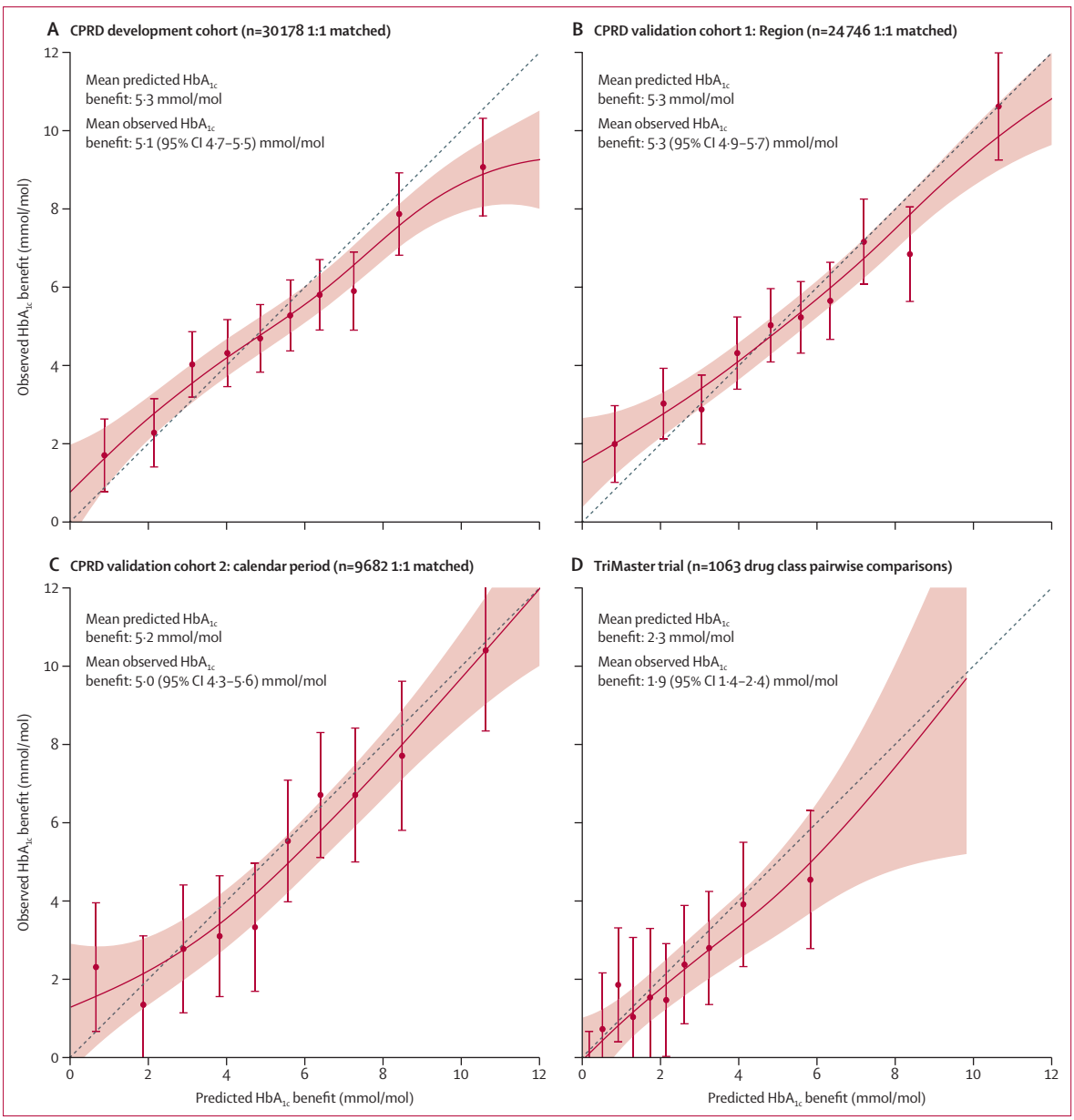

Figure 1: **Calibration of predicted HbA_{1c} benefit in the development and validation cohorts**

In parts A–C, n is the number of matched drug initiations (n/2 is the number of matched pairs). In part D, n is the number of drug class pair comparisons. Grey dashed lines represent perfect calibration. Point estimates represent predicted and observed mean benefits for subgroups defined by decile of predicted treatment benefit. Red lines represent LOESS of observed HbA_{1c} benefit versus predicted HbA_{1c} benefit. Error bars represent 95% CIs for point estimates and red shading represents 95% CIs for LOESS lines. Clinical features of CPRD matched cohorts are reported in the appendix (pp 11–14). In CPRD (A–C), observed HbA_{1c} benefits are unadjusted absolute mean differences in 12-month HbA_{1c} outcome in matched model-concordant compared with model-discordant drug initiations. In TriMaster (D), observed HbA_{1c} benefits are unadjusted absolute mean differences in 3–4-month HbA_{1c} outcome for model-concordant compared with model-discordant drug initiations in pairwise comparisons for individual participants. Mean predicted benefit (A–D) represents model-predicted HbA_{1c} outcome differences in model-concordant compared with model-discordant drug initiations. HbA_{1c}=glycated haemoglobin. LOESS=locally estimated scatterplot smoothing.

initiations, with 116 525 (54·9%) having one or two drug classes with a predicted HbA_{1c} benefit of at least 3 mmol/mol versus all other drug classes (appendix p 21). Illustrative associations between continuous predictive features and differences in 12-month HbA_{1c} by drug class are shown in the appendix (appendix pp 22–23).

