Precision Medicine in Type 2 Diabetes: Using Individualized Prediction Models to Optimize Selection of Treatment

John M. Dennis

Despite the known heterogeneity of type 2 diabetes and variable response to glucose lowering medications, current evidence on optimal treatment is predominantly based on average effects in clinical trials rather than individual-level characteristics. A precision medicine approach based on treatment response would aim to improve on this by identifying predictors of differential drug response for people based on their characteristics and then using this information to select optimal treatment. Recent research has demonstrated robust and clinically relevant differential drug response with all noninsulin treatments after metformin (sulfonylureas, thiazolidinediones, dipeptidyl peptidase 4 [DPP-4] inhibitors, glucagon-like peptide 1 [GLP-1] receptor agonists, and sodium–glucose cotransporter 2 [SGLT2] inhibitors) using routinely available clinical features. This Perspective reviews this current evidence and discusses how differences in drug response could inform selection of optimal type 2 diabetes treatment in the near future. It presents a novel framework for developing and testing precision medicine–based strategies to optimize treatment, harnessing existing routine clinical and trial data sources. This framework was recently applied to demonstrate that “subtype” approaches, in which people are classified into subgroups based on features reflecting underlying pathophysiology, are likely to have less clinical utility compared with approaches that combine the same features as continuous measures in probabilistic “individualized prediction” models.

Type 2 diabetes is a complex disease, characterized by hyperglycemia associated with varying degrees of insulin resistance and impaired insulin secretion and influenced by nongenetic and genetic factors. Despite this, glucose-lowering treatment is similar for most people. Current type 2 diabetes guidelines recommend the choice between glucose-lowering treatment options is based on clinical characteristics (1), an approach in line with the central goal of precision medicine: the tailoring of medical treatment to an individual. After initial metformin, the most recent guidelines recommend glucagon-like peptide 1 receptor agonists (GLP-1RA) or sodium–glucose cotransporter 2 inhibitors (SGLT2i) in people with established atherosclerotic cardiovascular disease, heart failure, or chronic kidney disease, but this stratification only applies to up to 15–20% of people with type 2 diabetes (2,3). For the remaining majority, evidence of benefit beyond glucose lowering with these drug classes has not been robustly demonstrated, and the optimal treatment pathway is not clear (1). Evidence on the key considerations, notably glucose-lowering efficacy, tolerability, and side effects, is mainly derived from average treatment effects from clinical trials. This means there is little information available on whether a specific person in the clinic is more or less likely than the average trial participant to respond well to a particular treatment or develop side effects. Given this knowledge gap, there is currently great interest in developing approaches that can characterize people beyond the standard type 2 diabetes phenotype and use this heterogeneity to optimize the selection of glucose-lowering treatment.

Any successful implementation of precision medicine in type 2 diabetes is likely to be very different from the most successful examples of precision medicine to date. These have been in cancer and single-gene diseases such as monogenic diabetes, where expensive genetic testing defines the etiology and the specific etiology helps to determine treatment (4,5). In type 2 diabetes, unlike cancer, tissue is not
available, and unlike rare forms of diabetes, current genetic testing does not allow clear definition of the underlying pathophysiology (6). This makes identification of discrete, non-overlapping subtypes of type 2 diabetes much less likely (7).

In this Perspective, I focus on a fundamental aim of precision medicine—the selection of optimal type 2 diabetes treatment based on likely differences in drug effect (henceforth, heterogeneity of treatment effect [HTE]). I provide an overview of the evidence from recent studies of HTE in type 2 diabetes and present a framework for using existing routine clinical and trial data sources to develop and test precision medicine–based strategies to optimize treatment. The focus is on glycemic response, as nearly all current evidence of HTE for diabetes drugs is for differences in HbA1c. However, the framework outlined can easily be extended to evaluate HTE for nonglycemic end points, including microvascular and macrovascular complications. Type 2 diabetes is a highly prevalent condition with relatively inexpensive treatment, meaning precision medicine approaches based on inexpensive markers have greatest potential to translate into clinical practice in the near future. As a result, this article concentrates on the use of routinely available clinical features to select optimal treatment, although the principles discussed equally apply to the use of genomic or nonroutine biomarkers (6). Recent reviews of the pharmacogenomics of type 2 diabetes drug response are available elsewhere (8,9).

