NIHR Development Funding Scheme Report: Developing a working platform for a treatment selection algorithm for glucose lowering in Type 2 diabetes

Authors: John M Dennis[1], Catherine Angwin[1], Maddy Geen[1], Andrew T Hattersley[1], Beverley M Shields[1]

[1] University of Exeter Medical School. Address: Institute of Biomedical & Clinical Science, RILD Building, Royal Devon & Exeter Hospital, Barrack Road, Exeter EX2 5DW, UK

Project details

NIHR reference number: NIHR129108

Project period: 01 Jun 2019 to 29 Feb 2020

Chief Investigator: Andrew T Hattersley

Research Type: Primary Research

Background

Type 2 diabetes treatment selection algorithm overview

We have developed a prototype treatment selection algorithm for type 2 diabetes therapy using United Kingdom (UK) primary care data. The purpose of the algorithm is to provide a decision support tool for clinicians and patients, by identifying the drug class, or drug classes, likely to lower glucose the most for a particular individual based on their clinical characteristics. The algorithm provides estimates of likely glycaemic response (HbA1c) for the four drug classes currently recommended at first intensification in National Institute for Health and Care Excellence (NICE) guidelines: sulfonylureas, dipeptidyl peptidase 4 [DPP-4] inhibitors, sodium–glucose cotransporter 2 [SGLT2] inhibitors, and pioglitazone (a thiazolidinedione).[1] We have proof of concept for the algorithm, by showing that response to each of these drug classes varies markedly by simple clinical characteristics in both real-world primary care data and randomised trial data.[2-4]

No head-to-head trials have been conducted to compare even the average glucose-lowering efficacy of the four NICE recommended drug classes. The one ongoing parallel group trial of glucose-lowering effectiveness, The Glycemia Reduction Approaches in Diabetes: A Comparative Effectiveness Study (GRADE), includes sulfonylureas and DPP-4 inhibitors alongside insulin and glucagon-like peptide-1 [GLP-1] receptor agonists, but does not include SGLT2-inhibitors or pioglitazone (estimated completion data July 2021).[5] In contrast, routine clinical datasets such as the UK's Clinical Practice Research Datalink (CPRD) primary care database provide information on both patient characteristics and glucose-lowering effectiveness for each drug class. Such datasets therefore offer the opportunity to 1) evaluate heterogeneity of treatment effect; 2) develop an algorithm for treatment selection based on likely glucose-lowering efficacy. Once developed, the algorithm can then be validated in independent routine clinical datasets, and, where individual level data are available, for specific drug combinations in existing head-to-head drug trials.

An important point is that the purpose of the treatment selection algorithm is not to accurately predict response at the individual level (we do not believe this is possible due to the impact of non-drug related factors such as short term changes in diet)). Instead, the key outcome is reliable prediction of likely *differences* in response by drug class. At the individual level this is unobservable (a patient can only be allocated one drug at a single point in time meaning their potential response on another therapy is unobservable). Therefore, the way to test efficacy of the algorithm is to evaluate, at population-level, whether allocation to the drug(s) identified as 'best' by the algorithm gives on average better glycaemic control than the other drug options. This means

standard measures of model performance which assess predictive value do not have value for assessing performance for treatment selection.[6] A more useful performance measure in this context is the overall effect on glycaemic control that use of the algorithm would have in the population (see below: Algorithm validation).