In matched model-concordant and model-discordant groups (n=42 977 drug initiations per group) in CPRD (appendix pp 24–25), 5-year risk of glycaemic failure was lower in the model-concordant group of model-optimal therapy initiations (*vs* model-discordant group, adjusted hazard ratio [aHR] 0·62 [95% CI 0·59–0·64]; figure 4). In the model-concordant group, the cumulative incidence

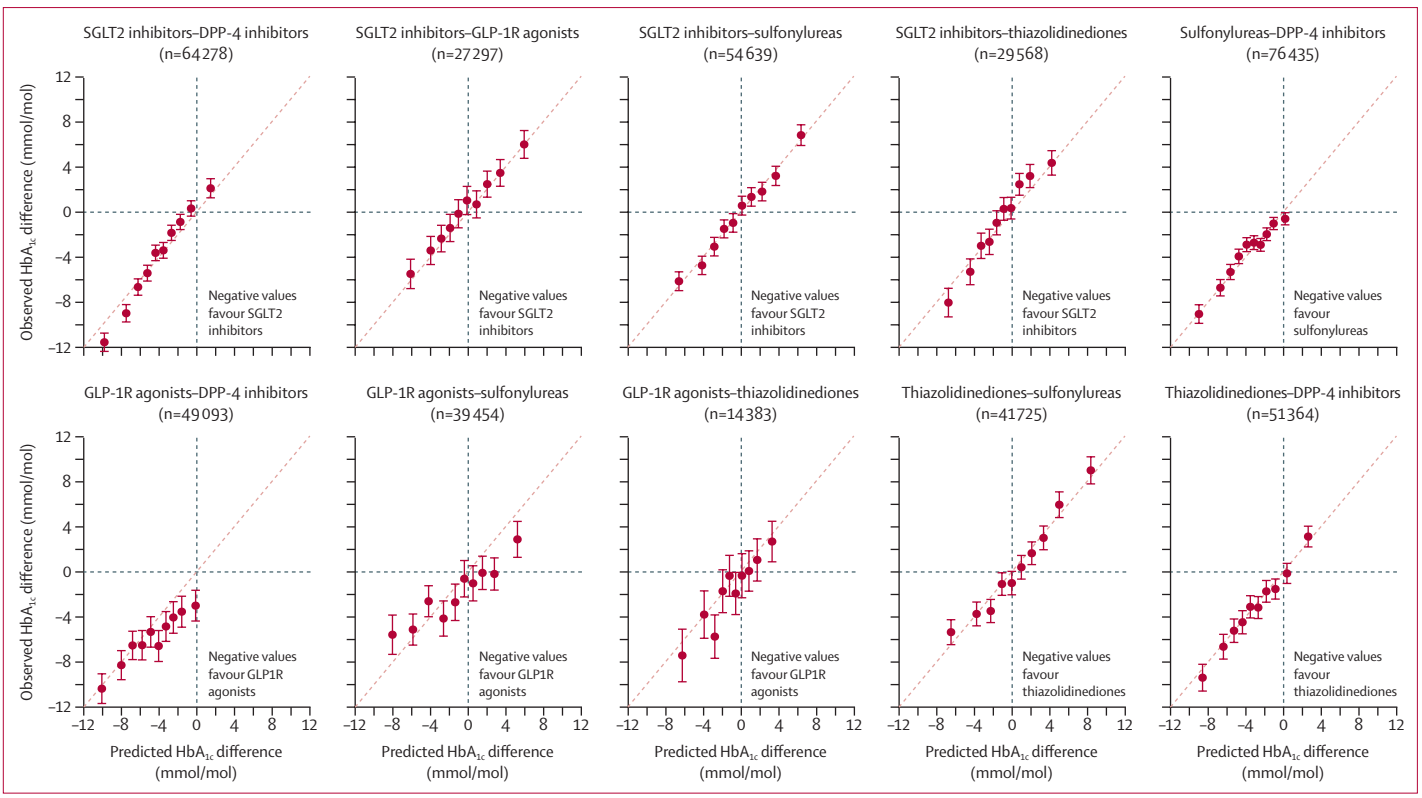

Figure 2: **Calibration of predicted heterogenous treatment effects across drug class pairs in the combined CPRD validation cohorts (n=112 059 drug initiations)**
Red dashed lines represent perfect calibration. Point estimates represent predicted and observed differences in 12-month HbA_{1c} comparing each drug class pair for subgroups defined by decile of predicted HbA_{1c} difference. Error bars represent 95% CIs. Observed HbA_{1c} differences are adjusted absolute mean differences in 12-month HbA_{1c} outcome between initiations of each drug class. n=number of drug initiations for each two-drug class analysis. DPP-4=dipeptidyl peptidase-4. GLP-1R=glucagon-like peptide-1 receptor. HbA_{1c}=glycated haemoglobin. SGLT2=sodium–glucose co-transporter-2.

of glycaemic failure was 34·7% (95% CI 33·8–35·6) at 5 years. In the model-discordant group, the same cumulative incidence (34·7% [34·0–35·4]) of glycaemic failure occurred at 2·5 years, and 5-year glycaemic failure was 46·3% (45·3–47·3). This lower risk was consistent in sensitivity analyses that used alternative definitions of glycaemic failure and when including the validation cohort only (appendix p 26). When assessed continuously, increased predicted 12-month HbA_{1c} benefits were associated with lower risks of glycaemic failure (predicted benefit 1 mmol/mol: aHR 0·84 [95% CI 0·78–0·91]; 3 mmol/mol: aHR 0·68 [0·65–0·71]; 5 mmol/mol: aHR 0·58 [0·56–0·61]; appendix p 27).