Why Type 2 Diabetes Glucose-Lowering Treatment Is an Excellent Candidate for a Precision Medicine Approach

Type 2 diabetes treatment is an excellent candidate for a precision medicine approach for the following reasons. 1) There are many different drug classes available after metformin with different mechanisms of action but the same principal aim: to lower blood glucose. 2) At the individual level, glucose-lowering response to each drug appears to vary greatly (Fig. 1). 3) There is not a clear “best” overall treatment outside a small proportion of individuals with specific complications. For the remainder, current treatment guidelines do not provide information on which drug class is best for lowering blood glucose, for which people (1). 4) There is great heterogeneity in the clinical phenotype of type 2 diabetes, making it plausible that people with different underlying pathophysiology will have varying responses to the different drug classes, depending on the mechanism of action of the drug.

Defining the Treatment Selection Approach in Type 2 Diabetes

Despite the large biological noise in HbA1c, the majority of people appear to respond when initiated on a glucose-lowering drug (Fig. 1), and it is unlikely that many who appear not to respond are true “nonresponders” (10). Therefore, the aim of precision medicine in type 2 diabetes is not to identify people who will and will not respond (which can only be achieved through repeated crossover trial designs [11,12]) but instead to identify people who are likely to have a greater relative benefit from one drug class over another. This means that the necessary first step is to identify whether there are markers robustly predictive of greater or lesser response to each drug class to a clinically significant degree. In the absence of single markers with huge effect sizes, which have not been found to date, the second step is to optimally use multiple markers in combination to select treatment for individuals.

Identifying Robust Predictors of Type 2 Diabetes Treatment Response Using Routine and Trial Data

A focus on identifying routine clinical markers means HTE can be evaluated using existing observational and trial data sets that capture information on the drug response of people initiating type 2 diabetes treatment. The conventional approach is to examine HTE in clinical trials using “one-at-a-time” subgroup analysis in which participants are subcategorized based on a set of single characteristics in turn, such as sex and age (old vs. young). This approach does not provide credible evidence of differences in response due to low statistical power, lack of multivariable adjustment, and the risk of false-negative and false-positive findings (13). This means very few “discovered” positive subgroups are subsequently replicated (14,15).

While subgroup analysis of trials is limited, a combination of large observational routine clinical data sets and trial data (increasingly available [16,17]) provides a powerful starting point to robustly evaluate HTE. Large anonymized routine clinical electronic health record databases, such as the U.K.’s Clinical Practice Research Datalink (18), provide a rich source of “real-world” information on demographics, clinical features, diagnoses, laboratory tests, and prescriptions. One two-step approach to “triangulate” routine and trial data sources is shown in Fig. 2, on the basis that the best evidence for robust HTE is replication of effect in multiple independent data sets with differing strengths and weaknesses. In step 1, due to the large sample size and availability of head-to-head data for all drug classes, routine clinical data are used for “discovery” analysis, with assessment of drug-by-marker interactions to identify candidate features associated with differential response across drug classes. As in these observational data drug selection is not random and there are likely to be large differences in baseline clinical features between treatment groups, careful identification of confounders and statistical adjustment is required. To further reduce bias, the use of causal inference methods such as inverse probability of treatment weighting (19), or target trial approaches where studies are set up to emulate the design of an “ideal” randomized trial, should be considered (20). Nonetheless, unmeasured confounding may still bias findings, meaning a second step of external validation is required to confirm findings. In step 2, specific markers associated with potentially clinically relevant differences in drug response can be tested for reproducibility as prespecified hypotheses in clinical trial data sets where treatment allocation is randomized and blinded and where there is...
systematic baseline assessment and follow-up, meaning the risk of confounding is much lower (21). This two-step approach takes advantage of the larger, more heterogeneous population in routine care data sets for feature discovery while minimizing the risk of data mining in the smaller, richer trial data sets.

What Clinical Features Alter Type 2 Diabetes Treatment Response?
Recent studies have demonstrated clinically relevant differences in response by clinical features for all noninsulin glucose-lowering drug classes commonly used after metformin. Studies that do not adjust for baseline HbA1c are not reported here, given the demonstrated risk of false associations in such analysis (22).

