Development of a treatment selection algorithm for SGLT2 and DPP-4 inhibitor therapies in people with type 2 diabetes

Dennis, John M.; Young, Katherine G.; McGovern, Andrew P.; Mateen, Bilal A.; Vollmer, Sebastian J.; Simpson, Michael D.

Published in:
The Lancet Digital Health

DOI:
10.1016/S2589-7500(22)00174-1

Publication date:
2022

Development of a treatment selection algorithm for SGLT2 and DPP-4 inhibitor therapies in people with type 2 diabetes: a retrospective cohort study

John M Dennis, Katherine G Young, Andrew P McGovern, Bilal A Mateen, Sebastian J Vollmer, Michael D Simpson, William E Henley, Rury R Holman, Naveed Sattar, Ewan R Pearson, Andrew T Hattersley, Angus G Jones, Beverley M Shields*, on behalf of the MASTERMIND consortium†*

Summary

Background Current treatment guidelines do not provide recommendations to support the selection of treatment for most people with type 2 diabetes. We aimed to develop and validate an algorithm to allow selection of optimal treatment based on glycaemic response, weight change, and tolerability outcomes when choosing between SGLT2 inhibitor or DPP-4 inhibitor therapies.

Methods In this retrospective cohort study, we identified patients initiating SGLT2 and DPP-4 inhibitor therapies after Jan 1, 2013, from the UK Clinical Practice Research Datalink (CPRD). We excluded those who received SGLT2 or DPP-4 inhibitors as first-line treatment or insulin at the same time, had estimated glomerular filtration rate (eGFR) of less than 45 mL/min per 1·73 m², or did not have a valid baseline glycated haemoglobin (HbA_{1c}) measure (<53 or ≥120 mmol/mol). The primary efficacy outcome was the HbA_{1c} value reached 6 months after drug initiation, adjusted for baseline HbA_{1c}. Clinical features associated with differential HbA_{1c} outcome on the two therapies were identified in CPRD (n=26 877), and replicated in reanalysis of 14 clinical trials (n=10 414). An algorithm to predict individual-level differential HbA_{1c} outcome on the two therapies was developed in CPRD (derivation; n=14 069) and validated in head-to-head trials (n=2499) and CPRD (independent validation; n=9376). In CPRD, we further explored heterogeneity in 6-month weight change and treatment discontinuation.

Findings Among 10 253 patients initiating SGLT2 inhibitors and 16 624 patients initiating DPP-4 inhibitors in CPRD, baseline HbA_{1c}, age, BMI, eGFR, and alanine aminotransferase were associated with differential HbA_{1c} outcome with SGLT2 inhibitor and DPP-4 inhibitor therapies. The median age of participants was 62·0 years (IQR 55·0–70·0). 10 016 (37·3%) were women and 16 861 (62·7%) were men. An algorithm based on these five features identified a subgroup, representing around four in ten CPRD patients, with a 5 mmol/mol or greater observed benefit with SGLT2 inhibitors in all validation cohorts (CPRD 8·8 mmol/mol [95% CI 7·8–9·8]; CANTATA-D and CANTATA-D2 trials 5·8 mmol/mol [3·9–7·7]; BI1245.20 trial 6·6 mmol/mol [2·2–11·0]). In CPRD, predicted differential HbA_{1c} response with SGLT2 inhibitor and DPP-4 inhibitor therapies was not associated with weight change. Overall treatment discontinuation within 6 months was similar in patients predicted to have an HbA_{1c} benefit with SGLT2 inhibitors over DPP-4 inhibitors (median 15·2% [13·2–20·3] *vs* 14·4% [12·9–16·7]). A smaller subgroup predicted to have greater HbA_{1c} reduction with DPP-4 inhibitors were twice as likely to discontinue SGLT2 inhibitors than DPP-4 inhibitors (median 26·8% [23·4–31·0] *vs* 14·8% [12·9–16·8]).

Interpretation A validated treatment selection algorithm for SGLT2 inhibitor and DPP-4 inhibitor therapies can support decisions on optimal treatment for people with type 2 diabetes.

Funding BHF-Turing Cardiovascular Data Science Award and the UK Medical Research Council.

Introduction

SGLT2 and DPP-4 inhibitors are recommended glucose-lowering treatment options for people with type 2 diabetes after metformin,[1] representing around 60% of second-line treatment initiations in the UK[2] and 27% in the US.[3] Trial data suggest that the average glucose-lowering efficacy of both therapies is similar, although SGLT2 inhibitors are associated with weight loss.[4] Differences in tolerability have not been evaluated in a large number of patients in routine practice. Although the American Diabetes Association and the European Association for the Study of Diabetes (ADA/EASD) guidelines recommend SGLT2 inhibitors or GLP-1 receptor agonists, or both, in people with established atherosclerotic cardiovascular disease, heart failure, or chronic kidney disease,[1] this stratification only applies to 15–20% of people with type 2 diabetes.[5] Thus, there is considerable uncertainty regarding optimal treatment for

Lancet Digit Health 2022; 4: e873–83

See **Comment** page e851

*Joint senior authors

†Members are listed in the appendix (p 35)

University of Exeter Medical School, Institute of Biomedical and Clinical Science, Royal Devon and Exeter Hospital, Exeter, UK (J M Dennis PhD, K G Young PhD, A P McGovern MD, Prof A T Hattersley DM, Prof A G Jones PhD, B M Shields PhD); **The Alan Turing Institute, British Library, London, UK** (B A Mateen MBBS); **Institute of Health Informatics, University College London, London, UK** (B A Mateen); **Department of Statistics, University of Warwick, Coventry, UK** (Prof S J Vollmer PhD); **Newcastle University, Newcastle upon Tyne, UK** (M D Simpson PhD); **Institute of Health Research, University of Exeter Medical School, Exeter, UK** (Prof W E Henley PhD); **Oxford Centre for Diabetes, Endocrinology, and Metabolism, University of Oxford, Churchill Hospital, Oxford, UK** (Prof R R Holman FMedSci); **Institute of Cardiovascular and Medical Sciences, University of Glasgow, Glasgow, UK** (Prof N Sattar MD); **Division of Molecular and Clinical Medicine, Ninewells Hospital and Medical School, University of Dundee, Dundee, UK** (Prof E R Pearson PhD)

Correspondence to: Dr John M Dennis, University of Exeter Medical School, Institute of Biomedical and Clinical Science, Royal Devon and Exeter Hospital, Exeter EX2 5DW, UK **j.dennis@exeter.ac.uk**

See **Online** for appendix

Research in context

Evidence before this study

Current type 2 diabetes treatment guidelines do not provide information to support targeted treatment based on differences in glycaemic outcomes. We searched PubMed and Medline for articles published from Jan 1, 2000, to June 29, 2022, using the search terms: Diabetes Mellitus[MeSH Major Topic] AND ("Precision" OR "stratification" OR "personalis*" OR "individualisation" OR "targeted" OR "effect heterogeneity" OR "individualis*" OR "prediction" OR "heterogenous treatment effects" OR "treatment selection" OR "effect modification") AND (Sodium-Glucose Transporter 2 Inhibitors [MeSH Major Topic]) OR (Dipeptidyl-Peptidase IV Inhibitors[MeSH Major Topic]) AND "HbA_{1c}". We found studies of more than 100 individuals that assessed whether patient characteristics are associated with differences in glucose-lowering response (change in glycated hemoglobin [HbA_{1c}] from baseline, with at least 3 months of follow-up) after initiating SGLT2 or DPP-4-inhibitor therapy. Previous studies have suggested that the glucose-lowering response to both therapies is heterogenous and might be associated with patient characteristics, but to the best of our knowledge, no studies have assessed whether differences in response are reproducible and of sufficient magnitude to support targeted treatment selection for individual patients.