Prototype algorithm development and validation

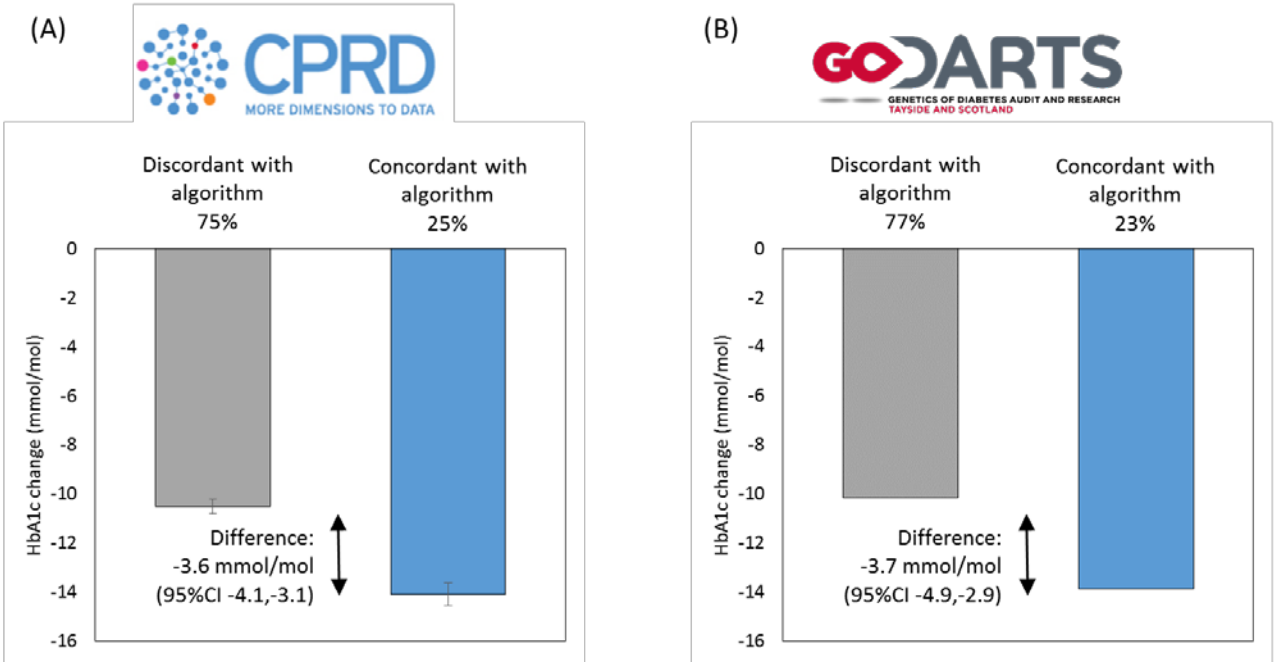
The first iteration of the type 2 diabetes treatment selection algorithm was developed in UK primary care data (CPRD, January 2018 download). It is based on the real-world treatment response (HbA1c) of nearly 50,000 individuals with type 2 diabetes to the four drug classes over January 2000 to December 2017 inclusive.

The algorithm itself is an ordinal regression model with achieved HbA1c at 12 months as a continuous measure as the outcome. An ordinal regression model has several advantages, including the ability to easily output the probability of a person achieving an individualised HbA1c target at 12 months. Inputs are clinical features and routine biomarkers easily available at the point a treatment decision is made (age at diagnosis, duration of diabetes, sex, baseline HbA1c, BMI, eGFR, HDL-c and ALT). All measures in the model have been demonstrated as robust predictors of response to at least one drug in both CPRD and in individual level data from existing randomised drug efficacy trials.[2-4] An important point is that some measures are associated with response to all drugs but not differential response. For example, higher baseline HbA1c is associated with increased response to all drug classes. In contrast, higher BMI is differential: it is associated with increased response to pioglitazone, lesser response to sulfonylureas and DPP4-inhibitors, but there is no association with SGLT2-inhibitor response.[2-4]

Early validation work has suggested the algorithm is likely to have clear clinical utility. For a treatment selection algorithm the key outcome is whether use of the drug(s) identified as 'best' by the algorithm gives better glycaemic control than the other drug options, not how accurate the algorithm is in predicting actual achieved HbA1c. It is this key outcome that we have tested in existing datasets. We have done so by comparing the glycaemic response of patients who actually received the drug the algorithm identified as best (people 'concordant' with the algorithm) against those who actually received a drug not indicated as best ('discordant' people). This is possible in trial datasets with active comparator arms as well as observational routine clinical datasets. In CPRD internal validation, concordant patients had a 3.6mmol/mol lower 12 month HbA1c compared with discordant patients (Figure 1A).

