MAYO CLINIC

ORIGINAL ARTICLE

An Electronic Health Record–Compatible Model to Predict Personalized Treatment Effects From the Diabetes Prevention Program: A Cross-Evidence Synthesis Approach Using Clinical Trial and Real-World Data

David M. Kent, MD, MS; Jason Nelson, MPH; Anastassios Pittas, MD, MS; Francis Colangelo, MD, MS; Carolyn Koenig, MD; David van Klaveren, PhD; Elizabeth Ciemins, PhD; and John Cuddeback, MD, PhD

Abstract

Objective: To develop an electronic health record (EHR)-based risk tool that provides point-of-care estimates of diabetes risk to support targeting interventions to patients most likely to benefit.

Patients and Methods: A risk prediction model was developed and validated in a large observational database of patients with an index visit date between January 1, 2012, and December 31, 2016, with treatment effect estimates from risk-based reanalysis of clinical trial data. The risk model development cohort included 1.1 million patients with prediabetes from the OptumLabs Data Warehouse (OLDW); the validation cohort included a distinct sample of 1.1 million patients in OLDW. The randomly assigned clinical trial cohort included 3081 people from the Diabetes Prevention Program (DPP) study.

Results: Eleven variables reliably obtainable from the EHR were used to predict diabetes risk. This model validated well in the OLDW (C statistic = 0.76; observed 3-year diabetes rate was 1.8% (95% confidence interval [CI], 1.7 to 1.9) in the lowest-risk quarter and 19.6% (19.4 to 19.8) in the highest-risk quarter). In the DPP, the hazard ratio (HR) for lifestyle modification was constant across all levels of risk (HR, 0.43; 95% CI, 0.35 to 0.53), whereas the HR for metformin was highly risk dependent (HR, 1.1; 95% CI, 0.61 to 2.0 in the lowest-risk quarter vs HR, 0.45; 95% CI, 0.35 to 0.59 in the highest-risk quarter). Fifty-three percent of the benefits of population-wide dissemination of the DPP lifestyle modification and 73% of the benefits of population-wide metformin therapy can be obtained by targeting the highest-risk quarter of patients.

Conclusion: The Tufts–Predictive Analytics and Comparative Effectiveness DPP Risk model is an EHR-compatible tool that might support targeted diabetes prevention to more efficiently realize the benefits of the DPP interventions.

 ■ Mayo Clin Proc. 2022;97(4):703-715

The Diabetes Prevention Program (DPP) Study showed that either an intensive program of lifestyle modification or pharmacotherapy with metformin substantially reduced the risk for developing type 2 diabetes in patients at high risk, compared with "usual care."[1] The findings have broad implications because "prediabetes" affects approximately 88 million adults in the United States.[2]

Strenuous calls to address the epidemic of diabetes with prevention[3,4] have been counterbalanced by concerns about the overmedicalization of prediabetes.[5] Almost 2 decades after publication of the DPP Study, it remains unclear how best to implement these

From the Predictive Analytics and Comparative Effectiveness Center (D.M.K., J.N., D.v.K.) and Division of Endocrinology

Affiliations continued at the end of this article.

interventions in such an overwhelmingly large, and mostly undiagnosed, population. A 2015 study examining a national sample of more than 17,000 working-age adults with prediabetes found that only 3.7% were receiving metformin.[6] Similarly, widespread use of the intensive lifestyle intervention remains largely unrealized despite evidence that rigorous diet and physical activity promotion reduces diabetes risk in the community setting.[7]

However, prediabetes is itself a heterogeneous condition. We previously showed that even among patients enrolled in the DPP Study itself, the risk for developing diabetes within 3 years varies widely and is highly skewed.[8] Some trial participants were estimated to have a 1% to 2% risk; others, 90%. Unsurprisingly, the degree of benefit from metformin therapy or from the lifestyle intervention was also distributed unevenly.

This prior proof-of-concept work had several limitations. Notably, the risk distribution within the DPP trial participants may differ from that of patients seen in routine practice, particularly since the American Diabetes Association (ADA) has subsequently broadened its definition of prediabetes to include a still more heterogeneous population.[9] Further, the application of prediction methods to data routinely collected in the electronic health record (EHR) provides a promising means to overcome some of the major barriers to the use of risk models.[10,11] For example, in addition to requiring manual ascertainment of variables, the previously reported DPP-based model required waist circumference and waist to hip ratio measurements that are not difficult to ascertain in routine practice. We describe the development of a clinical prediction model, the Tufts—Predictive Analytics and Comparative Effectiveness DPP risk model, using a hybrid approach that makes use of routinely collected EHR data to predict the risk for diabetes onset and clinical trial data to estimate unbiased risk-based effects of preventive interventions.

PATIENTS AND METHODS

Overview

We sought to develop and validate a diabetes risk prediction model using data elements readily available in the EHR for dissemination across health care systems as an EHR-embedded tool, to facilitate ease of use. The tool provides clinicians and their patients with an individualized risk for developing diabetes and the estimated benefit of applying a DPP treatment strategy, either an intensive lifestyle program or pharmacotherapy with metformin (the combination of both was not tested in the DPP Study).