The model-concordant group, compared with the model-discordant group, had a similar 5-year risk of all-cause mortality (aHR 0·95 [95% CI 0·83–1·09]) but lower risks of MACE-HF outcomes (aHR 0·85 [0·76–0·95]), renal progression (aHR 0·71 [0·64–0·79]), and microvascular complications (aHR 0·86 [0·78–0·96]; table 2). The direction of effect for risk differences was largely consistent in sensitivity analyses, although in the model-concordant group there was no evidence of a lower risk of MACE-HF outcomes for primary prevention when excluding those with a history of cardiovascular disease, myocardial infarction, stroke, or heart failure (appendix pp 28–29). Cumulative incidence plots for each outcome are reported in the appendix (p 30).

A calculator demonstrating the five-drug class model is available online.

For the **calculator demonstrating the five-drug class model** see https://www.diabetesgenes.org/t2-treatment/

Discussion

Our study showed that a five-drug class model based on nine routinely available clinical features could accurately predict optimal glucose-lowering therapy for people with type 2 diabetes. In the CPRD cohorts from England, we found an approximate 5 mmol/mol benefit in 12-month HbA_{1c} in taking model-predicted optimal therapy. Validation in CPRD and multinational clinical trial datasets showed the model appeared to accurately predict clinically relevant differences in HbA_{1c} outcome by drug class. Model-optimal therapy was also associated with a 38% reduction in the likelihood of glycaemic failure over 5 years, which would substantially increase the time on stable glucose-lowering therapies before additional intensification is required. However, even in recent years since 2019, only 3986 (17·8%) of 22 452 therapy initiations in England were of the model-predicted optimal drug class in clinical practice. Also in England, GLP-1R agonists or SGLT2 inhibitors were the predicted optimal therapy for glycaemia for 132 300 (62·4%) of

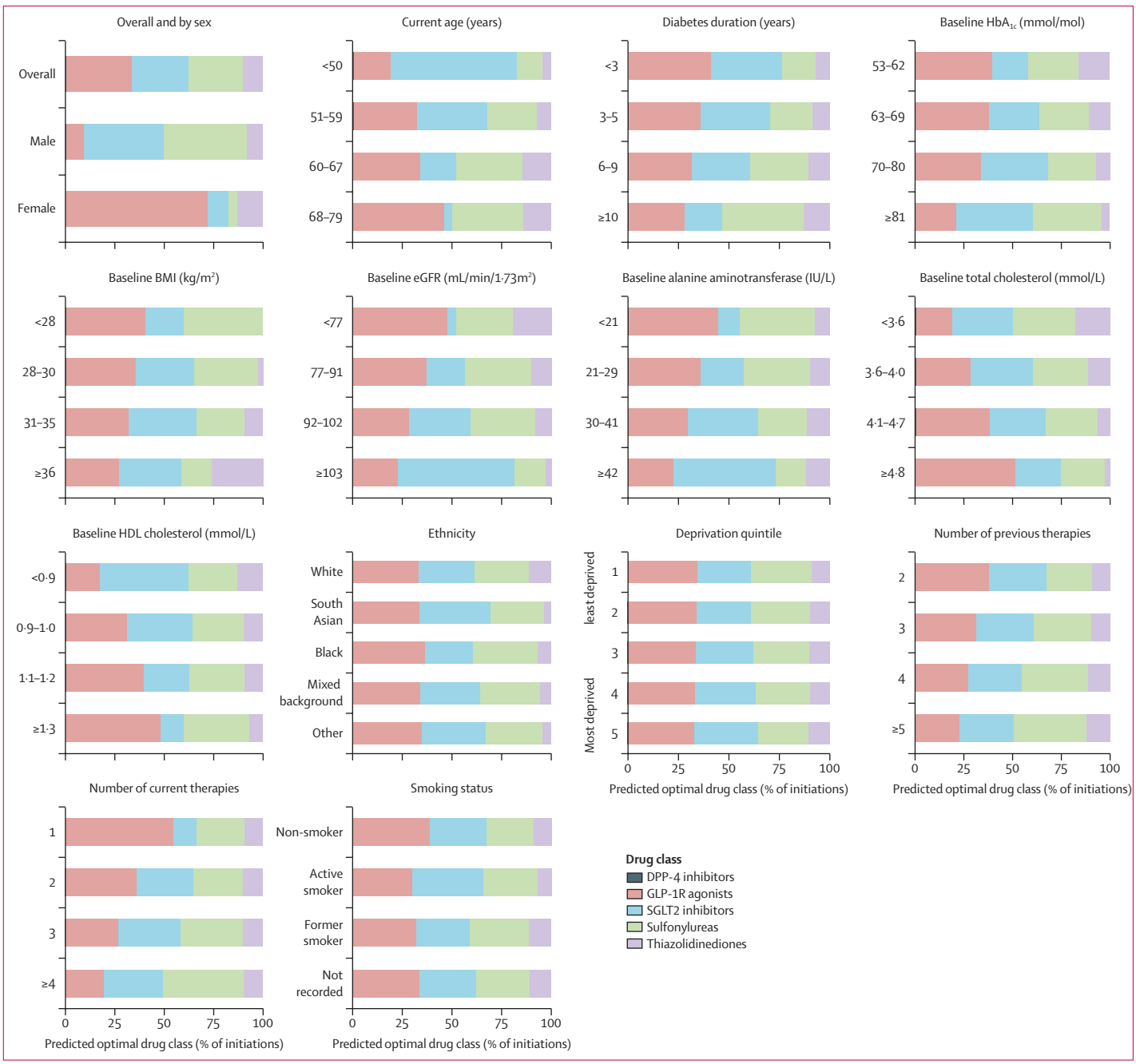

Figure 3: **Predicted optimal drug classes, overall and by clinical and demographic features at drug initiation, in the combined CPRD development and validation cohorts (n=212 166 drug initiations)**