Added value of this study

We show that routine clinical patient features are associated with clinically relevant differences in glucose-lowering response to SGLT2 and DPP-4 inhibitor therapies in observational and clinical trial data. We also developed the first treatment selection model that can provide individualised estimates of relative glucose-lowering benefit with these two therapies. Validation shows that the model can reliably identify novel strata for patients with type 2 diabetes and predict clinically relevant differences in glucose-lowering (HbA_{1c}) outcome and risk of early treatment discontinuation. People with type 2 diabetes who have the greatest predicted glycaemic benefit with SGLT2 inhibitors compared with DPP-4 inhibitors are usually not those who are preferentially recommended to receive SGLT2 inhibitors due to their cardiorenal risk status according to current treatment guidelines.

Implications of all the available evidence

Findings show that an individualised precision medicine approach based on routine clinical features can be used to support targeted SGLT2 and DPP-4 inhibitor therapy for people who are most likely to benefit. The validated treatment selection model provides individualised estimates of glycaemic response, weight change, and treatment discontinuation, for each therapy that can complement existing recommendations based on cardiorenal risk and could directly inform clinical decisions concerning optimal treatment choices for people with type 2 diabetes.

most people with type 2 diabetes after receiving metformin.

A possible precision medicine approach to treatment selection is to use individual-level patient characteristics to target specific glucose-lowering treatment for people with type 2 diabetes who are most likely to benefit.[6] Such a targeted approach might improve glucose-lowering efficacy and drug tolerability, reduce side-effects, and reduce the risk of diabetes complications when implemented. Studies[7–10] in the past 5 years have shown the potential of precision medicine for people with type 2 diabetes, identifying subgroups with different glycaemic response to specific agents, rates of glycaemic progression, and risk of complications. However, previous studies have not evaluated the clinical use of proposed precision medicine approaches and whether differential outcomes between drug classes can be predicted in a robust way and used for targeted therapy.[11]

Here, we aim to build on modelling approaches to detect patient-level treatment effects[12] and evaluate whether individual routinely collected clinical features are robustly associated with differential glycaemic response to SGLT2 and DPP-4 inhibitor therapies; whether combining these features to predict glycated haemoglobin (HbA_{1c}) responses can inform treatment selection based on optimal glucose-lowering; and whether this treatment selection relates to changes in weight and treatment discontinuation.

Methods

Study design and participants

In this retrospective cohort study, routine clinical data were used to explore differential glycaemic response to SGLT2 and DPP-4 inhibitor therapies and develop a treatment selection model, taking advantage of the large sample size and greater heterogeneity of patients in clinical practice. Replication analysis and model validation was performed by reanalysis of clinical trials with systematic follow-up of randomly assigned participants.

We identified patients initiating SGLT2 and DPP-4 inhibitor therapies after Jan 1, 2013, from the UK Clinical Practice Research Datalink (CPRD) 2013–19 GOLD after our published protocol.[13] We excluded those who received SGLT2 or DPP-4 inhibitors as first-line treatment (not recommended in UK guidelines)[14] or insulin at the same time, had eGFR of less than 45 mL/min per 1·73 m² (in which case use of SGLT2 inhibitors is usually contraindicated), or did not have a valid baseline HbA_{1c} measure (<53 or ≥120 mmol/mol; appendix p 3). To reflect underlying diabetes pathophysiology, we extracted baseline clinical features including HbA_{1c} (closest value within 6 months before to 7 days after treatment start),

current age, sex, diabetes duration, BMI, weight, eGFR (using the Chronic Kidney Disease Epidemiology Collaboration formula), HDL cholesterol, triglycerides, alanine aminotransferase, albumin, and bilirubin (closest values to treatment start in the previous 2 years). Ethnicity was extracted and used to describe the study population but was not analysed due to the limited numbers of patients of non-White ethnicity. We also identified the number of current and ever prescribed glucose-lowering treatments. To identify patients specifically recommended SGLT2 inhibitors in accordance with ADA/EASD treatment guidelines, we extracted records of atherosclerotic cardiovascular disease (myocardial infarction, stroke, peripheral artery disease, ischaemic heart disease, or revascularisation), heart failure, and chronic kidney disease (eGFR <60 mL/min per 1·3 m² or urinary albumin to creatinine ratio >30 mg/g, or both). To identify patients at higher risk of future cardiovascular disease, we applied the Systematic Coronary Risk Evaluation (SCORE) model to identify patients with a 10-year risk SCORE of 5% or greater.[15]

We also extracted individual participant data from 14 multi-country randomised clinical trials of SGLT2 and DPP-4 inhibitor therapies (n=10 414) from trial data sharing portals. These studies were three active comparator HbA_{1c} efficacy trials (CANTATA-D [NCT01106677] and CANTATA-D2 [NCT01137812] trials of canagliflozin [SGLT2 inhibitors] *vs* sitagliptin [DPP-4 inhibitors] or placebo [not analysed]; NCT01177813 trial of empagliflozin [SGLT2 inhibitors] *vs* sitagliptin), six trials of SGLT2 inhibitors versus placebo or sulfonylurea (canagliflozin [NCT00968812, NCT01081834, NCT01106651, and NCT01106625 trials] or empagliflozin [NCT01210001 and NCT0115960]; non-SGLT2 inhibitor treatment groups were not analysed), the EMPA-REG OUTCOME cardiovascular outcome trial (NCT01131676 empagliflozin *vs* placebo; placebo group not analysed, patients with insulin cotreatment excluded, only HbA_{1c} measures on unchanged glucose-lowering therapy analysed), and four efficacy trials of linagliptin (DPP-4 inhibitors) versus placebo or sulfonylurea (NCT00602472, NCT00621140, NCT00601250, and NCT00622284; non-DPP-4 inhibitor treatment groups not analysed). Participants randomly assigned to different doses of active agents were pooled for analysis. Full details of trial inclusion criteria, study cohorts, and trial references are provided in the appendix (pp 4–12). We extracted HbA_{1c} outcome and baseline assessment data for the same clinical features, as in CPRD. Approval for the study was granted by the CPRD Independent Scientific Advisory Committee (ISAC 13_177R), the YODA Project (number 2017–1816), and Vivli (ID00005959). Patient consent was not required.

Outcomes

The primary efficacy outcome was the HbA_{1c} value reached 6 months after drug initiation, adjusted for baseline HbA_{1c}.[16] In CPRD, this outcome was defined as the closest HbA_{1c} to 6 months after initiation (range 3–15 months) on unchanged therapy (no addition or cessation of other glucose-lowering medication and continued prescription of the drug of interest). In trials, all on treatment HbA_{1c} values at study visits from 3–6 months after randomisation were evaluated.

Secondary outcomes in CPRD were the weight reached 6 months after initiation (closest value within 3–15 months), adjusted for baseline weight, and treatment discontinuation within 6 months of drug initiation (a proxy of drug tolerability). Patients were required to have 3 months of follow-up time after their last prescription to confirm that the drug was discontinued.

Statistical analysis

We estimated the association between baseline HbA_{1c} and HbA_{1c} outcome by drug using linear regression with a drug-by-baseline HbA_{1c} interaction term, with baseline HbA_{1c} modelled as a 3-knot restricted cubic spline to allow for non-linearity. In this model and all subsequent CPRD models, we controlled for differences in drug order and potential adherence effects by adjusting for the number of current and ever prescribed glucose-lowering drug classes. We also adjusted for month of outcome measurement, defined as the month relative to baseline that HbA_{1c} outcome was recorded.