Data Sources and Participants

The model was developed and validated using EHR data from the OptumLabs Data Warehouse (OLDW). The OptumLabs EHR database is a geographically diverse sample of the US population with longitudinal clinical data on more than 33 million lives with at least 1 clinic visit during the study period. Using a retrospective observational cohort design, we geographically stratified the database by US Census Region into a development cohort of 1,076,983 patients (Northeast, South, and West) and a separate validation cohort of 1,075,833 patients (Midwest).

Eligibility criteria included age between 25 and 75 years on an "index" office or clinic encounter (index visit defined by *Current Procedural Terminology*/Healthcare Common Procedure Coding System codes; Supplemental Table 1, available online at http://www.mayoclinicproceedings.org) between January 1, 2012, and December 31, 2016, at which time they met laboratory-based criteria for the diagnosis of prediabetes. (The age enrollment was selected because it approximated the age distribution of the DPP trial, which enrolled patients $\geq$ 25 years.) Prediabetes was defined by current ADA criteria, that is, having no diagnosis of type 1 or type 2 diabetes on the problem list and one of the following within 12 months before the visit: hemoglobin A_{1c} (HbA_{1c}) level between 5.7% and 6.4% inclusive and/or fasting glucose (FG) level between 100 and 125 mg/dL (to convert to mmol/L, multiply by 0.0555) inclusive. Because labeling of fasting status may be incomplete, a glucose level drawn at the same time as a lipid panel or triglycerides was considered as fasting. We did not use the

Mayo Clin Proc. ■ April 2022;97(4):703-715 ■ https://doi.org/10.1016/j.mayocp.2021.09.012
www.mayoclinicproceedings.org

2-hour post–glucose load criterion because it is rarely used in clinical practice for prediabetes. Patients were excluded if they had random (nonfasting) glucose levels of 200 mg/dL or greater on 2 occasions within a 3-month period before the index visit. Women with documented pregnancy within 24 months of the index visit were also excluded. To ascertain the development of diabetes, patients also had to have some clinical activity 3 years after the index visit. Eligibility criteria are detailed in Supplemental Table 1.

The DPP data set was used to estimate treatment effect for metformin or the intensive lifestyle modification program. The design, rationale, outcomes, and loss to follow-up of the DPP have been described in detail elsewhere.[1,12]

Briefly, inclusion criteria included a body mass index (BMI; calculated as the weight in kilograms divided by the height in meters squared) of 24 or higher ($\geq$22 kg/m^2 in Asians) and a plasma FG concentration of 95 to 125 mg/dL inclusive (impaired FG) and a concentration of 140 to 199 mg/dL inclusive 2 hours after a 75-g oral glucose load (impaired glucose tolerance). We note that these criteria differ from the ADA's current diagnostic criteria for prediabetes that we used for the OLDW model; the ADA definition imposes no BMI requirement.[13]

The DPP participants were randomly assigned to: (1) standard lifestyle recommendations plus 850 mg of metformin twice daily, (2) an intensive program of lifestyle modification that included 16 lessons with a case manager and set goals of at least a 7% weight loss and at least 150 minutes of physical activity per week, or (3) standard lifestyle recommendations plus placebo twice daily. After a median follow-up period of 2.8 (range, 1.8-4.6) years, progression to diabetes was reduced by 58% (95% CI, 47% to 66%) in the lifestyle modification arm and 31% (17% to 43%) in the metformin arm, both compared with the placebo arm.[1] The National Institute of Diabetes and Digestive and Kidney Diseases Data Repository, from which we obtained data, includes 3081 of the 3234 DPP participants (95% of full population) because some local institutional review boards (IRBs) declined to participate in data distribution.

Outcome

For the OLDW cohort, the time-to-event outcome was defined as the time to the first patient encounter after the index visit with documented evidence of type 2 diabetes by any of the following criteria,[14] diagnosis codes *International Classification of Diseases, Ninth Revision* 250.x0 or 250.x2 or *International Classification of Diseases, Tenth Revision* E11.xx, pharmacotherapy or procedure for type 2 diabetes (as detailed in Supplemental Table 1), HbA_{1c} level greater than 6.4%, FG (or presumed fasting, as noted) level greater than 125 mg/dL, or 2-hour oral glucose tolerance test postload glucose level greater than 199 mg/dL. Laboratory-based criteria required confirmation by an additional laboratory in the diabetes range or by another method (ie, diagnosis or medication). Follow-up time for patients who did not meet the outcome definition was censored at the first occurrence of the last observed encounter or end of study period.

Candidate Predictors

A priori risk model predictors were identified by a systematic review conducted by Collins et al.[15] We selected the following 11 independent variables that were included in at least 3 prior diabetes risk models and were judged to be easily and reliably obtainable in EHR data: age, sex, race, smoking status, BMI, presence or absence of a diagnosis of hypertension, systolic blood pressure, high-density lipoprotein cholesterol level, triglyceride level, FG level, and HbA_{1c} level. Four variables included in 3 prior models were not considered based on the difficulty of ascertaining them in EHR data: physical activity, waist circumference, waist to hip ratio, and family history of diabetes.