A similar difference in 12 month HbA1c was also shown in an independent primary care dataset of people with type 2 diabetes in Scotland (GoDARTs, [n=3,024])). In GoDARTs, people with type 2 diabetes concordant with the CPRD algorithm had a -3.7 (95%CI -4.9,-2.5) mmol/mol (p<0.001) greater response than discordant people (Figure 1B).

Figure 1: Changes in HbA1c at 12 months in concordant and discordant treatment selection subgroups



External validation has also been undertaken in reanalysis of individual level data from existing randomised trials based on their primary outcomes:

1) ADOPT randomised trial of TZD and SU therapy: concordant groups had a reduced risk of glycaemic failure compared with the discordant subgroup (Hazard ratio 0.60, 95%CI 0.44-0.82, p=0.001).[7]

2) CANTATA-D/D2 RCTs: -2.7 (95%CI -4.1,-1.3) mmol/mol (p<0.001) greater response in concordant *vs.* discordant subgroups.[8 9]

Aim

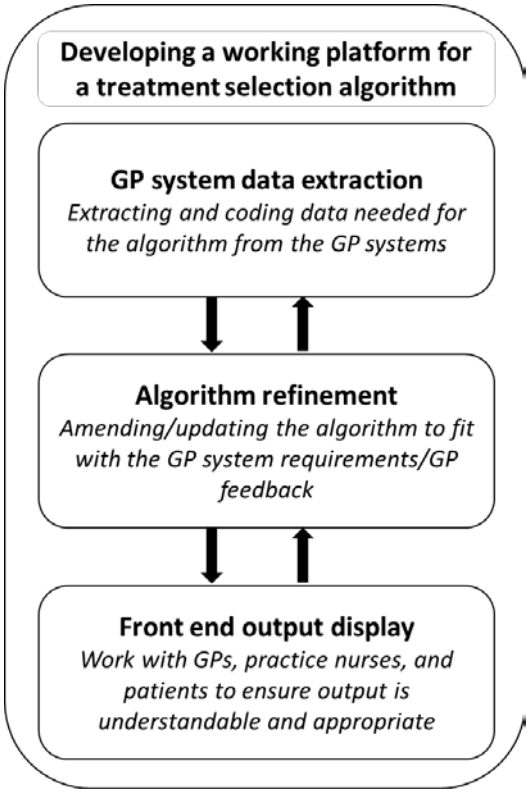
Study aims were as follows (**Figure 2**):

1) **General practice (GP) system data extraction**
   a) develop an infrastructure for extraction of patient data from primary care computer systems

2) **Front end output display**
   a) hold meetings with primary care physicians, practice nurses and patients to define the specification for the interface
   b) to develop a mock-up of the proposed interface
   c) to assess acceptability of and refine the proposed interface

3) **To further develop the algorithm**
   a) Continue algorithm validation
   b) Algorithm refinement based on Strands 1 and 2
   c) Algorithm update with new data

Figure 2: Schematic of DFS project



What was achieved

1) GP system data extraction:

We have established that the data is available within GP systems for the Type 2 diabetes treatment selection algorithm for glucose lowering and have developed the links and requirements for implementation.

Data: We estimated the number of clinical records with data required for the algorithm, shown in **Table 1**. We found that over 75% of people have all required data for the prototype algorithm available in their clinical record. Liver function measures were the data most missing from the record.

Work with the company First Databank, who provide GP alerts using real-time data extraction, has established systems that sit on top of GP software can retrieve the most recent event record but are not flexible. They are unable to search the entire record e.g. for age at diagnosis if based on first ever record of diabetes. Our conclusion is it will be most efficient to work with GP software companies directly.

Software links for implementation: We have developed links with the cancer research team (Exeter University) who are developing similar cancer risk algorithms, to find contacts for companies that provide GP software add-ons. This would allow us to develop directly within GP software which, as discussed above, will be the most efficient & effective way of implementing the algorithm.