Sulfonylurea and Thiazolidinedione Treatment
The first robust demonstration of HTE for type 2 diabetes therapy used the routine and trial data framework previously described to evaluate differential response to sulfonylurea (SU) and thiazolidinedione (TZD) treatment. Observational data from U.K. primary care data were used as a discovery data set, in which it was demonstrated that males without obesity (BMI $<30$) have on average a greater glucose-lowering response with SU compared with TZD treatment, while, conversely, females with obesity (BMI $\geq 30$) have a greater response to TZD than SU treatment (21). Differences in response in these subgroups were then validated, and confirmed to hold for long-term response, in randomized trial replication data, with differences in effect size within these subgroups equivalent to the addition of another glucose-lowering treatment (Fig. 3).

Dipeptidyl Peptidase 4 Inhibitors and GLP-1RA
With dipeptidyl peptidase 4 inhibitors (DPP-4i), the prospective Predicting Response to Incretin Based Agents (PRIBA) study demonstrated that markers of higher insulin resistance are consistently associated with lesser glucose-lowering response in non–insulin-treated participants (23). Differences were clinically relevant; a subgroup defined by obesity (BMI $\geq 30$) and high triglycerides ($\geq 2.3$ mmol/L) (31% of participants) had a response less than half that of a nonobese, low triglyceride ($<2.3$ mmol/L) subgroup (22% of participants) (6-month response $-5.3$ mmol/mol [−0.5%] and $-11.3$ mmol/mol [−1.0%], respectively). Conversely, there was no evidence of an association between markers of insulin resistance and glucose-lowering response for non–insulin-treated people initiating GLP-1RA (Fig. 4). Results were replicated in U.K. primary care data. Interestingly, in insulin-treated people but not in non–insulin-treated people, the same study found that with GLP-1RA, clinical markers of low β-cell function such as lower C-peptide and longer duration of diabetes were associated with reduced glucose-lowering efficacy (24). With DPP-4i, several other studies support the association between lower BMI, lower insulin resistance, and greater response and also suggest a benefit in glucose-lowering for people of Asian ethnicity (25,26).

SGLT2i
Analysis of trial data have reported markedly greater relative benefit with SGLT2i at higher baseline HbA1c levels compared with DPP-4i or SU treatment (27,28). Differences in response with SGLT2i have also been observed by baseline renal function. While the reduced efficacy
of SGLT2i at estimated glomerular filtration rates (GFRs) <60 mL/min/1.73 m² is well established (29), pooled trial analysis has demonstrated that this likely extends across the normal range, meaning that people with baseline eGFR >90 mL/min/1.73 m² have a greater response compared with those with eGFR 60–90 mL/min/1.73 m² (30). In contrast, with DPP-4i, response is likely maintained in people at lower eGFRs (31). Early work by our group suggests that these differential treatment effects for SGLT2i and DPP-4i are replicated in U.K. primary care data (Fig. 5).

Figure 2—A “triangulation” approach using routine clinical and trial data to evaluate differences in drug response, and the strengths and weaknesses of each data source.

Figure 3—Five-year glycemic response (change from baseline in HbA₁c) with TZD and SU treatment in males without obesity (BMI <30) (A) and females with obesity (BMI ≥30) (B) subgroups in 1,232 participants in the ADOPT clinical trial (21). Data are presented as means ± SD at each study visit from mixed-effects models. A reduction (improvement) in HbA₁c is represented as a negative value. For area-under-the-curve (AUC) difference estimates, positive values favor SU, and negative values favor TZD. Adapted from Dennis et al. (21).
Factors Altering Treatment Response May Relate to the Underlying Mechanism of Action of Different Drug Classes