Missing Data

Missing data is a common limitation when working with EHR data.[16] Although multiple imputation may improve estimates of parameter effects under a missing-at-random assumption, it does not provide a practical

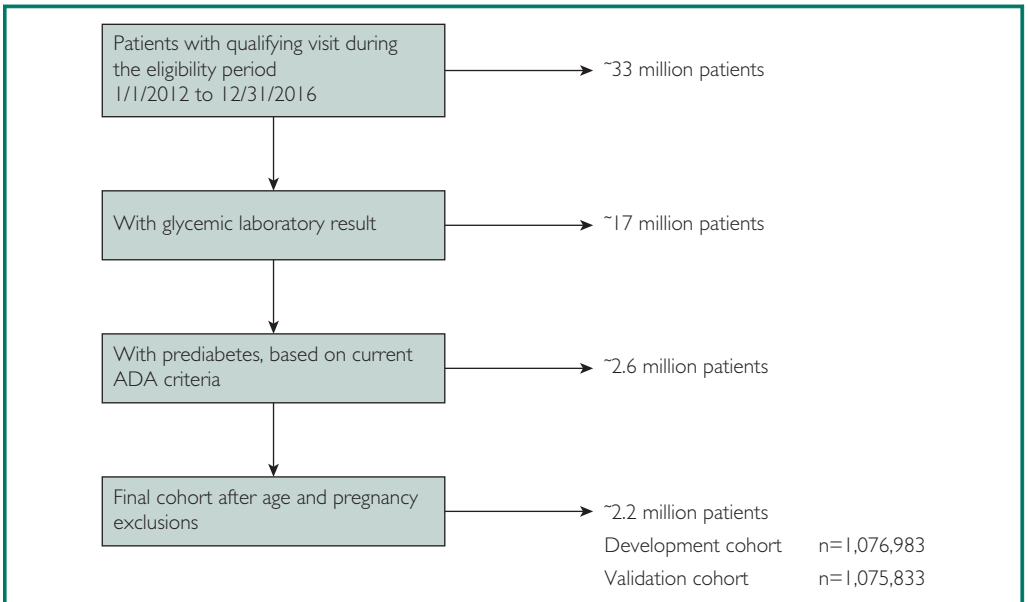

FIGURE 1. Consolidated Standards of Reporting Trials diagram for OptumLabs Data Warehouse derivation and validation cohort. The figure depicts the number of eligible patients as each inclusion criterion is applied. Starting with an initial population size of more than 33 million patients, our final study cohort consists of 2,152,816 eligible patients.

means to cope with missingness in actual patients for whom a prediction needs to be made. Thus, we used missing indicator variables to capture the predictive effects of missingness under the assumption that future and prior missingness are similarly informative. For each predictor, an additional dichotomous variable indicated the presence of missing values.[17,18] For continuous variables (eg, BMI and HbA_{1c} level), the missing value of the original variable was replaced by a fixed constant (the median) before model estimation, and the missing indicator variable appropriately adjusted for the "missing variable effect." For categorical variables (eg, race and smoking status), an additional level was added to define the missing category.

Model Development

We used multivariable Cox proportional hazards regression to estimate the predicted probability of developing type 2 diabetes. We included 2 a priori interactions, race × BMI and race × HbA_{1c} level, based on clinical judgment and the literature.[19,20] Model performance was assessed for discrimination and calibration. A bootstrap resampling procedure with 500 samples was used to internally validate the model, estimate optimism-corrected discrimination, and assess calibration.

Model Validation

Using the equation derived in the development cohort, we calculated the predicted probability of developing type 2 diabetes for patients in the validation cohort. Model performance on external validation was assessed for discrimination using Harrell's measure of concordance for censored response variable and calibration.[21]

Estimating Risk-Specific Treatment Effects

To estimate the risk-based treatment effect for metformin pharmacotherapy or the DPP lifestyle modification, we performed a risk-based heterogeneity of treatment effect analysis on the DPP.[22] The applicability of the OLDW model to the DPP data was anticipated to be limited by differences between predictor variable definitions and measurement within a trial context vs EHR data, differences in the pattern of missingness

Mayo Clin Proc. ■ April 2022;97(4):703-715 ■ https://doi.org/10.1016/j.mayocp.2021.09.012
www.mayoclinicproceedings.org

TABLE 1. Cohort Characteristics[a,b]

		Overall	OLDW		
			Development	Validation	DPP
	Missing, %	n=2,152,816	n=1,076,983	n=1,075,833	n=3081
Age (y), mean ± SD	0.0	54.9±11.7	55.1±11.9	55±11.5	50.6±9.0
Female sex, %	0.1	50.3	51.3	49.1	66.6
Race, %	8.2				
White		86.5	84.3	88.9	57.4
Black		10.2	10.8	9.1	20.9
Other non-White race (Optum = Asian)		3.4	4.9	1.9	5.2
Smoking status, %	15.6				
Current smoker		23.3	20.0	26.4	9.0
Never smoked		48.0	53.2	42.9	35.2
Former smoker		28.8	26.8	30.7	55.8
Height (cm), mean ± SD	15.9	170.1±10.1	169.5±10.1	170.7±10	166.8±9.2
Body mass index (kg/m^2), mean ± SD	12.2	31.1±7	30.8±6.7	31.8±6.9	33.5±5.8
Diagnosis of hypertension, %	0	44.5	44.4	45.0	27.1
Systolic blood pressure (mm Hg), mean ± SD	9.0	127.4±14.9	127.6±15.2	127.3±14.7	124.2±14.7
HDL cholesterol (mg/dL), mean ± SD	12.3	50.9±14.7	51.3±14.9	50.6±14.5	45.6±11.8
Triglycerides (mg/dL), mean ± SD	12.6	138.3±72.8	136.9±72.8	139.7±72.7	162.9±93.5
Hemoglobin A_{1c} (%), mean ± SD	54.7	5.8±0.3	5.8±0.3	5.8±0.3	5.9±0.5
Fasting plasma glucose (mg/dL), mean ± SD	3.8	103.7±10.8	103±11.1	104.5±10.4	107.2±7.7
Fasting plasma glucose, (fasting) (mg/dL), mean ± SD	86.3	103.3±9.2	101.3±10.5	105.3±7.3	
Fasting plasma glucose (random) (mg/dL), mean ± SD	13.0	103.7±11.4	103.1±11.4	104.4±11.2	