Regulatory requirements: A major aspect is the need for software to comply with medical device regulations. Under the new EU Medical Device Directive due in May 2021 (delayed a year due to COVID-19), our algorithm will likely be deemed a Class IIa medical device. We have sought advice from consultants and learned of the key steps needed to obtain MHRA approval ("notice of no objection") before any software can be used in a trial, and the requirements for CE marking. These discussions have highlighted the extensive work required including:

1) Software development within a specific quality management system to ensure design, coding, and testing is recorded and aligned with the appropriate standards (developer must have ISO13485 certification).
2) Risk assessments to mitigate all possible risks of errors/misuse/misinterpretation
3) A portfolio of clinical evaluation to demonstrate clinical validity
4) Clear labelling, guidance and information on intended use, data privacy.

5) Ongoing updating & monitoring is essential, & long-term monitoring of usage/maintenance/sustainability required.

Table 1: Availability of clinical information in GP computer systems (accessed in collaboration with First Databank), and CPRD

	First Databank OptimiseRx	CPRD		
Feature	**Feature extraction possible (Yes/No/unknown)***	**Feature extraction possible** (Yes/No/unknown)**	**N with feature available in previous 6 months** (N[% total])**	**N with feature available in previous 2 years** (N[% total])**
Type 2 diabetes diagnosis	YES	YES	17618 (100%)	17618 (100%)
HbA1c > 58mmol/mol	YES	YES	(95.8%) (95.8%)	NA***
HbA1c	YES	YES	16870 (95.8%)	NA***
2nd most recent HbA1c	NO	YES	7096 (40.3%)	15948 (90.5%)
Fasting glucose			5801 (32.9%)	9931 (56.4%)
Current treatment with metformin	YES	YES	17618 (100%)	17618 (100%)
Currently/Previously treated with sulfonylureas	YES	YES	17618 (100%)	17618 (100%)
Currently/Previously treated with pioglitazone	YES	YES	17618 (100%)	17618 (100%)
Currently/Previously treated with DPP4-inhibitors	YES	YES	17618 (100%)	17618 (100%)
Currently/Previously treated with SGLT2-inhibitors	YES	YES	17618 (100%)	17618 (100%)
Current age	YES	YES	17618 (100%)	17618 (100%)
Sex	YES	YES	17618 (100%)	17618 (100%)
Ethnicity	Unknown	YES	9971 (56.6%)	9971 (56.6%)
Age at diabetes diagnosis	NO	YES	17618 (100%)	17618 (100%)
BMI	YES	YES	13694 (77.7%)	16635 (94.4%)
Weight	YES	YES	13810 (78.4%)	16783 (95.3%)
Creatinine	YES	YES	15905 (90.3%)	17519 (99.4%)
eGFR	YES	YES	15905 (90.3%)	17519 (99.4%)
Triglycerides	YES	YES	11538 (65.5%)	15167 (86.1%)
HDL-c (HDL cholesterol)	YES	YES	13479 (76.5%)	16844 (95.6%)
LDL-c (LDL cholesterol)	YES	YES	11534 (65.5%)	14862 (84.4%)
Albumin	YES	YES	13725 (77.9%)	16354 (92.8%)
Bilirubin	Unknown	YES	13984 (79.4%)	16707 (94.8%)
ALT (Alanine aminotransferase)	YES	YES	13331 (75.7%)	15906 (90.3%)
AST (Aspartate Aminotransferase)	Unknown	YES	1685 (9.6%)	2463 (14.0%)