We then sequentially assessed associations by drug for other baseline clinical features, by adding each in turn as drug-by-feature interactions to the baseline HbA_{1c} adjusted model in complete case analysis. A complete case approach was used because missing baseline data were considered likely missing not at random, meaning imputation approaches might lack validity.[17] Each feature was standardised to allow comparison of effect size. To evaluate model fit we examined normality of residuals and linearity of associations for continuous variables. We performed a series of sensitivity analyses to evaluate consistency of the same associations. To explore potential modifying effects of background therapy, we included only patients on metformin with or without sulfonylurea (the two biggest background therapy groups). To evaluate the effect of combination therapy with SGLT2 and DPP-4 inhibitors, we excluded patients initiating SGLT2 inhibitors while receiving DPP-4 or initiating DPP-4 inhibitors while receiving SGLT2 inhibitors. To explore the effect of defining clinical features using the closest values to treatment start in the previous 2 years, we reduced this time period to 1 year.

In trials, we estimated associations for the same clinical features using mixed-effect models with patient-level random effects to allow for repeated HbA_{1c} outcome measures. Features were standardised to CPRD distributions. For active comparator trials, we estimated drug-specific effects using drug-by-feature interaction terms. Trial estimates were pooled using two-stage random effect meta-analysis.[18]

For more on the **YODA Project** see https://yoda.yale.edu/

For more on the **Vivli Center for Global Clinical Research Data** see https://vivli.org/

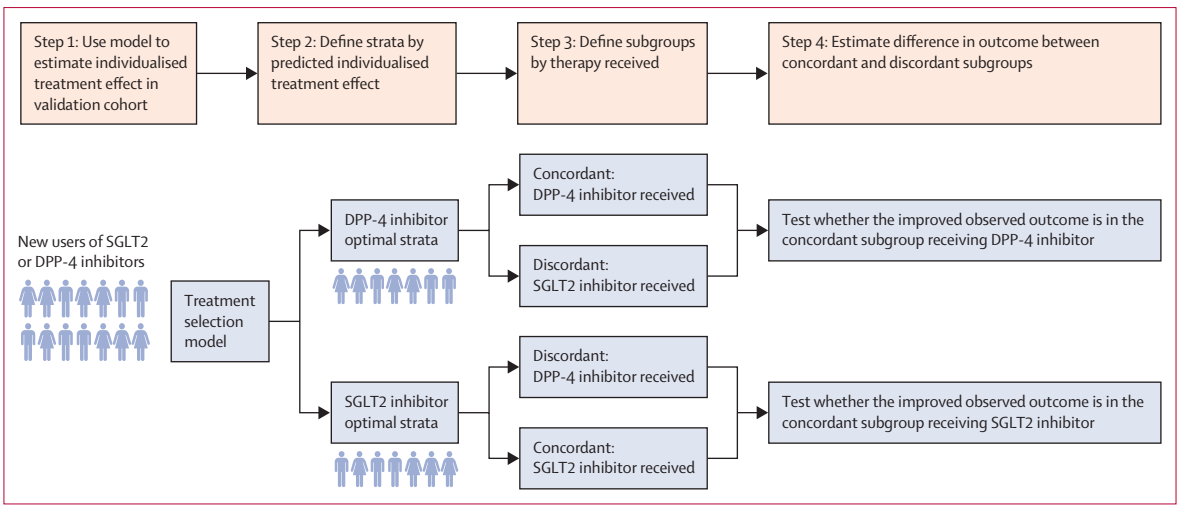

Figure 1: **SGLT2 inhibitor versus DPP-4 inhibitor treatment selection model**
The treatment selection model assigns each patient to either DPP-4 inhibitors or SGLT2 inhibitors as the predicted optimal therapy. Concordant (therapy received is the predicted optimal therapy) and discordant (therapy received is not the predicted optimal therapy) subgroups can then be defined based on the therapy actually received by each patient. To evaluate treatment selection model performance, improvement of the outcome in the concordant compared with discordant subgroups can be estimated. More granular subgroups can be defined by the size of predicted treatment effect (eg, defined by a predicted HbA_{1c} of more than 5 mmol/mol benefit with SGLT2 *vs* DPP-4 inhibitors).

In CPRD, we combined clinical features to develop a treatment selection model to predict HbA_{1c} outcome for an individual patient if they were to receive SGLT2 or DPP-4 inhibitor therapy. A multivariable linear regression model was developed in a 60% random sample (derivation cohort), with 40% of patients held back for model validation. By including treatment-by-feature interaction terms, the model facilitated prediction of the outcome on each therapy, conditional on the features included as interaction terms, and thus enabled prediction of individualised treatment effects. For each person, the difference between the predicted HbA_{1c} outcome on the two therapies provided an estimate of their individualised treatment effect.

To inform variable selection, an initial linear regression model for 6-month HbA_{1c} with interaction terms between treatment and all baseline clinical features (continuous features modelled as 3-knot restricted cubic splines) as explanatory variables was fitted, adjusting for number of current and ever prescribed glucose-lowering drug classes, and month of outcome measurement. Collinearity between clinical features was assessed to ensure variables were not highly correlated. To rank the relative importance of each baseline clinical feature for estimating individualised treatment effects, we estimated the proportion of χ^2 explained by the interaction term for each feature (which represents the differential drug-specific effect of the feature on HbA_{1c} outcome), with bootstrapped CIs. Stepwise forward selection was used to define features for the final model, by adding drug-by-feature interaction terms in order of relative importance to a base model that included all clinical features without interaction terms (ie, $p<0{\cdot}01$ threshold). Non-differential features were omitted from the final model because they explained little variation in predicting overall HbA_{1c} outcome (additional R^2 0·004). To adjust for overfitting, penalised ridge regression was used to optimise for Akaike information criterion.[19] Standard performance metrics (optimism-adjusted R^2, root mean square error, calibration slope, and calibration in the large) were estimated to assess model performance for predicting HbA_{1c} outcome directly.[20]

For treatment selection models, accurately predicting the magnitude of difference between therapies (the individualised treatment effect) is more important than accurately predicting the outcome.[21] Therefore, standard model performance metrics that test the ability of a model to predict the outcome directly are of limited use in the context of evaluating a model estimating individualised treatment effects.[21,22] The challenge for validation is that the difference in treatment effect cannot be measured directly within an individual, because for each individual the counterfactual outcome on the treatment they did not initiate cannot be observed.

Given this limitation, our approach to treatment selection model validation was to assess differences in observed HbA_{1c} outcome in patient strata defined by model-predicted individualised treatment effects. Evaluation using this approach is described in figure 1. This validation approach was applied in three active comparator trials of SGLT2 inhibitors versus DPP-4 inhibitors (a pooled analysis of trials CANTATA-D and CANTATA-D2, and separate analysis of trial BI1245.20) and the CPRD validation set. First, we used the model to predict individual-level HbA_{1c} outcomes for all individuals. Second, predictions were used to estimate individualised