[a]DPP = Diabetes Prevention Program; OLDW = OptumLabs Data Warehouse; SD = standard deviation.

[b]SI conversion factors: To convert HDL cholesterol values to mmol/L, multiply by 0.0259; to convert glucose values to mmol/L, multiply by 0.0555; to convert triglyceride values to mmol/L, multiply by 0.0113.

between these contexts (ie, there was essentially no data missingness in the DPP), differences in patient enrollment in the 2 settings, and differences in outcome definition and ascertainment.[23] Thus, we refit the OLDW model to the DPP, using the same variables and interaction terms. Consistent with methodological recommendations,[24,25] all 3 DPP arms were used because research has shown that overfitting to a control arm can induce spurious heterogeneity of treatment effects.[26-28] The treatment effect was then estimated by incorporating this linear predictor into a Cox proportional hazards model with the following terms: treatment (metformin or DPP lifestyle modification), the linear predictor of risk from the refitted model, and (potentially) an interaction between these to account for important changes in relative risk reduction across different levels of baseline risk. Based on a previous analysis,[8] we anticipated a risk-by-treatment interaction with metformin pharmacotherapy and a consistent relative effect with the DPP lifestyle modification, but we examined interactions for both treatment arms. We also performed a sensitivity analysis, examining the risk-by-treatment interactions, stratifying the DPP by the OLDW model without any refitting, and examining the distribution of predicted effects using this model.

Incorporation of Decision Support in EHR

To facilitate use in clinical decision making, based on patient and provider focus groups and interviews, we implemented the model in 2 different ways: (1) a hard coded calculation in an Allscripts EHR, and (2) a cloud-hosted SMART on FHIR[29] app that can be incorporated into any EHR, leveraging interoperability standards recently promulgated

TABLE 2. Final Model for Incident Diabetes[a,b,c]

	Hazard Ratio	Lower	Upper		Hazard Ratio	Lower	Upper
Age, per 10 y	1.08	1.08	1.08	Adjustments for missing data			
Female sex	1.21	1.19	1.23	Race (missing) vs White	0.16	0.07	0.38
Black vs White	2.73	1.32	5.64	Smoking (missing) vs never	1.08	1.06	1.11
Asian vs White	0.01	0.00	0.02	HbA_{1c} (missing)	0.75	0.74	0.77
Current smoker vs never	1.22	1.19	1.24	Fasting plasma glucose (missing)	1.03	0.99	1.07
Former smoker vs never	1.11	1.09	1.13	Triglycerides (missing)	1.08	1.03	1.12
Hypertension	1.23	1.21	1.25	BMI (missing)	1.22	1.18	1.26
HbA_{1c}, per 0.1%	1.24	1.18	1.29	Systolic blood pressure (missing)	1.22	1.17	1.26
Fasting plasma glucose, per 10 mg/dL	1.29	1.29	1.29	HDL cholesterol (missing)	1.23	1.17	1.28
Triglycerides, per 10 mg/dL	1.01	1.01	1.02	AA × BMI (missing)	0.97	0.91	1.03
BMI, per 5 units	1.24	1.24	1.24	AA × HbA_{1c} (missing)	1.56	1.48	1.64
Systolic blood pressure, per 20 mm Hg	1.05	1.05	1.05	Asian × BMI (missing)	0.77	0.70	0.84
HDL cholesterol, per 10 mg/dL	0.85	0.85	0.85	Asian × HbA_{1c} (missing)	2.03	1.85	2.23
Black × BMI	0.98	0.98	0.99	Race (missing) × BMI	0.99	0.99	1.00
Black × HbA_{1c}	0.95	0.84	1.07	Race (missing) × BMI (missing)	0.82	0.77	0.87
Asian × BMI	1.00	0.99	1.01	Race (missing) × HbA_{1c}	1.40	1.22	1.61
Asian × HbA_{1c}	2.32	1.91	2.83	Race (missing) × HbA_{1c} (missing)	1.46	1.37	1.55

[a]AA = Black/African American; BMI = body mass index; HbA1c = hemoglobin A1c; HDL = high-density lipoprotein.
[b]SI conversion factors: To convert HDL cholesterol values to mmol/L, multiply by 0.0259; to convert glucose values to mmol/L, multiply by 0.0555; to convert triglyceride values to mmol/L, multiply by 0.0113.
[c]Baseline hazard at 5 years (S0): 1 year = 0.02470, 2 years = 0.04757, 3 years = 0.07044.

by the US office of the National Coordinator of Health Information.