The identified clinical features associated with HTE in many cases relate to the known mechanisms of action of the different drug classes. Such “plausibility of effect modification” greatly strengthens the credibility of HTE analysis (13). For TZD, in addition to the increased insulin resistance with higher BMI, variation in response by sex and obesity is likely to reflect associated differences in adipocyte distribution and function, as these drugs primarily act on adipose tissue (32,33). For SU and DPP-4i, which stimulate insulin secretion by the β-cell, the association between reduced insulin sensitivity and higher BMI possibly explains greater response in nonobese people. However, this does not explain the lack of association between insulin resistance and glucose lowering for the other incretin-based drug class, GLP1-RA; it is possible this difference could relate to the added weight-loss effects of this medication class or that GLP-1RA response was studied in an almost entirely obese (and therefore insulin resistant) population (23). The lack of GLP-1RA glycemic benefit in insulin-treated participants with very severe endogenous insulin deficiency is also consistent with the known role of potentiation of endogenous insulin secretion in their action. Effects on urinary glucose excretion provide a likely explanation for the variation in glucose-lowering efficacy of SGLT2i with baseline HbA1c and cotreatment change. Error bars denote 95% CI. Adapted from Dennis et al. (23). HOMA2-IR, HOMA of insulin resistance; SHBG, sex hormone binding globulin; UCPCR, urine C-peptide creatinine ratio.

How Can Differences in Treatment Response Inform Selection of Optimal Treatment?

While evidence of robust differences in type 2 diabetes treatment response is growing, there is current debate and considerable uncertainty about how to translate this to inform decision-making in clinical practice. Recent literature has focused on the following two approaches (Fig. 6).

The first approach is a “subtypes” approach, in which people with type 2 diabetes are subclassified based on their underlying pathophysiology (whether clinical, genetic, phenotypic, or biomarker traits) on the assumption that once subtypes are defined, they will have utility to stratify therapeutic decisions and other outcomes such as progression to complications. This was recently and notably proposed by Ahlqvist et al. (34) in a sex-stratified, data-driven cluster
analysis of people close to diabetes diagnosis that grouped individuals with similar underlying pathophysiology using five clinical features (age at diagnosis, BMI, HbA1c, and HOMA-measured insulin resistance and insulin sensitivity) in Scandinavian registry data. Importantly, similar-looking subgroups were identified when the analysis was repeated in multiple international population-based cohorts (35,36). Subgroups showed differences in outcomes in observational follow-up, although differential treatment response was not assessed. Several other data-driven classifications have recently been proposed with substantial variation in the features used for classification and the numbers of subgroups identified (37–39), including genetically defined clusters (40,41), but their utility to stratify treatment response has similarly not been assessed.

The second approach is to use a person’s specific clinical information in a probabilistic “individualized prediction” approach. In this approach, markers reflecting underlying pathophysiology are used as continuous traits to directly predict an individual’s treatment response for each drug. An individual’s specific information can then be used to predict their likely best drug in terms of glucose-lowering response (or, alternatively, to identify the absence of clinically relevant differences in response across treatments), and these predictions can guide selection of optimal treatment. The model developed is specific to the outcome of treatment response and can be deployed based on a person’s current information at the point a decision to escalate treatment is made. Although subtypes could then in theory be specified based on the prediction of differential response or optimal therapy, this would make little sense, as the subtypes would be based on clinical parameters that vary over time and are affected by treatment, meaning that for an individual, subtype assignment is unlikely to be stable. This proposed approach is consistent with the ideas underlying the recently proposed “palette model” of diabetes (7), which, at a specific point in time, positions an individual with diabetes on a spectrum of phenotypic variation and uses this position to predict likely outcome.

While the advantages and disadvantages of each approach in the context of selecting optimal treatment are shown in Fig. 6, the fundamental difference between the two approaches is that the subtypes approach assumes homogeneity of differential treatment response for all individuals within a subtype, whereas the individualized prediction approach allows for estimation of differential treatment effects at the individual level. The use of individual-level data means that the individualized prediction approach will almost certainly provide more precise estimates of treatment response, and thus more accurately guide optimal treatment selection, than approaches that lose information by classifying individuals into subgroups (42). The same principles will apply to prediction of any other outcome, for example, predicting disease progression or development of microvascular and macrovascular complications.

Evaluating Performance of Strategies for Selecting Optimal Treatment

Our group has recently applied a novel framework to evaluate treatment selection models in type 2 diabetes. Novel approaches are required in this context; conventional measures of prediction model performance are of
limited utility when evaluating treatment selection models (13), as the focus is not the overall ability of a model to predict response but rather accurate identification of treatment-by-covariate interactions that predict differences in response between treatments. At the individual level these differences are unobservable (13), as at one point in time the response of a person to multiple different therapies cannot simultaneously be evaluated.