	DPP-4 inhibitor (n=16 624)	SGLT2 inhibitor (n=10 253)	Standardised mean difference
Age, years	63·9 (10·8)	60·0 (9·2)	0·39
Duration of diabetes, years	8·0 (5·3)	8·5 (5·1)	0·11
Sex			
Female	6265 (37·7%)	3751 (36·6%)	0·023
Male	10359 (62·3%)	6502 (63·4%)	..
Ethnicity			
White	8089 (48·7%)	4793 (46·7%)	0·061
Asian	650 (3·9%)	354 (3·5%)	..
Black	216 (1·3%)	99 (1·0%)	..
Mixed or other	158 (1·0%)	90 (0·9%)	..
Missing	7511 (45·2%)	4917 (48·0%)	..
DPP-4 inhibitor type			
Alogliptin	1957 (11·8%)	..	..
Linagliptin	3331 (20·0%)	..	..
Saxagliptin	2207 (13·3%)	..	..
Sitagliptin	9025 (54·3%)	..	..
Vildagliptin	104 (0·6%)	..	..
SGLT2 inhibitor type			
Canagliflozin	..	1588 (15·5%)	..
Dapagliflozin	..	5861 (57·2%)	..
Empagliflozin	..	2804 (27·3%)	..
Number of glucose-lowering drug classes ever prescribed			
1	7355 (44·2%)	2198 (21·4%)	0·55
2	7024 (42·3%)	3187 (31·1%)	..
3	1901 (11·4%)	3027 (29·5%)	..
4+	344 (2·1%)	1841 (18·0%)	..
Number of other current glucose-lowering drugs			
0	947 (5·7%)	286 (2·8%)	0·84
1	9702 (58·4%)	4117 (40·2%)	..
2	5770 (34·7%)	4701 (45·8%)	..
3+	205 (1·2%)	1149 (11·2%)	..
Background therapy			
Metformin	14 667 (88·2%)	9296 (90·7%)	0·08
Sulfonylurea	6320 (38·0%)	3981 (38·8%)	0·017
DPP-4 inhibitor	..	2712 (26·5%)	..
SGLT2 inhibitor	361 (2·2%)	..	..
Thiazolidinedione	422 (2·5%)	339 (3·3%)	0·046
GLP-1 receptor agonist	56 (0·3%)	650 (6·3%)	0·34
Baseline biomarkers			
HbA_{1c}, mmol/mol†	73·0 (13·4)	76·9 (14·2)	0·29
BMI, kg/m²	32·3 (6·4)	34·4 (6·6)	0·31
eGFR, mL/min per 1·3 m²	83·1 (17·3)	88·8 (14·7)	0·36
HDL cholesterol, mmol/L	1·2 (0·3)	1·1 (0·3)	0·15
Triglycerides, mmol/L	2·2 (1·6)	2·4 (1·6)	0·078
Alanine transaminase, IU/L	33·7 (52·4)	36·6 (40·2)	0·061
Albumin, g/L	42·2 (4·0)	42·2 (4·0)	0·014
Bilirubin, µmol/L	10·0 (5·1)	9·8 (5·0)	0·040
Cardiovascular disease			
Atherosclerotic cardiovascular disease‡	3314 (19·9%)	1646 (16·1%)	0·10
Heart failure	535 (3·2%)	213 (2·1%)	0·071
Cardiovascular disease risk§			
SCORE 10-year cardiovascular disease risk, %	1·8 (2·0)	1·2 (1·3)	0·37
High or very high cardiovascular disease risk (SCORE ≥5%)	1067 (6·9%)	186 (1·9%)	0·24
Renal disease			
Chronic kidney disease¶	2257 (13·6%)	476 (4·6%)	0·31
SGLT2 inhibitor eligible due to cardiovascular disease or chronic kidney disease status	5475 (32·9%)	2184 (21·3%)	0·26
HbA_{1c} outcome			
HbA_{1c}, mmol/mol	65·1 (16·1)	65·1 (14·4)	0·0010
Month of HbA_{1c} measure	6·4 (2·7)	6·3 (2·6)	0·050

Data are mean (SD) or n (%). Clinical trial cohort characteristics are reported in the appendix (pp 4–6). eGFR=estimated glomerular filtration rate. HbA_{1c}=glycated haemoglobin. SCORE=Systematic Coronary Risk Evaluation. *Standardised mean difference of 0·1 or higher is a metric for meaningful imbalance. †Median time before baseline of HbA_{1c} measurements was 15 days (IQR 4–30 days). 4·8% of patients had a baseline HbA_{1c} more than 91 days before baseline. ‡Composite of history of myocardial infarction, stroke, ischaemic heart disease, peripheral artery disease, and revascularisation. §Systematic Coronary Risk Evaluation, excluding patients with a history of myocardial infarction (n=1902 excluded). Derived using equations from Conroy and colleagues.[15] ¶eGFR of less than 60 mL/min per 1·3 m² or urinary albumin to creatinine ratio of more than 30 mg, or both.

Table: **Baseline clinical characteristics of patients from the UK Clinical Practice Research Datalink**

treatment effects (the difference in predicted HbA_{1c} outcomes between the two therapies). Third, strata were defined by clinically defined HbA_{1c} cutoffs of predicted individualised treatment effect (SGLT2 inhibitor benefit of ≥10, 5–10, 3–5, or 0–3 mmol/mol; DPP-4 inhibitor benefit of ≥3 or 0–3 mmol/mol). Fourth, for each strata linear regression models were fitted to compare HbA_{1c} outcome in concordant (therapy received is the predicted optimal therapy for HbA_{1c}) versus discordant (therapy received is the predicted non-optimal therapy for HbA_{1c}) subgroups. In the trials, the primary outcome assessed was the last observation carried forward 6-month HbA_{1c}, with 12-month HbA_{1c} outcome evaluated as a sensitivity analysis. We also evaluated performance for treatment selection in the CPRD validation set excluding patients with established atherosclerotic cardiovascular disease,

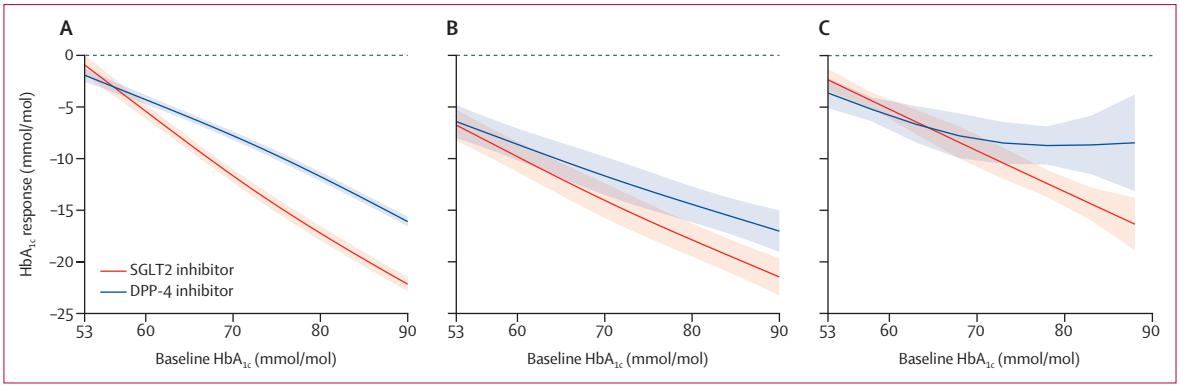

Figure 2: **Association between baseline HbA_{1c} and 6-month HbA_{1c} response with SGLT2 inhibitor versus DPP-4 inhibitor treatment**
Shading represents 95% CIs. (A) CPRD routine clinical data (n=26 877). (B) CANTATA-D (NCT01106677) and CANTATA-D2 trials (NCT01137812; n=1755). (C) BI1245.20 trial (NCT01177813; n=630). CPRD=Clinical Practice Research Datalink. HbA_{1c}=glycated haemoglobin.

heart failure, chronic kidney disease, or SCORE risk of 5% or higher.[15]

In the whole CPRD cohort, we assessed whether selecting treatment based on predicted HbA_{1c} outcome altered 6-month weight change and risk of treatment discontinuation. For both analyses, we included all patients with valid data for the treatment selection model, whether a valid HbA_{1c} outcome was recorded or not (to avoid selection bias because patients with a valid HbA_{1c} outcome measure would have remained on treatment for at least 3 months) and applied the same concordant–discordant approach used to evaluate HbA_{1c} outcome. For treatment discontinuation, a logistic regression model was fitted with predicted HbA_{1c} difference between therapies as the exposure (3-knot restricted cubic spline), adjusting for baseline HbA_{1c}, the number of current glucose-lowering medications, and the number of previously initiated medications. For weight change, linear regression was used with additional adjustment for baseline weight. Longer term 12-month weight and treatment discontinuation outcomes were evaluated in sensitivity analyses. All analyses were conducted using R (version 4.0.2). We followed TRIPOD prediction model reporting guidance.[20] A web tool including the algorithm developed for this study is available online.