IRB Approval

This study was reviewed and approved by the Tufts Health Sciences IRB before accessing the deidentified data from the DPP and OLDW data sets.

RESULTS

Figure 1 shows the development of the derivation and validation OLDW data sets. Approximately 1.1 million people with prediabetes from the Northeast, South, and West were included in the derivation cohort, and a similar number from the Midwest were included in the validation cohort. Characteristics of these cohorts are shown in Table 1.

Model Development and Validation: Risk Stratification

The coefficients for each of the variable and interaction terms included in the model are shown in Table 2. The optimism-corrected C statistic on the derivation sample was 0.73. When the model was tested on the validation cohort, the C statistic was slightly higher at 0.76. Calibration on the validation cohort was very good (Figure 2). Harrell's E statistic was 1.63% and the calibration intercept and slope were −0.27 and 1.12, respectively. Among the 268,959 patients in the lowest-risk quartile, the predicted diabetes rate was 3.1% (95% CI, 3.0% to 3.2%), while the observed rate was 1.8% (95% CI, 1.7% to 1.9%); among the 268,958 patients in the highest-risk quartile, the predicted diabetes rate was 19.2% (95% CI, 18.6% to 19.9%), while the observed rate was 19.6% (95% CI, 19.4% to 19.8%).

Calculation of Relative Treatment Effects in the DPP Study

Prior work demonstrated a consistent relative treatment effect across risk groups with the DPP lifestyle modification and an increasing relative effect with progressively higher risk for metformin pharmacotherapy.[8] Using the OLDW model refit to the DPP data (Supplemental Table 2, available online at

Mayo Clin Proc. ■ April 2022;97(4):703-715 ■ https://doi.org/10.1016/j.mayocp.2021.09.012
www.mayoclinicproceedings.org

http://www.mayoclinicproceedings.org; C statistic, 0.719), we confirmed the absence of a treatment-by-risk interaction for lifestyle modification (*P* for interaction = .68); thus, we applied a constant relative risk reduction in the prediction model (hazard ratio [HR], 0.43; 95% CI, 0.35 to 0.53) to estimate the diabetes outcome with lifestyle modification. We also confirmed the presence of a treatment-by-risk interaction with metformin pharmacotherapy (*P* for interaction = .003; using the continuous risk on the logit scale): low-risk patients had outcomes with metformin that were similar to usual care (in lowest-risk quarter, observed HR, 1.1; 95% CI, 0.61 to 2.0), and high-risk patients have outcomes with metformin that were similar to the DPP lifestyle modification (in highest-risk quarter, observed HR, 0.45; 95% CI, 0.35 to 0.59).

Figure 3 shows observed and predicted benefits across quartiles for the DPP for both lifestyle and metformin therapy. A look-up table showing the relative risk reduction with metformin for each level of risk is shown in Supplemental Table 3 (available online at http://www.mayoclinicproceedings.org), truncated at a low value of 0% (no harm or benefit) and a high value of 60%.

Distribution of Risks and Benefits in OLDW

The overall average 3-year predicted risk for developing diabetes for patients in the validation OLDW cohort was 9.0%, 3.9%, and 6.0% with usual care, the DPP lifestyle diabetes, and metformin therapy, respectively. For lifestyle modification, 53% of the total preventable cases of diabetes could be prevented by treating the 25% of patients at highest risk; 76%, by treating the 50% at highest risk; and 91%, by treating the 75% at highest risk. For metformin therapy, 73% of the total preventable cases could be prevented by treating the 25% of patients at highest risk; 93%, by treating the 50% at highest risk; and 100%, by treating the 75% at highest risk.

Sensitivity Analyses

Direct application of the OLDW model (not refit) on the DPP showed a moderately diminished discrimination (C statistic = 0.68). There was no risk-by treatment interaction with lifestyle (*P*=.69). The risk-by-treatment interaction with metformin therapy was qualitatively similar to that with the refit model (*P*=.08), and the distribution of predicted benefits with this model was also similar. For lifestyle modification, 53% of the total cases of preventable diabetes could be prevented by treating the 25% of patients at highest risk; 76%, by treating the 50% at highest risk. For metformin therapy, 65% of the total cases of preventable diabetes could be prevented by treating the 25% of patients at highest risk; 86%, by treating the 50% at highest risk.

Implementation of the Final Model

Figure 4 shows the user interface of the SMART app in an EHR. Predictions are generated automatically based on the data available and retrieved from the patient's record, using appropriate indicators in the model for missingness when necessary.