For continuous features, subgroups are defined by quartiles of observed values. Optimal drug class proportions are derived from marginal predicted effects for drug initiations within each subgroup, with optimal therapy for a drug initiation defined as the drug class with the lowest absolute predicted 12-month HbA_{1c}. DPP-4 inhibitors were the predicted optimal therapy for 12 (<0·01%) of 212 166 drug initiations (not visible). DPP-4=dipeptidyl peptidase-4. GLP-1R=glucagon-like peptide-1 receptor. HbA_{1c}=glycated haemoglobin. IU=international units. SGLT2=sodium–glucose co-transporter-2.

212 166 therapy initiations. Our evaluation of long-term outcomes suggests that glycaemia-based treatment optimisation would associate with a lowered risk of long-term microvascular, cardiovascular, and renal endpoints. Overall, the five-drug class model represents a safe, low-cost, individualised approach to targeting specific type 2 diabetes drug classes to individuals which could benefit both glycaemia and related complications.

The heterogenous treatment effects identified in this study reflect previously described differences in both average efficacy, and mechanism of action, of the five drug classes. In keeping with the GRADE

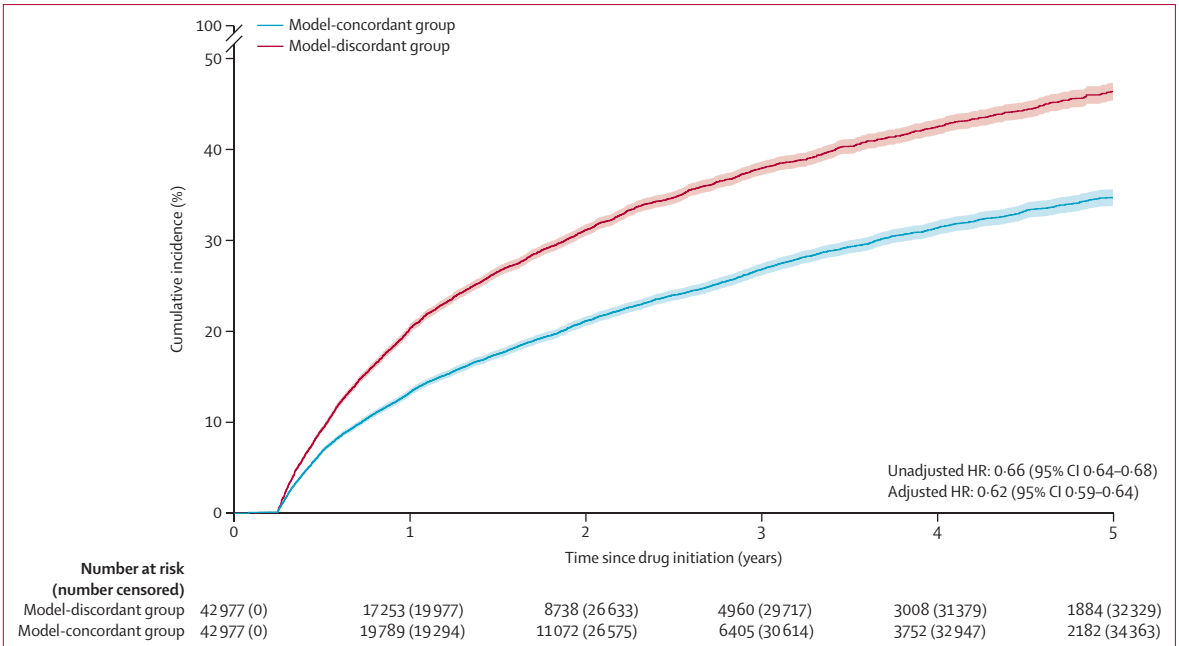

Figure 4: **5-year Kaplan–Meier curve of time to glycaemic failure (confirmed HbA_{1c} ≥69 mmol/mol) in matched model-concordant and model-discordant groups in CPRD (n=42 977 drug initiations per group)**
Shading represents 95% CIs. HbA_{1c}=glycated haemoglobin. HR=hazard ratio. n=drug initiations.

randomised controlled trial,[27] GLP-1R agonists had the greatest glycaemic efficacy for the majority of individuals, although our data suggested that this benefit was largely restricted to female patients. A greater efficacy of GLP-1R agonists in female versus male patients is supported by secondary analysis of clinical trial data for liraglutide,[9,28] and might relate to higher circulating drug concentrations in female patients at a given dose.[29] DPP-4 inhibitors were identified as optimal therapy for less than 0·01% of therapy initiations in our English cohort, in accordance with their lower efficacy in GRADE and moderate efficacy status in current international guidance.[4,27] As found in the present study and in the TriMaster trial, DPP-4 inhibitors appeared less effective in individuals with higher BMI (appendix pp 22–23).[10] Considering SGLT2 inhibitors, our finding of increased relative glycaemic benefit at high baseline HbA_{1c} and eGFR (figure 3) corresponds with previous secondary analyses of clinical trials comparing SGLT2 inhibitors with placebo and DPP-4 inhibitors,[6,8] and likely reflects increased urinary glucose excretion with high blood glucose and kidney function. The higher glycaemic effectiveness of thiazolidinediones than sulfonylureas in female patients and individuals with higher BMI was first identified in reanalysis of the ADOPT randomised controlled trial,[7] and likely relates to the increased total number of subcutaneous adipocytes, the site of thiazolidinedione action,[30] in these groups.