For the **Type 2 Diabetes Treatment Selection Decision Aid** see https://www.diabetesgenes.org/sglt2DPP-4decisionaid/

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

Between Jan 1, 2013, and July 1, 2019, we identified 41 807 study-eligible participants initiating SGLT2 or DPP-4 inhibitors in CPRD (appendix p 3). For analysis of glycaemic response, the CPRD cohort included 10 253 participants initiating SGLT2 inhibitor therapy and 16 624 participants initiating DPP-4 inhibitor therapy with valid HbA_{1c} outcome data (appendix p 3). The median age of participants was 62·0 years (IQR 55·0–70·0). 10 016 (37·3%) were women and 16 861 (62·7%) were men. Baseline clinical characteristics by initiated drug class are reported in the table for CPRD and appendix (pp 4–6) for the 14 clinical trial cohorts (n=10 414). In CPRD, higher baseline HbA_{1c} was associated with a markedly greater HbA_{1c} reduction at 6 months with SGLT2 inhibitors compared with DPP-4 inhibitors (figure 2). This association was replicated in trials. Adjusted for baseline HbA_{1c}, multiple individual features showed evidence of differential responses in CPRD (figure 3A). Higher age and longer diabetes duration were associated with greater response to DPP-4 inhibitors, but not SGLT2 inhibitors. Higher eGFR and alanine transaminase were associated with greater response to SGLT2 inhibitors and lesser response to DPP-4 inhibitors. Higher BMI was associated with a lesser response to DPP-4 inhibitors, but not associated with SGLT2 inhibitor response. Associations were consistent in sensitivity analyses assessing patients receiving only background metformin (n=12 421), background metformin and sulfonylurea dual therapy (n=7579), when excluding patients receiving concurrent SGLT2 inhibitors or DPP-4 inhibitors from each treatment group (analysis cohort; n=23 804), and when defining baseline clinical features using a 1-year rather than 2-year time period before drug initiation (appendix pp 18–19). Differential treatment effects for BMI and alanine transaminase were replicated in trials, as was the association between higher eGFR and greater SGLT2 inhibitor response (figure 3B). Associations between higher age and lower eGFR with greater DPP-4 inhibitor response were not replicated.

In the CPRD model derivation cohort (n=16 126 for baseline characteristics in derivation and validation cohorts; appendix pp 20–21), when including all clinical features (n=12 034 with valid data for all features), the features with the highest explained variation for predicting differential treatment effects were baseline HbA_{1c} and eGFR, followed by alanine transaminase, BMI, and current age (appendix p 22). Only these five differential features significantly improved prediction of

HbA_{1c} outcome and were included in the final model, which included 14 069 patients with valid data for all five features. The full model equation and plots illustrating non-linear associations for continuous clinical features are reported in the appendix (pp 23–24). Model performance for predicting HbA_{1c} outcome in the derivation cohort is reported in the appendix (pp 25–26). Internal validation showed that the final model explained 29% of the variation in HbA_{1c} outcome (R^2 0·29) and was well calibrated for predicting absolute HbA_{1c} (calibration slope 0·9967).

There was evidence of marked heterogeneity in predicted individualised treatment effects, with the model predicting an HbA_{1c} benefit with SGLT2 inhibitors for 11 814 (84%) of 14 069 patients and a benefit with DPP-4 inhibitors for 2255 (16%) in the derivation cohort (appendix p 27). Across deciles of predicted individualised treatment effect, observed HbA_{1c} differences and predictions were similar, indicating predicted individualised treatment effects were well calibrated (appendix p 27).

In validation, calibration between observed HbA_{1c} differences and predictions were good in the CANTATA-D and CANTATA-D2 trials and the CPRD validation cohort (appendix p 27). In the smaller BI1245.20 trial deciles with the greatest predicted HbA_{1c} benefit on SGLT2 inhibitors had a clear observed benefit in line with predictions. In the CPRD validation cohort, 9376 of 10 751 had valid data to estimate model-predicted individualised treatment effects. 3756 (40·1%) of 9376 patients had a predicted benefit of 5 mmol/mol or greater with SGLT2 inhibitors and an observed benefit of 8·8 mmol/mol (95% CI 7·8–9·8) if they received SGLT2 inhibitors compared with DPP-4 inhibitors. These results were similar in the CANTATA-D and CANTATA-D2 trials (5·8 mmol/mol [3·9–7·7] observed benefit) and the BI1245.20 trial (6·6 mmol/mol [2·2–11·0] observed benefit; figure 4).

The model also identified a smaller group of patients with a potential HbA_{1c} benefit on DPP-4 inhibitor therapy. In CPRD validation, DPP-4 inhibitor concordant patients with a predicted benefit of 3 mmol/mol or greater who received DPP-4 inhibitors (450 [4·8%] of 9376) had a 1·8 mmol/mol (95% CI –2·5 to 6·0) observed benefit, and this benefit was similar in trials (figure 4). Differences in HbA_{1c} benefit were consistent when excluding 2518 (26·9%) of 9376 patients with established atherosclerotic cardiovascular disease, heart failure, chronic kidney disease, or SCORE risk of 5% or greater from the CPRD validation set (appendix p 30) and in trials for the 12-month HbA_{1c} outcome (appendix p 31).

Model performance for predicting HbA_{1c} outcome in all validation sets is reported in the appendix (pp 25–26). Calibration was good in CPRD validation but observed HbA_{1c} outcome was consistently lower than predicted HbA_{1c} outcome in the trials.

In CPRD, at 6 months patients initiating SGLT2 inhibitors had a greater median weight loss (–3·7 kg [IQR –4·3 to –3·2]) than patients initiating DPP-4 inhibitors (–1·0 kg [–1·6 to –0·5]; appendix p 32) and predicted differential HbA_{1c} response was not associated with weight change (figure 5A). Overall treatment discontinuation within 6 months was similar on SGLT2 inhibitors (median 16·1% [13·5 to 20·3]) versus DPP-4