DISCUSSION

We present the Tufts—Predictive Analytics and Comparative Effectiveness DPP risk model, an EHR-compatible model that predicts diabetes onset based on 11 variables routinely collected in clinical practice. A major strength of the risk model is that it was derived on the OLDW, which reflects people with prediabetes defined by the most commonly used ADA criteria, from heterogeneous EHRs and more than 30 US health care systems. The risk model derived in 3 US Census regions performed very well in a geographically distinct cohort. Compatible risk-specific estimates of treatment effect were then obtained directly from the DPP. By prioritizing care based on the risk for diabetes, this "hybrid" model might help optimize the efficiency of diabetes prevention: treating just the highest-risk half of people with prediabetes would capture 77% of the benefit of population-wide lifestyle modification or 93% of the benefit of population-wide metformin pharmacotherapy. This is important because lifestyle programs are resource intensive and require a high level of commitment from the patient. Pharmacotherapy is not without adverse effects and

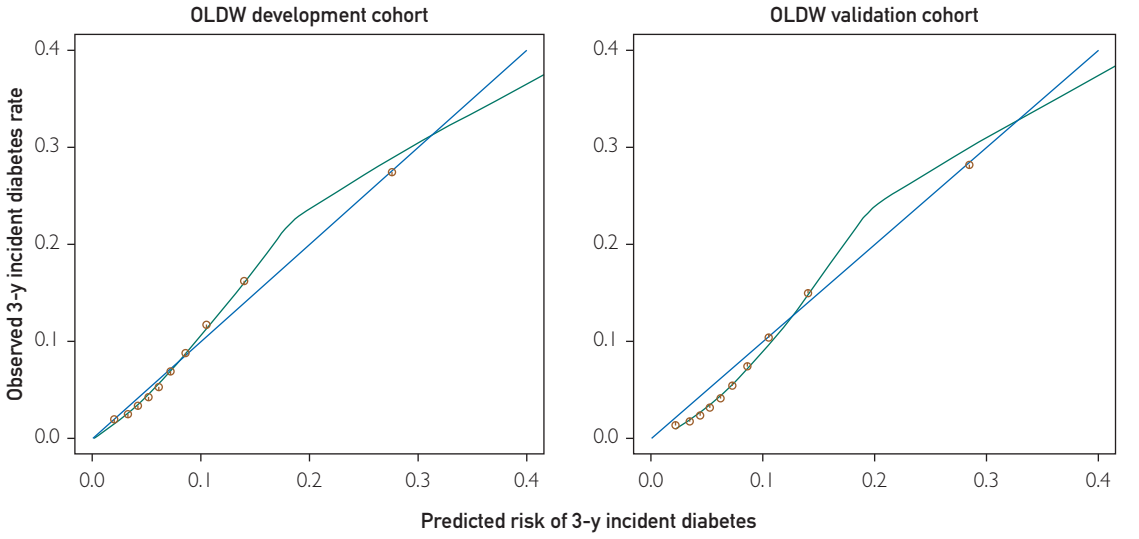

FIGURE 2. Calibration curves. The figure on the left depicts the observed vs predicted 3-year rate of developing diabetes in the 1,076,983 million patients in the derivation cohort (Northeast, South, and West regions) divided into equal-sized tenths. The figure on the right depicts the observed vs predicted 3-year rate of developing diabetes in the 1,075,833 million patients in the validation cohort (Midwest). OLDW = OptumLabs Data Warehouse.

overtreatment should be avoided, especially in low-risk patients who do not appear to benefit.

The issue of how to address prediabetes has grown in importance as broader diabetes screening has been recommended and promoted.[13,30] For every patient with diabetes identified, screening identifies 6 patients with prediabetes; health systems are thus confronted with a growing number of patients who have prediabetes, without the capacity to treat everybody, reserving limited resources to improving cardiometabolic control for patients with diabetes.

Although the ADA has lowered the HbA_{1c} and FG thresholds to define prediabetes,[9,31] some have argued that the value of medicalizing prediabetes and defining an ever-growing proportion of the population as diseased is of dubious value.[5] Most patients who are classified as prediabetic do not develop diabetes even in a decade, and risks for developing end-organ damage are low for those developing diabetes later in life.[32] Risk stratification offers an approach that promises more focused resources specifically on those who are likely to benefit. Although our prior research results provided proof of concept that risk stratification could support providers and health systems prioritize these patients,[8] the present EHR-compatible model is designed to be used at point of care, and it has been incorporated into the EHRs at several locations in the United States.

A long-standing concern regarding limitations of randomized clinical trial results is that they might not be applicable to "real-world" populations when there is nonrandom selection into the trial and treatment effects are heterogeneous.[33] Here, for example, we found that the real-world at-risk population was at substantially lower overall risk than patients included in the DPP and that treatment effects were risk dependent. The lower overall risk in the OLDW cohort is the result of multiple factors, including: (1) different inclusion criteria for the DPP (including a high BMI and elevated 2-hour glucose level after a 75-g glucose load), (2) differences in the distribution of risk variables (see Table 1), and (3) different outcome ascertainment, which is substantially more rigorous in the trial

Mayo Clin Proc. ■ April 2022;97(4):703-715 ■ https://doi.org/10.1016/j.mayocp.2021.09.012
www.mayoclinicproceedings.org