The optimisation of treatment for glycaemic response was associated with reduced 5-year risks of MACE-HF outcomes (15% lower risk), renal progression (29% lower risk), and microvascular complications (14% lower risk). The estimated cardiorenal protection is mainly likely to be reflective of GLP-1R agonists and

	Number of drug initiations	Number of events	Person-years at-risk	Unadjusted HR (95% CI), model-concordant *vs* model-discordant	Adjusted HR (95% CI)*, model-concordant *vs* model-discordant
All-cause mortality					
Model-discordant	42 977	443	69 976	..	..
Model-concordant	42 977	448	76 880	0·91 (0·80–1·03)	0·95 (0·83–1·09)
MACE-HF†					
Model-discordant	42 977	708	69 369	..	..
Model-concordant	42 977	621	76 291	0·79 (0·71–0·88)	0·85 (0·76–0·95)
Renal progression‡					
Model-discordant	42 977	841	68 917	..	..
Model-concordant	42 977	610	76 073	0·64 (0·57–0·70)	0·71 (0·64–0·79)
Microvascular complications§					
Model-discordant	42 977	896	68 670	..	..
Model-concordant	42 977	747	75 755	0·76 (0·69–0·84)	0·86 (0·78–0·96)

HR=hazard ratio. MACE-HF=major adverse cardiovascular events or heart failure outcomes. *Adjusted for full covariate set (appendix p 4). †MACE-HF defined as hospital admission for myocardial infarction, stroke, or heart failure as the primary reason, or death from cardiovascular disease or heart failure as the primary cause. ‡>40% estimated glomerular filtration rate decrease or end-stage kidney disease. §Earliest of progression to clinically significant albuminuria (urinary albumin-to-creatinine ratio >30 mg/g) or severe retinopathy (secondary care admission for vitreous haemorrhage or laser photocoagulation); exclusion of drug initiations in individuals with a history of both outcomes of interest was planned for the composite outcome; however, there were no such exclusions.

Table 2: **5-year secondary outcomes in model-concordant and model-discordant matched groups in CPRD**

SGLT2 inhibitors being predicted as the optimal therapies for glucose lowering in 62·4% of drug initiations, given that these therapies have well established glycaemia-independent cardiorenal benefits.[31,32] Although we did not find an association between glycaemic optimisation and reduced all-cause mortality, this was not unanticipated as previous studies have shown that the benefits of glycaemic reduction are longer-term than 5 years.[1] Future analysis, when additional follow-up of newer GLP-1R agonists and SGLT2 inhibitors is available, will help to clarify the long-term effects of model-based glycaemic optimisation on mortality.

Our model represents a pragmatic and minimal cost alternative to recently proposed precision medicine approaches focused on subclassifying type 2 diabetes based on phenotypic or genetic heterogeneity.[33–35] Our model is aimed at directly informing clinical decisions on treatment, and unlike the alternative approaches, is not aimed at advancing aetiological understanding of diabetes.[33] Direct outcome modelling has the advantage of not requiring assignment of an individual to a type 2 diabetes subtype, which, if based on clinical features, might not be stable at the individual-level over time. Outputs are also generated at the individual patient level rather than subtype level, based on contemporary clinical information available at the time a decision to intensify treatment needs to be made, likely increasing accuracy.[36]

The availability of several independent datasets, including from two multinational clinical trials and a UK-based trial, meant we were able to validate the accuracy of predicted heterogenous treatment effects on outcome HbA_{1c} at high granularity. Validation suggested that the model had both a clinically relevant overall benefit if used to target antihyperglycaemic therapies, and was well calibrated when predicting absolute differences in HbA_{1c} outcome between each of the five drug classes.

We recognise limitations of our work. Although we validated the model in clinical trial data for DPP-4 inhibitors, SGLT2 inhibitors, sulfonylureas, and thiazolidinediones, we were unable to access active comparator trial data for GLP-1R agonists to perform validation. We were also not able to include the GLP-1R agonist semaglutide (not initiated in sufficient numbers in our English cohort), nor the dual glucose-dependent insulinotropic peptide and GLP-1R agonist tirzepatide (not licensed or prescribed in the UK during our study period). Our priority is to include new agents in future versions of the model, including next-generation incretin-based therapies, as soon as there is sufficient individual-level data available for model development and validation. Based on HbA_{1c} changes estimated in a network meta-analysis,[15] we anticipate model estimates of GLP-1R agonist response will be accurate for oral semaglutide, as this has similar average efficacy to the GLP-1R agonists we did include. However, the model will likely need recalibration before incorporating subcutaneous semaglutide and tirzepatide as highly potent glucose-lowering therapies.[37,38] Other limitations of our analysis include the potential for bias or misclassification in our CPRD analysis due to the use of routine clinical data to define predictors and study outcomes, missing data, and absence of data on medication adherence. Our clinical trial validation provides reassurance that, despite these limitations, the identified heterogenous treatment effects are likely to be robust. Ongoing model development will allow for flexible imputation of missing predictor data when needed for a particular setting or individual,[39] and future research could explore optimal approaches for variable selection and the added value of non-routine clinical data including genetic variants.[40] A further limitation is the exclusion from the model of paediatric patients and individuals aged 80 years or older, groups for whom real-world data on treatment outcomes were limited and trial data were unavailable or too scarce to conduct independent validation.