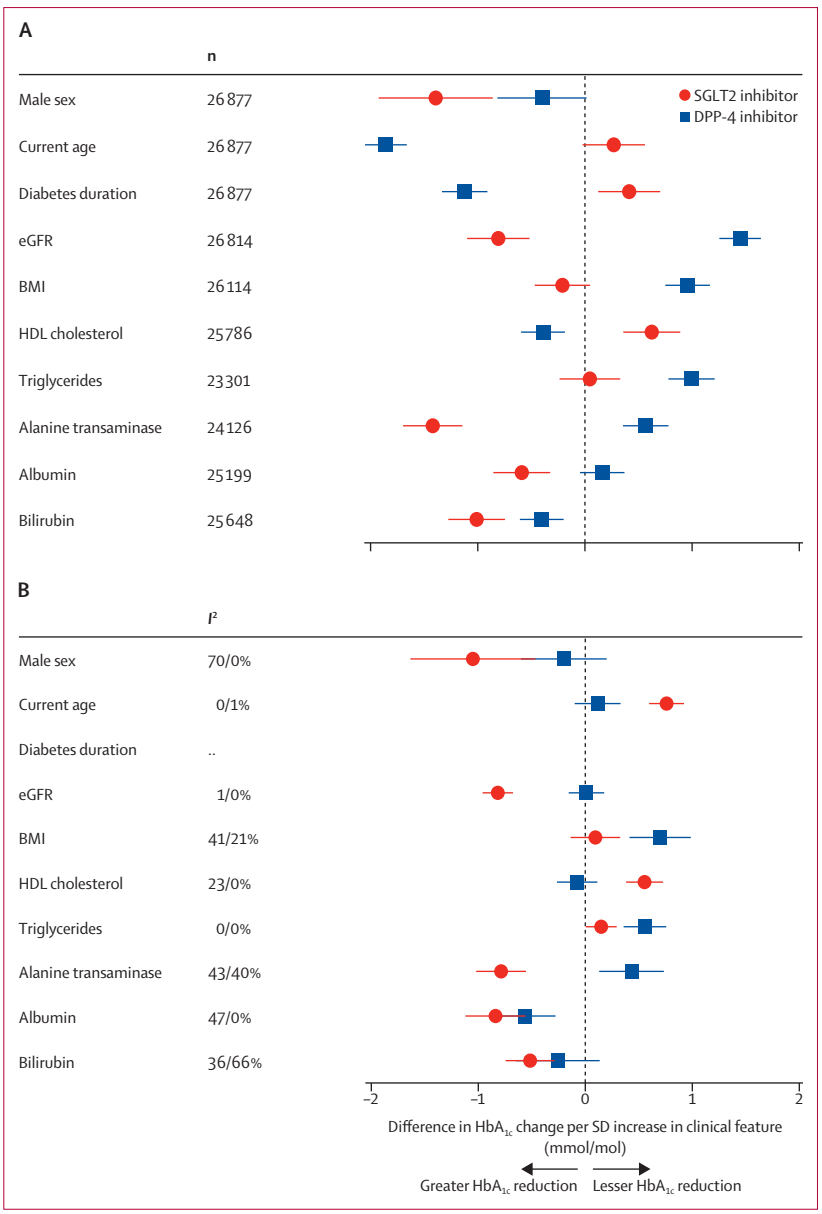

Figure 3: **Associations between clinical features and baseline adjusted 6-month HbA_{1c} response with SGLT2 inhibitor versus DPP-4 inhibitor treatment**

Bars show 95% CI. Underlying plot data, including estimates from each individual trial and the correlation matrix, are shown in the appendix (pp 13–16, 17). Estimates are derived from separate models for each clinical feature. (A) CPRD routine clinical data (n=28 877) adjusted for baseline HbA_{1c} by drug interaction, number of glucose-lowering drug classes ever prescribed and currently, and month of HbA_{1c} outcome measurement. (B) Meta-analysis data (14 clinical trials; n=10 414) adjusted for baseline HbA_{1c} by drug interaction. The association for diabetes duration is not reported in the trials as it was not available in most trial datasets. I^2 values are presented separately for SGLT2 and DPP-4 inhibitors. CPRD=Clinical Practice Research Datalink. eGFR=estimated glomerular filtration rate. HbA_{1c}=glycated haemoglobin.

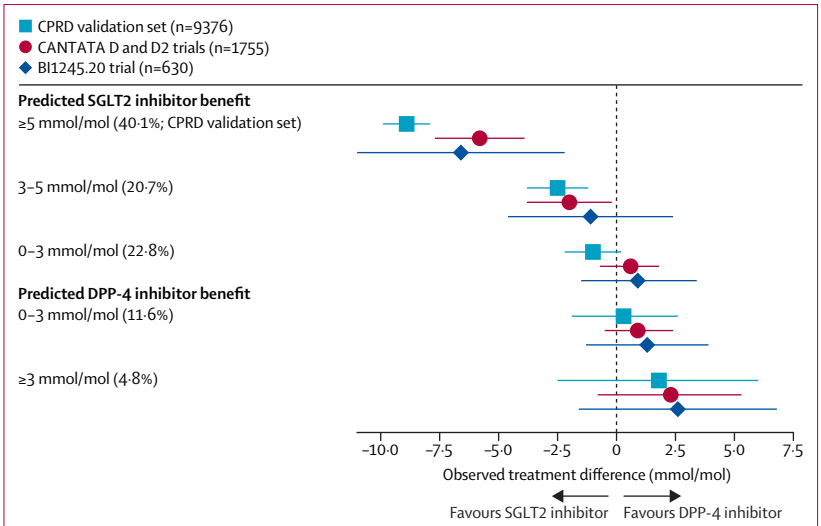

Figure 4: **Observed treatment effects across subgroups defined by clinical cutoffs of predicted treatment benefit in cohorts**

Bars show 95% CI. In CPRD, estimates are adjusted for clinical features in the treatment selection model, and trial estimates are unadjusted. All predicted treatment difference estimates and underlying plot data are shown in the appendix (pp 28, 29). CPRD=Clinical Practice Research Datalink.

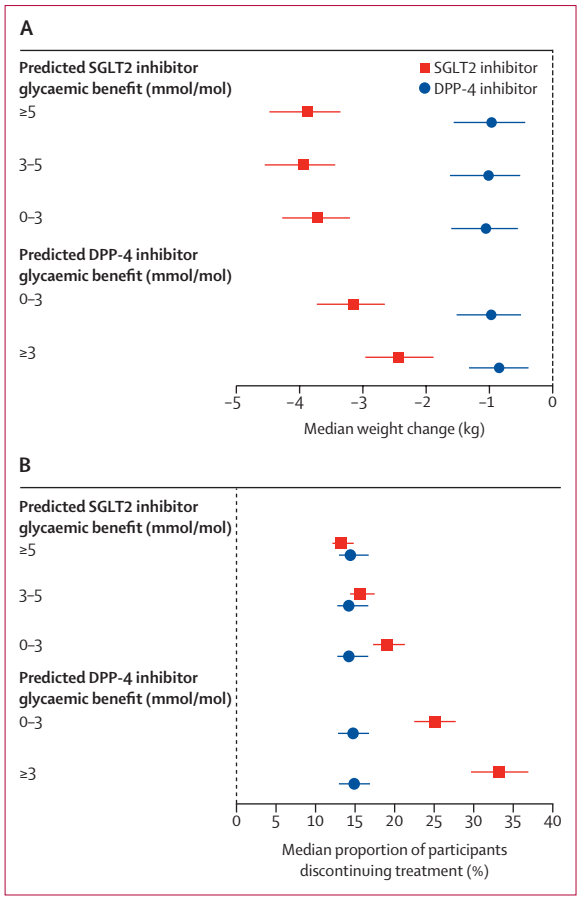

Figure 5: **6-month weight change and risk of treatment discontinuation across subgroups defined by clinical cutoffs of predicted treatment benefit in CPRD clinical data**

Data are median (IQR). 12-month outcomes are shown in the appendix (p 33). (A) Weight change at 6 months (n=15 627) for all participants with valid baseline data in the treatment selection model and with weight recorded 3–15 months after drug initiation (the closest measure to 6 months was used) in those receiving unchanged glucose-lowering therapy. (B) Risk of treatment discontinuation within 6 months (n=28 514) for all patients with valid baseline data in the glucose-lowering treatment selection model and with 3 additional months of follow-up to confirm that treatment has been discontinued.

inhibitors (14·4% [12·9 to 16·7]), and in patients predicted to have an HbA_{1c} benefit with SGLT2 inhibitors over DPP-4 inhibitors (median 15·2% [13·2 to 20·3] *vs* 14·4% [12·9 to 16·7]; figure 5B; appendix p 32). In patients with a predicted HbA_{1c} benefit of 3 mmol/mol or greater on DPP-4 inhibitors, discontinuation on DPP-4 inhibitors was lower than that observed for SGLT2 inhibitors (14·9% [13·0 to 16·9] *vs* 33·1% [29·7 to 36·9]; figure 5B; appendix p 32). Differences were consistent at 12 months (appendix p 33).

In 36 454 (87·2%) of 41 807 in the study eligible CPRD cohort (appendix p 3) with valid baseline data to fit the treatment selection model, patients with a predicted HbA_{1c} benefit of 5 mmol/mol or greater with SGLT2 inhibitor over DPP-4 inhibitor therapy (14 860 [40·8%]; around four in ten patients) were younger (median age 55 years [IQR 50–61]), predominantly male (66·6%), with a higher BMI (median 34·1 kg/m² [30·7–38·3]), baseline HbA_{1c} (median 80·2 mmol/mol [72–91]), eGFR (median 97 mL/min per 1·3 m² [89–104]), and alanine transaminase (median 37 IU/L [27–52]; appendix p 34). Only 2125 (14·3%; around one in seven patients) of 14 860 in this patient subgroup would be recommended SGLT2 inhibitors, due to their lower prevalence of atherosclerotic cardiovascular disease (1806 [12·2%]), heart failure (240 [1·6%]), and chronic kidney disease (257 [1·7%]).