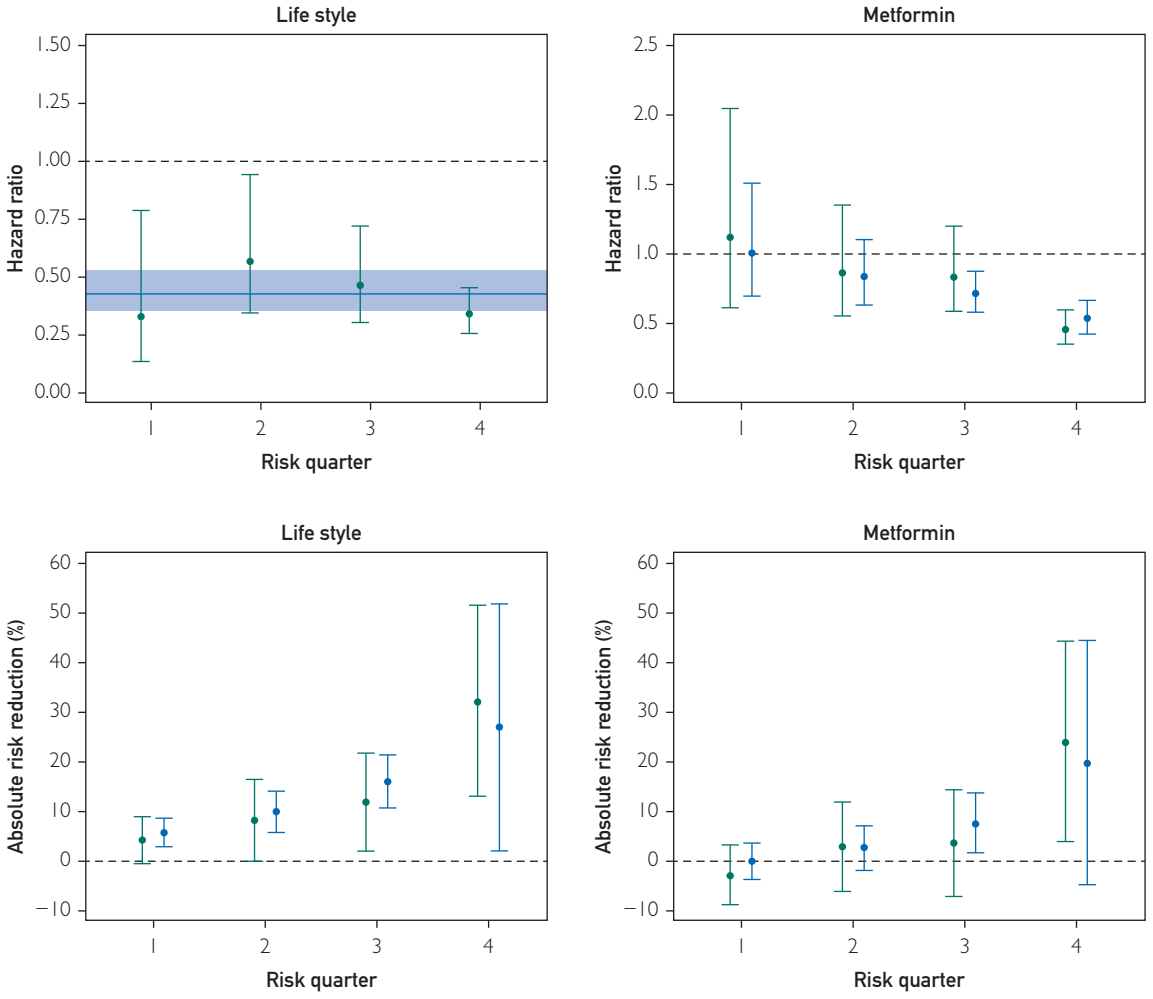

FIGURE 3. Observed and predicted treatment effects in the Diabetes Prevention Program (DPP) Study across risk groups. Green dot and bar (95% confidence interval) are observed treatment effect. Blue dot and bar are predicted treatment effect. Figure 3 depicts the observed treatment effects (green dots) in patients in the DPP Study when patients are stratified into quarters based on predicted risk for the DPP lifestyle modification intervention (left) and for metformin (right). Predicted effects across risk groups are shown in blue. The top set of graphs displays relative effects and shows a consistency of effects across risk groups for lifestyle modification but heterogeneous treatment effects for metformin (P=.003). The bottom graphs show effects on the absolute risk difference scale, which shows increasing benefits for higher-risk patients for both interventions.

setting. Cross-design synthesis has been proposed as a means of addressing the potential problems of external validity of trial evidence by combining the strengths of both designs; observational designs to capture the full range of patients and randomized trials for unbiased treatment effects.[34,35] Beyond differences in risk, a related concern is whether the relative effects seen in the trial would apply in routine clinical care, for which the patient/provider commitment may be less than ideal. We believe it is appropriate to provide these estimates in shared decision making because they estimate effects that patients should anticipate if they have trial-like adherence to the interventions.

Although several different methods for cross-design synthesis have been proposed,[36,37] all approaches depend on the ability to adjust results based on patient characteristics across designs. A seldom