*Confirmed by First Databank via their OptimiseRx platform

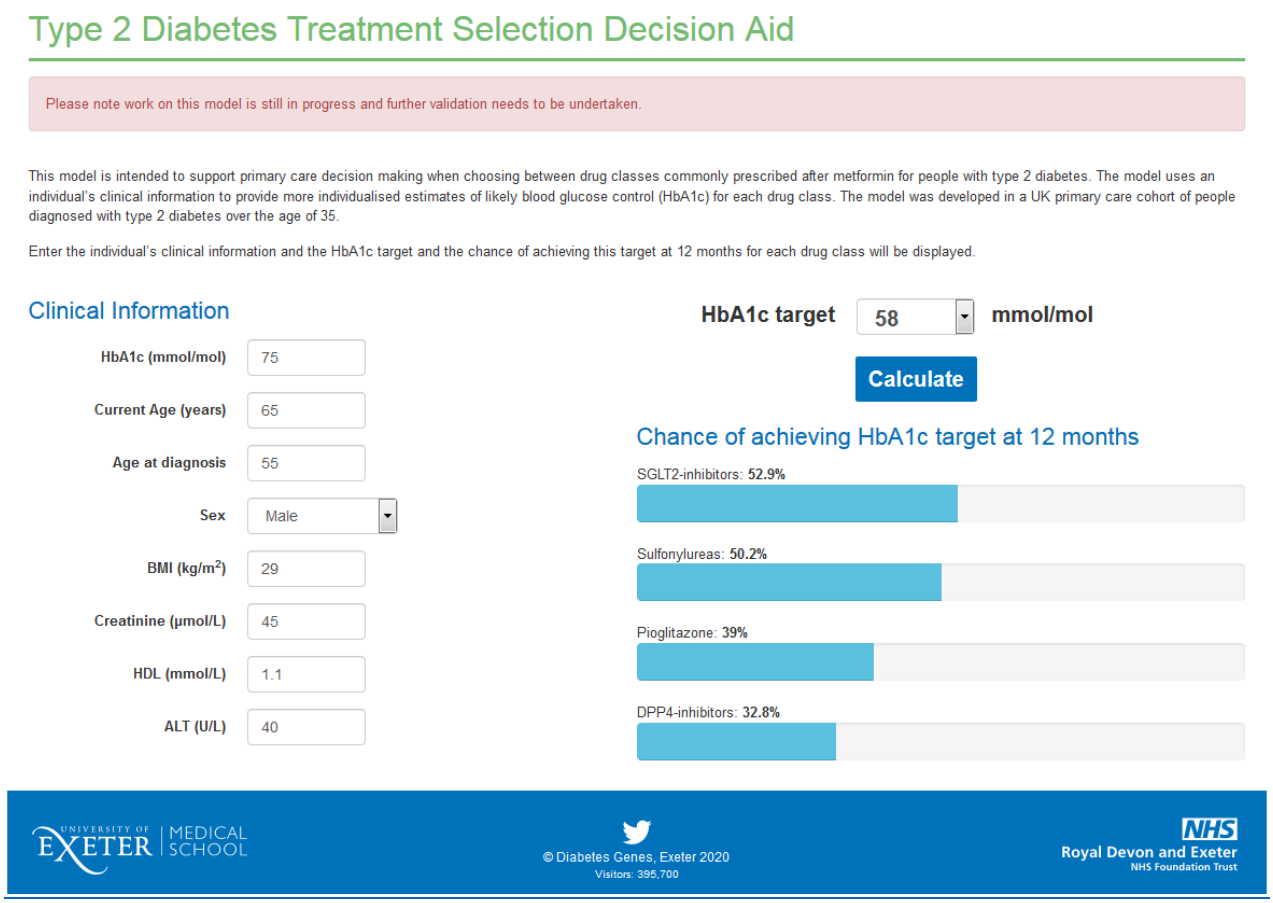
**At date of initiation of oral second-line treatment (all treatment initiations 2015 onwards)

*** HbA1c within 6 months prior to date of drug initiation to provide a valid baseline

2) Front end output display:

We have developed a potential web interface for the algorithm which was based on feedback from Healthcare professionals during focus groups. To inform the interface we conducted focus groups with GPs, practice nurses and patients. These were very positive and feedback suggested the algorithm was useful and acceptable. We gained critical information on how to display the output: with the user first setting an individualised target HbA1c and then the output displayed as the probability of achieving that target for each drug (**Figure 3**).