Conversely, patients with a predicted HbA_{1c} benefit of 3 mmol/mol or greater with DPP-4 inhibitor over SGLT2 inhibitor therapy (1814 [5·0%]) were older (median age 79 years [IQR 75–84]), had an equal proportion of males (51·5%) and females (48·5%), with lower BMI (median 26·5 [24·0–30·0]), baseline HbA_{1c} (median 61 [57–66]), eGFR (median 59 [51–68]), and alanine transaminase (median 15 [12–20]; appendix p 34). Despite the predicted benefit of DPP-4 inhibitors for HbA_{1c} outcome, 1360 (75·0%) of this group would be recommended SGLT2 inhibitors based on current guidelines, reflecting their older age and high prevalence of atherosclerotic cardiovascular disease (619 [34·1%]), heart failure (152 [8·4%]), and chronic kidney disease (712 [39·3%]). In practice, 1609 (88·7%) of these patients received DPP-4 inhibitor therapy.

Discussion

Our study shows that the clinical features of people with type 2 diabetes are robustly associated with differential HbA_{1c} responses to SGLT2 inhibitor and DPP-4 inhibitor therapies. Combining just five routinely measured clinical features into a treatment selection algorithm can

identify a large patient subgroup (around four in ten of UK patients; appendix p 34) with a predicted glycaemic benefit of 5 mmol/mol or greater on SGLT2 inhibitors compared with DPP-4 inhibitors, and a similar risk of early discontinuation for both agents. We estimate only one in seven of this group would currently be recommended SGLT2 inhibitors based on their cardiorenal disease status or risk according to ADA/EASD treatment guidelines. We also identify a smaller group of patients who might have a greater glycaemic reduction and a lower risk of short-term discontinuation on DPP-4 inhibitors than those on SGLT2 inhibitors. However, due to older age and more severe cardiorenal disease profile, most of this group would be preferentially recommended SGLT2 inhibitors based on current guidelines.

Our analysis gives an example of translational precision medicine to inform the selection of type 2 diabetes therapy. Validation of findings in randomised clinical trial data provides a robust demonstration of the treatment selection algorithm. Although not all associations for individual clinical features observed in routine clinical data were replicated in the trials, specifically associations between higher age and lower eGFR with a greater DPP-4 inhibitor response, the algorithm performed well in validation. The validation framework based on strata defined by model predicted benefit is analogous to cardiovascular risk prediction models, such as QRISK, which are used routinely in current practice,[23] in which effectiveness reflects the ability of the model to accurately quantify risk at a population level, rather than to precisely define the time to a cardiovascular event for an individual. Individual-level estimates provided by the algorithm are not intended to be prescriptive, but instead to support more informed discussion between patients and clinicians on the benefits and risks of SGLT2 inhibitor and DPP-4 inhibitor treatment for an individual, alongside understanding of average level class effects, and the known cardiorenal benefit of SGLT2 inhibitors for those with, or at high-risk of, cardiorenal disease.

The combination of clinical features predicting differential glycaemic response most likely relates to differences in the underlying mechanism of action of the two therapies, although this hypothesis needs to be studied further. For example, increased urinary glucose excretion offers a possible explanation for the greater response to SGLT2 inhibitors in patients with higher baseline HbA1c and higher kidney function, and it is biologically plausible that factors associated with insulin resistance (ie, BMI and alanine transaminase concentrations) would reduce response to DPP-4 inhibitors that act primarily through potentiating insulin secretion, but not SGLT2 inhibitors which act though an independent mechanism.[24]

Our study has limitations. Notably, we were not able to validate differential treatment effects at the individual level, as while the outcome on the treatment received was observed for each individual, the counterfactual outcome the individual would have had if they received the other treatment was not observed (the fundamental problem of causal inference).[25] Although randomised crossover trials provide the optimal way to study differential treatment effects,[26] our validation in independent datasets, including head-to-head efficacy trials of the two therapies in which participants were randomly assigned and protocol driven follow-up was available, provides the next best evidence. As we developed the treatment selection model in routine clinical data, the potential of selection bias due to non-random treatment assignment means differential treatment effects for individual clinical features do not have a causal interpretation. However, our aim was not to uncover causal effects, but instead to develop a clinically useful treatment selection model and establish the generalisability and prediction accuracy of the model. We were unable to evaluate differential treatment effects by ethnicity, due to the low numbers of non-White individuals in routine and trial datasets. Although the trial validation in multicountry trial cohorts shows that the model is probably generalisable to other high-income countries, we acknowledge that further validation of the model in low-income and middle-income countries is needed before deployment in these settings. Although we incorporated cardiorenal risk stratification proposed in current ADA-EASD treatment guidelines in our evaluation,[1] we did not investigate individual-level heterogeneity in cardiovascular or renal outcomes. This investigation would ideally require long-term trial data of SGLT2 inhibitors compared with DPP-4 inhibitors, and such trials have not yet been conducted. In the absence of comparative trials, further population-based studies applying causal inference methods to evaluate whether differences in short-term efficacy shown in this study alter long-term cardiorenal outcomes would be of considerable interest. Because our aim was to develop a model based on a simple set of clinical features, our variable selection approach omitted routine clinical features such as diabetes duration and biomarkers such as HDL cholesterol which were modest predictors of differential HbA_{1c} outcome. We did not evaluate non-routine biomarkers, diet or nutrition, physical activity, or genetics and did not fully explore differential effects of different background therapy combinations (for which a larger dataset would be required) or medication adherence (because only information on the prescriptions issued is electronically captured). Future research to identify whether these factors are associated with differential treatment effects could improve model performance and provide insight into underlying mechanisms of heterogeneous drug action.[27] Finally, the relative performance of machine learning approaches designed to explicitly identify treatment effect heterogeneity[28] compared with the regression-based approach used in this study has yet to be evaluated.

By providing validated patient-level estimates of differences in glucose-lowering efficacy for two major

type 2 diabetes treatments, and assessing weight and discontinuation outcomes, our study has implications for clinical practice. The clinical features required to provide these estimates are available to most health professionals in high-income countries. Research to extend the prototype treatment selection decision aid to other type 2 diabetes treatment options is ongoing, informed by previously shown differential treatment effects for sulfonylurea, thiazolidinedione, and GLP-1 receptor agonist therapy based on routine clinical features.[9,29]

In conclusion, routine clinical features can identify a large group of people with type 2 diabetes with a marked glycaemic benefit on SGLT2 inhibitor therapy and no increased risk of discontinuing treatment with SGLT2 inhibitors compared with DPP-4 inhibitors. Only one in seven of this group would be preferentially recommended SGLT2 inhibitors based on current guidelines. A smaller group of predominantly older patients have a small glycaemic benefit on DPP-4 inhibitors and lower discontinuation risk on DPP-4 inhibitors compared with SGLT2 inhibitors. These findings show the potential of a precision medicine approach for people with type 2 diabetes based on routine clinical features and evaluating multiple patient-centred outcomes to inform clinical decisions concerning choice of optimal glucose-lowering treatment.

Contributors

JMD, AGJ, BMS, APM, and ATH conceptualised and designed the study. JMD, KGY, and BMS accessed and verified all the raw data in the study. JMD and KGY accessed the Clinical Practice Research Datalink (CPRD) and trial datasets and did the analysis with support from BMS. JMD, KGY, and BMS verified the final analysis. MDS developed the web calculator. All authors had full access to complete analysis results of the CPRD and trial datasets, provided support for the analysis and interpretation of results, and critically revised and approved the final manuscript. All authors were responsible for the decision to submit for publication.