The use of only nine routinely collected clinical features as treatment effect modifiers supports potential low-cost implementation of our model worldwide for people with type 2 diabetes. The model can be used to support evidence-based decisions on the likely optimal treatment to achieve an individual's target HbA_{1c}, whether as second-line therapy after initial metformin or subsequently when further glucose control intensification is needed. It is important to emphasise that type 2 diabetes treatment decisions are multifactorial, and model predictions should always be considered alongside other information including non-glycaemic cardiorenal benefits, the risk of side-effects of a drug, and availability and cost of medication. Current guideline recommendations on the initiation of SGLT2 inhibitors or GLP-1R agonists for individuals with or at high risk of cardiorenal disease are important,[4] and these guidelines might lead to a clinician prioritising these glucose-independent benefits over the optimal medication for glycaemia, at least for the initial post-metformin drug choice. An advantage of the model is that it can easily be adapted to both a particular setting and individual to maximise clinical utility. This local adaptation can include alignment with treatment guideline recommendations on drug-specific cardiorenal benefits, and omission of treatments that are not available due to local prescribing rules, cost, or supply issues. Individual adaptations can include omitting contraindicated drug classes and those that have been previously poorly tolerated. In future research, our methodological framework could be applied to evaluate heterogeneity in other clinical outcomes including cardiorenal events, side-effects, and weight change, allowing model extensions beyond glycaemia, to comprehensively support personalised diabetes treatment decisions.

Before implementation in a specific population or setting, we recommend that model performance is carefully assessed, with a particular focus on the average glycaemic effectiveness of each drug class, which might be population-specific. This is highlighted by our validation in all three clinical trials, in which model recalibration was necessary to account for greater observed than predicted glycaemic responses with all drug classes, an effect that has been previously reported to predominantly reflect increased medication adherence in clinical trial participants.[22] Such recalibration is standard when deploying prediction models in new settings. Ultimately, a clinical trial would further support implementation, although to date trials have rarely been done before deployment of clinical prediction models in practice.

In conclusion, we have developed and validated an individualised treatment selection model for people with type 2 diabetes, based on glycaemic effectiveness and considering five major glucose-lowering drug classes. The model is based solely on routinely collected clinical parameters, supporting low-cost application worldwide. Implementation could lead to improvements in glycaemic control, a marked increase in time on stable glucose-lowering therapy before additional intensification, and a potential reduction in diabetes complications.

Contributors

The study concept and design were conceived and developed by JMD, ERP, AGJ, BMS, and ATH. JMD, KGY, PC, LMG, and BMS had access to all the raw datasets used for the study. JMD directly accessed the Clinical Practice Research Datalink (CPRD) and trial datasets and undertook the analysis, with support from KGY, PC, LMG, APM, TJM, and BMS. JMD, KGY, and PC accessed and verified the underlying data used in the study. PC developed the five-drug class model online calculator. All authors had full access to the complete results of the analysis of the CPRD and trial datasets. Per university-specific data access agreements, authors at the University of Exeter (JMD, KGY, PC, LMG, APM, and BMS) had access to individual-level patient data. All authors provided support for the analysis and interpretation of results, critically revised the Article, and saw and approved the final Article. JMD and ATH attest that all listed authors meet authorship criteria and that no others meeting the criteria have been omitted. All authors had final responsibility for the decision to submit for publication.

Declaration of interests

JMD is supported by a Wellcome Trust Early-Career Award (227070/Z/23/Z). KGY and TJM were previously supported by Research England's Expanding Excellence in England (E3) fund. APM declares previous research funding from Eli Lilly, Pfizer, and AstraZeneca. ERP declares personal fees from Illumina, Eli Lilly, and Novo Nordisk. RRH is an Emeritus National Institute for Health and Care Research (NIHR) Senior Investigator, and reports personal fees from AstraZeneca, Eli Lilly, and Novartis. NS declares personal fees from Abbott Laboratories, AbbVie, Amgen, AstraZeneca, Boehringer Ingelheim, Eli Lilly, Hanmi Pharmaceuticals, Janssen, Menarini-Ricerche, Novartis, Novo Nordisk, Pfizer, Roche Diagnostics, and Sanofi, and grants to his university from AstraZeneca, Boehringer Ingelheim, Novartis, and Roche Diagnostics. ATH and BMS are supported by the NIHR Exeter Clinical Research Facility and AF is supported by the NIHR Oxford Biomedical Research Centre. AGJ was previously supported by an NIHR Clinician Scientist Fellowship (CS-2015-15-018) and declares research funding to his university from the UK Medical Research Council, Diabetes UK, Breakthrough T1D (formerly JDRF), the European Foundation for the Study of Diabetes, and the Novo Nordisk Foundation. PC and LMG declare no competing interests. Representatives from GSK, Takeda, Janssen, Quintiles, AstraZeneca, and Sanofi attended meetings as part of the industry group involved with the MASTERMIND Consortium. No industry representatives were involved in the writing of the manuscript or analysis of data. For all authors, all aforementioned declarations are outside the submitted work; all authors declare that there are no other relationships or activities that could appear to have influenced the submitted work.