Figure 3: Screenshot of prototype web interface for the type 2 diabetes treatment selection decision aid for primary care



Feedback from the focus groups has also established the interface needs to include more than just estimation of outcome prediction:

- Provide information on risks (e.g. side-effects, treatment discontinuation, weight gain).
- GPs would like a tab explaining the evidence in order to trust the results.
- GLP1-receptor agonists should be added.
- Information on which drugs were contraindicated & previous drug history should be automatically populated
- Ability to adapt algorithm for local practice is essential (e.g. refining drug lists according to local CCG).

The general consensus was that it would be most helpful to see the algorithm before rather than during a consultation to allow preparation. There was concern use of the algorithm as an audit tool for identifying patients who needed treatment could create additional work for already hard pressed practices.

3) Algorithm development and validation:

We updated the algorithm using an updated CPRD download (July 2019). We extracted clinical information on people with type 2 diabetes in CPRD initiating one of the drug classes of interest (DPP-4 inhibitors, SGLT2 inhibitors, sulfonylureas, pioglitazone), identified and extracted as previously reported.[10] Baseline recorded clinical features for the model comprised HbA1c, sex, age, duration of diabetes, BMI, eGFR, HDL-c and ALT. Only people with all valid baseline clinical information to inform the algorithm (valid records of all measures required within six months of the drug start date) and a valid outcome HbA1c (closest HbA1c to 12 months between 6 month and 15 months after drug initiation whilst on unchanged therapy). As missing data were considered likely to be missing not at random multiple imputation was not used for model development. The final study population (n=56,851) was randomly split into a development cohort (80%, n=45,530) and validation cohort (n=11,321) with a proportionally equal allocation of each drug class to each cohort.

The model outcome was achieved 12 month HbA1c. For model development, a penalised ordinal regression model was fitted with 12 month HbA1c modelled as a continuous response variable and all baseline clinical features described above, with additional adjustment for line of therapy (the number of glucose-lowering drug classes previously initiated), and number of current glucose-lowering drugs. Heterogeneity of treatment effects was modelled using clinical features by drug

class interaction terms. All continuous clinical features were modelled using non-linear restricted cubic splines with 3 knots. Separate penalty terms were fitted for main effects, non-linear effects, and interaction effects.

Initial model validation was performed in the CPRD validation cohort. The algorithm shows good performance as assessed by the greater HbA1c response in the subgroup of people concordant with the treatment selection model (those who were actually prescribed the drug identified as optimal in terms of HbA1c response by the algorithm), compared to the subgroup of discordant people (those prescribed a drug identified as non-optimal), with similar absolute differences to those observed in the development cohort (**Figure 4**).

In sensitivity analysis we assessed the impact of the time frame for retrieval of data such as blood test results. We found the algorithm performance to be robust with blood test results up to 2 years prior to the drug start date, suggesting the required clinical information can be automatically populated from the electronic health record without additional testing for over 75% of contemporary people with type 2 diabetes. We are currently evaluating how to handle the remaining missing data – focus groups reported it is important for the algorithm to be able to run from only partial data.

Difficulties encountered

As detailed above, the algorithm will be harder to implement if it needs to operate within a system that sits on top of GP software systems and is unable to store data. This is mitigated if the algorithm is developed within GP software rather than as an add-on. This will require partnership with the GP software companies. This is something that is likely to require further funding and resources in the near future, to help work with commercial partners.

Our work with consultants on medical device regulations has highlighted the considerable amount of work required in ensuring that any software developed incorporating the algorithms is compliant with regulations in order to be approved by MHRA for pilot trials. We have established what is required and which partners we will need for this.