Declaration of interests

JMD is supported by an independent fellowship funded by the UK Research and Innovation Expanding Excellence in England. APM received research funding from Eli Lilly and Company, Pfizer, and AstraZeneca. BAM is an employee of the Wellcome Trust; holds an honorary post at the Alan Turing Institute and University College London; and declares payments from the Medical Research Council, Health Data Research UK, British Heart Foundation, and Engineering and Physical Sciences Research Council (grant EP/N510129/). ASJV declares support from the University of Warwick, University of Kaiserslautern, and German Research Center for Artificial Intelligence; consulting fees from PUMAS; stock from Freshflow; and grant funding from The Alan Turing Institute (EP/N510129), Engineering and Physical Sciences Research Council, and Massachusetts Institute of Technology. WEH declares a grant from IQVIA and support from the National Institute for Health and Care Research (NIHR) Applied Research Collaboration South West Peninsula. ERP declares personal fees from Sanofi, Illumina, and Lilly. RRH reports research support from AstraZeneca, Bayer, and MSD; and personal fees from Anji Pharmaceuticals, Bayer, Novartis, and Novo Nordisk. NS declares personal fees from Abbott Diagnostics, Afimmune, Amgen, AstraZeneca, Boehringer Ingelheim, Eli Lilly, Hanmi Pharmaceuticals, MSD, Novartis, Novo Nordisk, Pfizer, and Sanofi; and grants from AstraZeneca, Boehringer Ingelheim, Novartis, and Roche Diagnostics. ATH and BMS are supported by the NIHR Exeter Clinical Research Facility. AGJ was supported by an NIHR Clinician Scientist fellowship (CS-2015–15–018) and declares research funding from the UK Medical Research Council, Diabetes UK (charity), Juvenile Diabetes Research Foundation (charity), and the European Foundation for the Study of Diabetes (charity). Representatives from GSK, Takeda, Janssen, Quintiles, AstraZeneca, and Sanofi attend meetings as part of the industry group involved with the MASTERMIND consortium. All declarations are outside of this study.

Data sharing

CPRD data are available by application to the CPRD Independent Scientific Advisory Committee and clinical trial data are accessible by application to the Yale University Open Data Access Project and Vivli.

Acknowledgments

This study was funded by a BHF-Turing Cardiovascular Data Science Award (SP/19/6/34809) and the UK Medical Research Council (MR/N00633X/1). This Article is partly based on data from the CPRD obtained under licence from the UK Medicines and Healthcare products Regulatory Agency, provided by patients, and collected by the NHS as part of their care and support. This publication is based on research using 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. This study, carried out under YODA Project (number 2017–1816), used data obtained from the Yale University Open Data Access Project, 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 does not necessarily represent the official views of the Yale University Open Data Access Project or Janssen Research & Development.

References

1 Buse JB, Wexler DJ, Tsapas A, et al. 2019 update to: management of hyperglycemia in type 2 diabetes, 2018. A consensus report by the American Diabetes Association (ADA) and the European Association for the Study of Diabetes (EASD). *Diabetes Care* 2020; **43:** 487–93.

2 Dennis JM, Henley WE, McGovern AP, et al. Time trends in prescribing of type 2 diabetes drugs, glycaemic response and risk factors: a retrospective analysis of primary care data, 2010–2017. *Diabetes Obes Metab* 2019; **21:** 1576–84.

3 Montvida O, Shaw J, Atherton JJ, Stringer F, Paul SK. Long-term trends in antidiabetes drug usage in the US: real-world evidence in patients newly diagnosed with type 2 diabetes. *Diabetes Care* 2018; **41:** 69–78.

4 Scheen AJ. SGLT2 versus DPP-4 inhibitors for type 2 diabetes. *Lancet Diabetes Endocrinol* 2013; **1:** 168–70.

5 McGovern A, Feher M, Munro N, de Lusignan S. Sodium-glucose co-transporter 2 (SGLT2) inhibitor: comparing trial data and real-world use. *Diabetes Ther* 2017; **8:** 365–76.

6 Chung WK, Erion K, Florez JC, et al. 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* 2020; **43:** 1617–35.

7 Ahlqvist E, Storm P, Käräjämäki A, et al. Novel subgroups of adult-onset diabetes and their association with outcomes: a data-driven cluster analysis of six variables. *Lancet Diabetes Endocrinol* 2018; **6:** 361–69.

8 Dennis JM, Shields BM, Henley WE, Jones AG, Hattersley AT. Disease progression and treatment response in data-driven subgroups of type 2 diabetes compared with models based on simple clinical features: an analysis using clinical trial data. *Lancet Diabetes Endocrinol* 2019; **7:** 442–51.

9 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.

10 Dennis JM, Shields BM, Hill AV, et al. Precision medicine in type 2 diabetes: clinical markers of insulin resistance are associated with altered short- and long-term glycemic response to DPP-4 inhibitor therapy. *Diabetes Care* 2018; **41:** 705–12.

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

12 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.

13 Rodgers LR, Weedon MN, Henley WE, Hattersley AT, Shields BM. 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.

14 National Institute for Health and Care Excellence. Type 2 diabetes in adults: management. 2015. https://www.nice.org.uk/guidance/ng28 (accessed June 28, 2021).

15 Conroy RM, Pyörälä K, Fitzgerald AP, et al. Estimation of ten-year risk of fatal cardiovascular disease in Europe: the SCORE project. *Eur Heart J* 2003; **24:** 987–1003.

16 Jones AG, Lonergan M, Henley WE, Pearson ER, Hattersley AT, Shields BM. Should studies of diabetes treatment stratification correct for baseline HbA1c? *PLoS One* 2016; **11:** e0152428.

17 Marston L, Carpenter JR, Walters KR, Morris RW, Nazareth I, Petersen I. Issues in multiple imputation of missing data for large general practice clinical databases. *Pharmacoepidemiol Drug Saf* 2010; **19:** 618–26.

18 Riley RD, Lambert PC, Abo-Zaid G. Meta-analysis of individual participant data: rationale, conduct, and reporting. *BMJ* 2010; **340:** c221.

19 Harrell FE Jr. Regression modeling strategies. Cham, Switzerland: Springer International Publishing, 2016.

20 Moons KG, Altman DG, Reitsma JB, et al. Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): explanation and elaboration. *Ann Intern Med* 2015; **162:** W1-73.

21 Bossuyt PM, Parvin T. Evaluating biomarkers for guiding treatment decisions. *EJIFCC* 2015; **26:** 63–70.

22 Janes H, Pepe MS, Bossuyt PM, Barlow WE. Measuring the performance of markers for guiding treatment decisions. *Ann Intern Med* 2011; **154:** 253–59.

23 Hippisley-Cox J, Coupland C, Brindle P. Development and validation of QRISK3 risk prediction algorithms to estimate future risk of cardiovascular disease: prospective cohort study. *BMJ* 2017; **357:** j2099.

24 Kim Y, Babu AR. Clinical potential of sodium-glucose cotransporter 2 inhibitors in the management of type 2 diabetes. *Diabetes Metab Syndr Obes* 2012; **5:** 313–27.

25 Holland PW. Statistics and causal inference. *J Am Stat Assoc* 1986; **81:** 945–60.

26 Senn S. Cross-over trials in clinical research. Chichester, UK: John Wiley & Sons, 2002.

27 Pearson ER. Diabetes: is there a future for pharmacogenomics guided treatment? *Clin Pharmacol Ther* 2019; **106:** 329–37.

28 Wager S, Athey S. Estimation and inference of heterogeneous treatment effects using random forests. *J Am Stat Assoc* 2018; **113:** 1228–42.

29 Jones AG, McDonald TJ, Shields BM, et al. Markers of β-cell failure predict poor glycemic response to GLP-1 receptor agonist therapy in type 2 diabetes. *Diabetes Care* 2016; **39:** 250–57.