Data sharing

Routine clinical data used for model development was from CPRD. All the CPRD data are available by application to the CPRD Independent Scientific Advisory Committee (https://cprd.com/data-access). The NCT00622284 and NCT01167881 clinical trial data are accessible via application to Vivli (https://vivli.org/faq/how-do-i-access-data-available-in-the-vivli-platform/), and the TriMaster clinical trial data are accessible via application to the Peninsula Research Bank (https://exetercrfnihr.org/about/exeter-10000-prb/). Code to develop the CPRD cohorts used in the study is available at: https://github.com/Exeter-Diabetes/CPRD-Cohort-scripts/tree/main/03-Treatment-response-(MASTERMIND).

Acknowledgments

This research was supported the UK Medical Research Council (grant number MR/N00633X/1). This Article is based partly on data from the CPRD obtained under licence from the UK Medicines and Healthcare products Regulatory Agency. CPRD data are provided by patients and collected by the NHS as part of their care and support. This publication is also based on research data from data contributor Boehringer Ingelheim that has been made available through Vivli. Vivli has not contributed to or approved, and is not in any way responsible for, the contents of this publication. Rhian Hopkins and Robert Challen (both of the University of Exeter) supported processing of the CPRD cohort data used in the study. The authors acknowledge support from Diabetes UK for related work developing diabetes research cohorts in CPRD, and the NIHR Exeter Biomedical Research Centre. The views expressed are those of the authors and not necessarily those of the NIHR or the Department of Health and Social Care.

References

1 Adler AI, Coleman RL, Leal J, Whiteley WN, Clarke P, Holman RR. Post-trial monitoring of a randomised controlled trial of intensive glycaemic control in type 2 diabetes extended from 10 years to 24 years (UKPDS 91). *Lancet* 2024; **404:** 145–55.

2 UK Prospective Diabetes Study (UKPDS) Group. Intensive blood-glucose control with sulphonylureas or insulin compared with conventional treatment and risk of complications in patients with type 2 diabetes (UKPDS 33). *Lancet* 1998; **352:** 837–53.

3 Tobias DK, Merino J, Ahmad A, et al. Second international consensus report on gaps and opportunities for the clinical translation of precision diabetes medicine. *Nat Med* 2023; **29:** 2438–57.

4 Davies MJ, Aroda VR, Collins BS, et al. Management of hyperglycemia in type 2 diabetes, 2022. A consensus report by the American Diabetes Association (ADA) and the European Association for the Study of Diabetes (EASD). *Diabetes Care* 2022; **45:** 2753–86.

5 Dennis JM. Precision medicine in type 2 diabetes: using individualized prediction models to optimize selection of treatment. *Diabetes* 2020; **69:** 2075–85.

6 Cherney DZI, Cooper ME, Tikkanen I, et al. Pooled analysis of phase III trials indicate contrasting influences of renal function on blood pressure, body weight, and HbA_{1c} reductions with empagliflozin. *Kidney Int* 2018; **93:** 231–44.

7 Dennis JM, Henley WE, Weedon MN, et al. Sex and BMI alter the benefits and risks of sulfonylureas and thiazolidinediones in type 2 diabetes: a framework for evaluating stratification using routine clinical and individual trial data. *Diabetes Care* 2018; **41:** 1844–53.

8 Dennis JM, Young KG, McGovern AP, et al. Development of a treatment selection algorithm for SGLT2 and DPP-4 inhibitor therapies in people with type 2 diabetes: a retrospective cohort study. *Lancet Digit Health* 2022; **4:** e873–83.

9 Cardoso P, Young KG, Nair ATN, et al. Phenotype-based targeted treatment of SGLT2 inhibitors and GLP-1 receptor agonists in type 2 diabetes. *Diabetologia* 2024; **67:** 822–36.

10 Shields BM, Dennis JM, Angwin CD, et al. Patient stratification for determining optimal second-line and third-line therapy for type 2 diabetes: the TriMaster study. *Nat Med* 2023; **29:** 376–83.

11 Wolf A, Dedman D, Campbell J, et al. Data resource profile: Clinical Practice Research Datalink (CPRD) Aurum. *Int J Epidemiol* 2019; **48:** 1740–40g.

12 Ministry of Housing, Communities and Local Government. English indices of deprivation 2015. Sept 30, 2015. https://www.gov.uk/government/statistics/english-indices-of-deprivation-2015 (accessed Dec 14, 2024).

13 Gallwitz B, Rosenstock J, Rauch T, et al. 2-year efficacy and safety of linagliptin compared with glimepiride in patients with type 2 diabetes inadequately controlled on metformin: a randomised, double-blind, non-inferiority trial. *Lancet* 2012; **380:** 475–83.

14 Ridderstråle M, Andersen KR, Zeller C, Kim G, Woerle HJ, Broedl UC. Comparison of empagliflozin and glimepiride as add-on to metformin in patients with type 2 diabetes: a 104-week randomised, active-controlled, double-blind, phase 3 trial. *Lancet Diabetes Endocrinol* 2014; **2:** 691–700.

15 Tsapas A, Avgerinos I, Karagiannis T, et al. Comparative effectiveness of glucose-lowering drugs for type 2 diabetes: a systematic review and network meta-analysis. *Ann Intern Med* 2020; **173:** 278–86.

16 Rodgers LR, Weedon MN, Henley WE, et al. Cohort profile for the MASTERMIND study: using the Clinical Practice Research Datalink (CPRD) to investigate stratification of response to treatment in patients with type 2 diabetes. *BMJ Open* 2017; **7:** e017989.