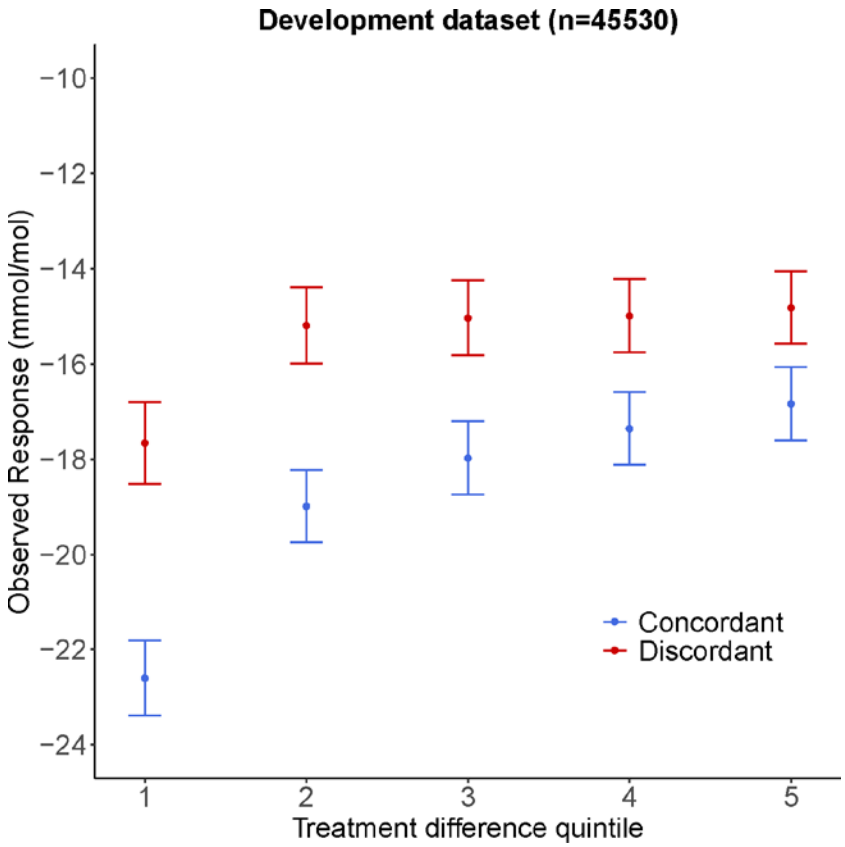
Conclusions

NIHR Development Funding Scheme (DFS) funding has allowed us to perform essential preliminary work to 1) refine and validate algorithm for type 2 diabetes treatment selection based on glycaemic response, 2) establish the necessary steps required for implementing the algorithm as a decision aid into GP software, 3) consult with clinicians to identify the most important features for a decision aid to maximise clinical utility. This work has provided a strong foundation in which to apply for the next

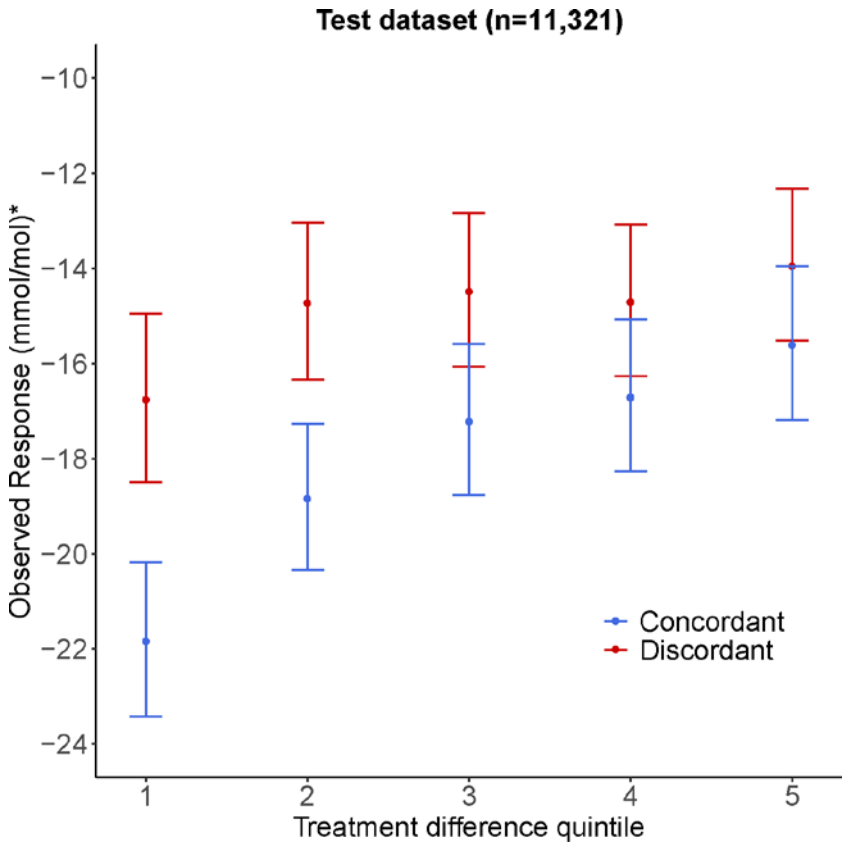
stage of funding to develop final clinically implemented models and to conduct a prospective trial of the type 2 diabetes treatment selection algorithm.