Our framework was applied to test head-to-head the Ahlqvist clusters strategy against an individualized prediction strategy for selecting optimal treatment, in post hoc analysis of individual level data from two large clinical trials (A Diabetes Outcome Progression Trial [ADOPT] and Rosiglitazone Evaluated for Cardiac Outcomes and Regulation of Glycaemia in Diabetes [RECORD]; n = 8,798) (43–45). This was important, as a key discussion point raised in the Ahlqvist et al. study was that the clusters identified could be used to “guide therapy” (34). In both trials, participants were randomized to either SU, TZD, or metformin treatment. The same five subtypes proposed using the Scandinavian data were reproduced in ADOPT using the same data-driven cluster analysis approach.
(34,46). Then, within each subtype, average glycemic response for each of the three treatments was estimated, and the treatment associated with the greatest average glycemic response was allocated as the optimal treatment for all people assigned to that subtype. The utility of the subtypes was compared with an individualized prediction strategy that assigned optimal treatment on an individual rather than subtype level, using a model that estimated response for each drug for each participant based on their specific features. Notably, only the simple routine clinical features (sex, and BMI, HbA1c, and age at diagnosis as continuous markers) were used for the individualized prediction model; two features used to inform the cluster analysis, HOMA-IR and HOMA-B (respectively, measures of insulin resistance and insulin secretion), were not included, as they are not routinely available in clinical practice.

Despite including only simple markers, the individualized prediction strategy markedly outperformed the subtypes strategy in the external validation trial data set (RECORD trial; \(n = 4,057\)) (Fig. 7) (43). For each strategy, the approach used was to define two subgroups of participants: 1) a concordant subgroup whose randomized treatment was the same as their predicted optimal treatment and 2) a discordant subgroup whose randomized treatment differed from their predicted optimal treatment (47). The difference between the concordant versus discordant subgroups was then contrasted for each strategy, with a bigger difference indicating a more useful treatment selection strategy. Where external test data sets are available, this evaluative framework represents a novel and cost-effective means of evaluating the utility of treatment selection models, whether on their own or in head-to-head comparison, and can be applied for other outcomes as well as treatment response.

**Future Directions: “omics” and Beyond HbA1c**

While this Perspective has focused only on glycemic response to diabetes treatment, the approaches outlined can easily be extended to nonglycemic end points including microvascular and macrovascular complications. The ideal precision medicine approach in type 2 diabetes will maximize therapeutic benefit while limiting risks (48), which will also require evaluation of HTE for side effects, glycemic progression, and risk of microvascular or macrovascular complications. Particular subgroups at higher risk of common treatment-specific side effects are already established for several drug classes; for example, the risk of fracture with TZD is limited mainly to females (49), and with SGLT2i females and those with a history of prior infection are at greatly increased risk of genital infections (50). Methods to overcome unmeasured confounding, such as the prior event rate ratio, may have particular utility for evaluating side-effect risk in observational routine care data where allocation to therapy is not randomized (51,52). A related but overlooked question for precision medicine, with great clinical relevance, is whether the benefits and risks of a treatment are positively associated. This is likely the case for TZD; the risk of edema and likelihood of weight gain increase with greater glucose-lowering response (21,53), and this should be an important consideration when choosing treatment. A further extension of the current work would be evaluation of effects of higher-order drug combinations. This will be possible in large routine clinical data sets where substantial numbers of patients are on specific combination therapies, although robust validation approaches will be required.

A key question is how genetics can inform precision medicine in type 2 diabetes. Proposed genetically defined type 2 diabetes subtypes reflect and help to understand
underlying pathophysiology (40,41). The clear advantage of using genetics is stability, as subtypes defined solely by genetics will be constant throughout life. At the moment it is unknown whether the continuous polygenic scores underlying genetic subtypes can improve prediction models that are based solely on routine clinical features and biomarkers. For treatment response, individual genetic markers have shown differences for specific treatments and may be of clinical utility when genetic information is routinely available in the medical records (54,55). If clinically relevant benefit can be demonstrated for polygenic scores and implementation is cost-effective, such scores can similarly be integrated into models based on routine clinical features.