17 Levey AS, Inker LA, Matsushita K, et al. GFR decline as an end point for clinical trials in CKD: a scientific workshop sponsored by the National Kidney Foundation and the US Food and Drug Administration. *Am J Kidney Dis* 2014; **64:** 821–35.

18 Inker LA, Eneanya ND, Coresh J, et al. New creatinine- and cystatin C-based equations to estimate GFR without race. *N Engl J Med* 2021; **385:** 1737–49.

19 Lipkovich I, Svensson D, Ratitch B, Dmitrienko A. Modern approaches for evaluating treatment effect heterogeneity from clinical trials and observational data. *Stat Med* 2024; **43:** 4388–436.

20 Harrell FE Jr. Regression modeling strategies: with applications to linear models, logistic regression, and survival analysis. Springer, 2001.

21 Holland PW. Statistics and causal inference. *J Am Stat Assoc* 1985; **81:** 945–60.

22 Carls GS, Tuttle E, Tan RD, et al. Understanding the gap between efficacy in randomized controlled trials and effectiveness in real-world use of GLP-1 RA and DPP-4 therapies in patients with type 2 diabetes. *Diabetes Care* 2017; **40:** 1469–78.

23 Steyerberg EW, Vergouwe Y. Towards better clinical prediction models: seven steps for development and an ABCD for validation. *Eur Heart J* 2014; **35:** 1925–31.

24 European Medicines Agency. Guideline on clinical investigation of medicinal products in the treatment or prevention of diabetes mellitus. June 22, 2023. https://www.ema.europa.eu/en/documents/scientific-guideline/guideline-clinical-investigation-medicinal-products-treatment-or-prevention-diabetes-mellitus-revision-2_en.pdf (accessed Sept 3, 2024).

25 Kent DM, Paulus JK, van Klaveren D, et al. The Predictive Approaches to Treatment effect Heterogeneity (PATH) statement. *Ann Intern Med* 2020; **172:** 35–45.

26 Collins GS, Moons KGM, Dhiman P, et al. TRIPOD+AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods. *BMJ* 2024; **385:** e078378.

27 Nathan DM, Lachin JM, Balasubramanyam A, et al. Glycemia reduction in type 2 diabetes—glycemic outcomes. *N Engl J Med* 2022; **387:** 1063–74.

28 Garvey WT, Cohen RM, Butera NM, et al. Association of baseline factors with glycemic outcomes in GRADE: a comparative effectiveness randomized clinical trial. *Diabetes Care* 2024; **47:** 562–70.

29 Overgaard RV, Hertz CL, Ingwersen SH, Navarria A, Drucker DJ. Levels of circulating semaglutide determine reductions in HbA_{1c} and body weight in people with type 2 diabetes. *Cell Rep Med* 2021; **2:** 100387.

30 Spiegelman BM. PPAR-gamma: adipogenic regulator and thiazolidinedione receptor. *Diabetes* 1998; **47:** 507–14.

31 McGuire DK, Shih WJ, Cosentino F, et al. Association of SGLT2 inhibitors with cardiovascular and kidney outcomes in patients with type 2 diabetes: a meta-analysis. *JAMA Cardiol* 2021; **6:** 148–58.

32 Sattar N, Lee MMY, Kristensen SL, et al. Cardiovascular, mortality, and kidney outcomes with GLP-1 receptor agonists in patients with type 2 diabetes: a systematic review and meta-analysis of randomised trials. *Lancet Diabetes Endocrinol* 2021; **9:** 653–62.

33 Misra S, Wagner R, Ozkan B, et al. Precision subclassification of type 2 diabetes: a systematic review. *Commun Med (Lond)* 2023; **3:** 138.

34 Smith K, Deutsch AJ, McGrail C, et al. Multi-ancestry polygenic mechanisms of type 2 diabetes. *Nat Med* 2024; **30:** 1065–74.

35 Nair ATN, Wesolowska-Andersen A, Brorsson C, et al. Heterogeneity in phenotype, disease progression and drug response in type 2 diabetes. *Nat Med* 2022; **28:** 982–88.

36 van Smeden M, Harrell FE Jr, Dahly DL. Novel diabetes subgroups. *Lancet Diabetes Endocrinol* 2018; **6:** 439–40.

37 Sorli C, Harashima SI, Tsoukas GM, et al. Efficacy and safety of once-weekly semaglutide monotherapy versus placebo in patients with type 2 diabetes (SUSTAIN 1): a double-blind, randomised, placebo-controlled, parallel-group, multinational, multicentre phase 3a trial. *Lancet Diabetes Endocrinol* 2017; **5:** 251–60.

38 Rosenstock J, Wysham C, Frías JP, et al. Efficacy and safety of a novel dual GIP and GLP-1 receptor agonist tirzepatide in patients with type 2 diabetes (SURPASS-1): a double-blind, randomised, phase 3 trial. *Lancet* 2021; **398:** 143–55.

39 Cardoso P, Dennis JM, Bowden J, Shields BM, McKinley TJ. Dirichlet process mixture models to impute missing predictor data in counterfactual prediction models: an application to predict optimal type 2 diabetes therapy. *BMC Med Inform Decis Mak* 2024; **24:** 12.

40 Dawed AY, Mari A, Brown A, et al. Pharmacogenomics of GLP-1 receptor agonists: a genome-wide analysis of observational data and large randomised controlled trials. *Lancet Diabetes Endocrinol* 2023; **11:** 33–41.