Figure 4: Observed response in people prescribed their optimal therapy (as predicted by the algorithm) [concordant group] compared to people prescribed a non-optimal therapy [discordant group], across quintiles defined by predicted difference in HbA1c (quintile 1=highest predicted difference in HbA1c). A lower (more negative) observed response indicate greater glucose-lowering.

A) Development set (CPRD)



B) Test set (CPRD)



Acknowledgements

We thank the NIHR for their Development Funding Scheme award to help fund this project (ref 129108). Much of the initial work which this was based on was carried out as part of the Medical Research Council's MASTERMIND programme (Stratified medicine in Type 2 diabetes).

References

1. National Institute for Health and Care Excellence (2015) Type 2 diabetes in adults: management. NICE guideline (NG28).
2. 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(4):705-12. doi: 10.2337/dc17-1827 [published Online First: 2018/02/02]
3. 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(9):1844-53. doi: 10.2337/dc18-0344 [published Online First: 2018/08/04]
4. Kimmitt RA. DJM, Weedon M., Rodgers L.R., Jones A.G., Pearson E.R., Hattersley A.T., Oram R.A. and Shields B.M. Science in clinical practice 2. *Diabetic Medicine* 2018;35(S1):33-35. doi: doi:10.1111/dme.14_13570
5. Nathan DM, Buse JB, Kahn SE, et al. Rationale and Design of the Glycemia Reduction Approaches in Diabetes: A Comparative Effectiveness Study (GRADE). *Diabetes care* 2013;36(8):2254-61. doi: 10.2337/dc13-0356
6. Janes H, Pepe MS, Bossuyt PM, et al. Measuring the performance of markers for guiding treatment decisions. *Annals of internal medicine* 2011;154(4):253-9. doi: 10.7326/0003-4819-154-4-201102150-00006 [published Online First: 2011/02/16]
7. Kahn SE, Haffner SM, Heise MA, et al. Glycemic durability of rosiglitazone, metformin, or glyburide monotherapy. *The New England journal of medicine* 2006;355(23):2427-43. doi: 10.1056/NEJMoa066224 [published Online First: 2006/12/06]
8. Lavalle-González FJ, Januszewicz A, Davidson J, et al. Efficacy and safety of canagliflozin compared with placebo and sitagliptin in patients with type 2 diabetes on background metformin monotherapy: a randomised trial. *Diabetologia* 2013;56(12):2582-92. doi: 10.1007/s00125-013-3039-1
9. Schernthaner G, Gross JL, Rosenstock J, et al. Canagliflozin compared with sitagliptin for patients with type 2 diabetes who do not have adequate glycemic control with metformin plus sulfonylurea: a 52-week randomized trial. *Diabetes care* 2013;36(9):2508-15. doi: 10.2337/dc12-2491 [published Online First: 2013/04/09]
10. 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(10):e017989. doi: 10.1136/bmjopen-2017-017989 [published Online First: 2017/10/14]