A further exciting opportunity is the application of causal inference, data-driven machine learning, and artificial intelligence–based approaches to improve HTE prediction accuracy and generalizability of findings from large data sources such as electronic health records. Data-driven approaches may be of particular utility when databases start to incorporate high-dimensional genetic information (56). One possibility is that individualized prediction models developed with standard statistical methods based on classical risk factors could be augmented with data-driven classification approaches, if data-driven approaches are able to improve prediction by capturing higher-order complex traits missed by the standard methods. Although existing data can be used to develop and test candidate type 2 diabetes precision medicine approaches, ultimately, prospective trials, as done in cancer and monogenic diabetes (4,57,58), will likely be needed to demonstrate clinical utility. TriMaster, an ongoing three-way crossover randomized trial due to report in May 2021, is one such study in type 2 diabetes (NCT02653209). TriMaster will directly test the hypotheses that simple subgroups defined by baseline BMI and eGFR alter response with DPP-4i, SGLT2i, and TZD treatment (59). Not only will this provide the first prospective randomized evaluation of a precision medicine approach for glycemic response, the three-way crossover design will allow an “n of 1” analysis of patient preferences regarding the three treatments when they are tried in randomized order in blinded conditions. However, running prospective trials to test potential candidate factors one at a time for personalization is not a feasible, cost-effective, or efficient strategy. Future trials could instead test specific precision medicine algorithms based on multiple factors (potentially both clinical and genetic features), to test whether use of an algorithm results in improved outcomes for patients. One simple trial design for this would be to cluster randomize health centers (e.g., general practitioner practices in the U.K.) to either receive or not receive an algorithm—comparing centers with and without the algorithm would enable evaluation of its effectiveness and efficacy. If two competing algorithms or strategies need to be tested, this could be done using three-way cluster randomization.

A final key challenge is implementation of algorithms, which, to ensure patient benefit, should be not only effective but transparent, reproducible, and ethically sound (60) and which should be equally and freely accessible to all health professionals and patients. A type 2 diabetes treatment selection model would likely be most appropriately positioned within clinical practice software systems, so that it can be automatically populated with relevant clinical information from the electronic health record and function as a decision aid at the point of care. Development of software infrastructure that can utilize routinely collected health records to support delivery of such probabilistic algorithms will be required before precision medicine can truly become a reality for common diseases such as type 2 diabetes.

CONCLUSIONS

Recent demonstration of robust, clinically relevant differences in glycemic response suggest that a precision medicine approach to selecting optimal type 2 diabetes treatment will soon be possible. The most practical way to implement this in the near future will be to focus on routine clinical markers, and the most accurate approach will be integration of continuous features into individualized, probabilistic prediction models that can be deployed at the point a decision to escalate treatment is made, rather than subtyping. Estimates of differences in treatment response can augment the limited existing stratification of people with type 2 diabetes based on cardiovascular and renal comorbidity and will be applicable to everyone requiring glucose-lowering treatment. For people for whom differences in response between treatments are modest, this information is still important, as it can facilitate selection of treatment based on other criteria. A framework of discovery in routine data, followed by replication and testing in existing clinical trial data sets, offers a low-cost and principled way to evaluate the potential of precision medicine, applicable to other chronic diseases in addition to type 2 diabetes.

Acknowledgments. The author is grateful to colleagues Dr. Beverley Shields, Dr. Angus Jones, Professor Andrew Hattersley, Dr. Andrew McGovern, and Dr. Nick Thomas for insightful discussions and recommendations when preparing this article. The author also thanks the MASTERMIND (MRC APBI Stratification and Extreme Response Mechanism IN Diabetes) consortium who have contributed to many of the studies discussed in this article. Parts of this work were presented at the 80th Scientific Sessions of the American Diabetes Association, 12–16 June 2020.
References


31. Goldenberg RM. Choosing dipeptidyl peptidase-4 inhibitors, sodium-glucose co-transporter-2 inhibitors, or both, as add-ons to metformin: patient baseline characteristics are crucial. Clin Ther 2017;39:2438–2447


43. 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;7:442–451