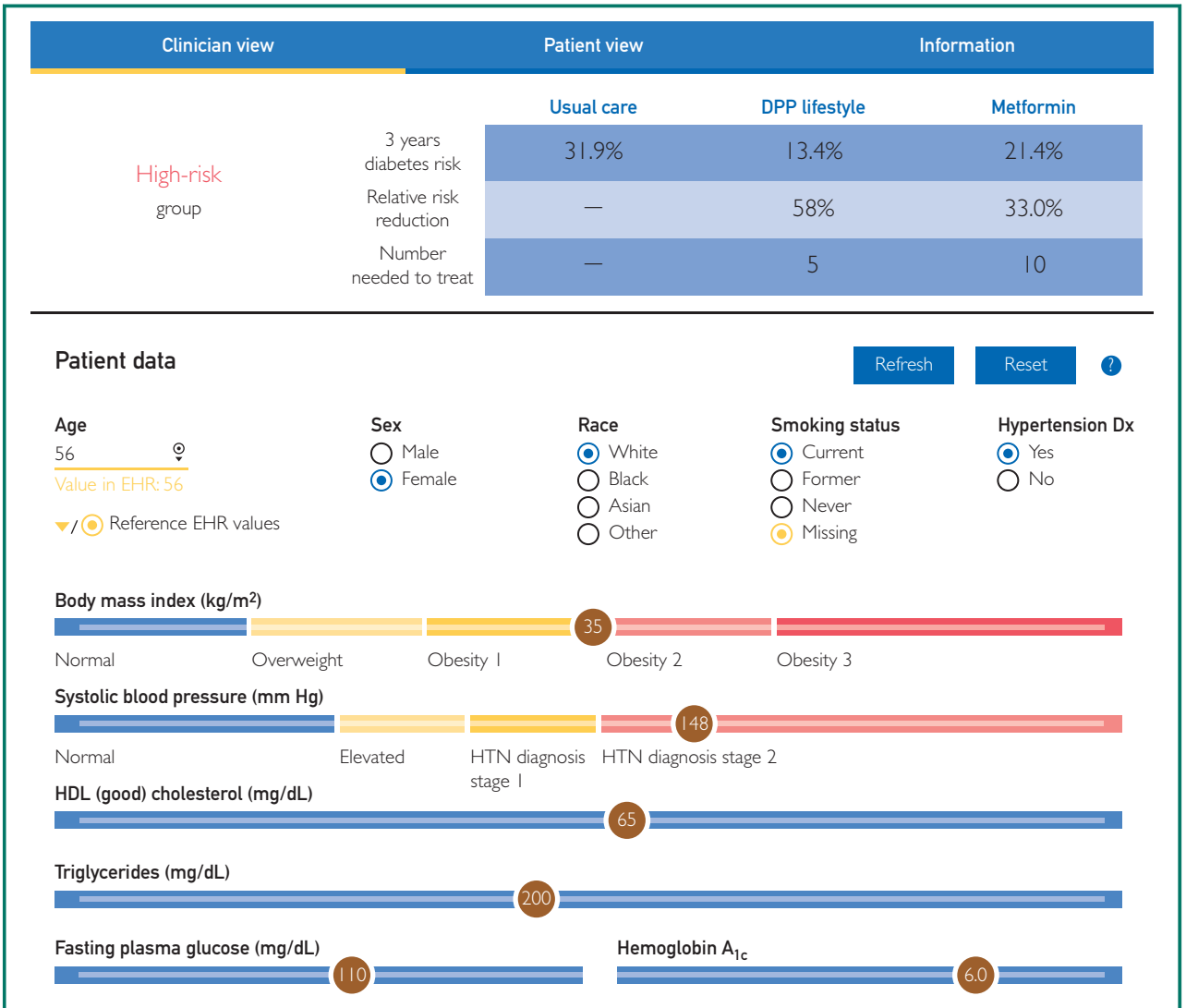

FIGURE 4. User interface for decision support tool. This figure depicts the interface of a clinical decision support tool currently implementing the Tufts–Predictive Analytics and Comparative Effectiveness Diabetes Prevention Program (DPP) risk model in clinical care. EHR = electronic health record; HDL = high-density lipoprotein; HTN = hypertension. SI conversion factors: To convert HDL cholesterol values to mmol/L, multiply by 0.0259; to convert glucose values to mmol/L, multiply by 0.0555; to convert triglyceride values to mmol/L, multiply by 0.0113.

discussed barrier is that variable definitions and ascertainment can differ considerably between clinical trial data and routinely collected observational data. Our approach was designed to address these barriers in a pragmatic way, by estimating risk-specific treatment effects in the clinical trial using the same set of variables as used in the observational risk model. This approach was driven in part by our novel aim, to predict effects in patients in clinical care based on individual patient characteristics, rather than estimating average treatment effects in broad target populations.

A related issue that has received limited attention is how to deploy clinical prediction models in an EHR. There is a proliferation of clinical prediction models; use of routinely collected EHR data to automatically generate individual patient predictions is an appealing approach to disseminate these into the clinic. However, most published clinical

Mayo Clin Proc. ■ April 2022;97(4):703-715 ■ https://doi.org/10.1016/j.mayocp.2021.09.012
www.mayoclinicproceedings.org

prediction models are developed on research cohorts or clinical trials. Predictor variables collected in a trial are not consistently and rigorously captured in the EHR. Recent work has highlighted that heterogeneity in predictor measurement across different settings can substantially degrade model performance.[23,38]

Finally, use of trial or registry data cannot yield a model robust to missing values in the EHR database used for clinical prediction because the pattern of missingness present across research and EHR environments is expected to differ. The usual approaches addressing potential bias arising from missingness (eg, multiple imputation) are not designed to cope with missingness in variables used to generate predictions. These issues guided our decision to derive separate models in the EHR and trial setting, using a common set of variables that were well ascertained in both settings.

There are some limitations. The methods we used for "cross-walking" between the 2 very different types of data (trial and EHR real-world data) potentially introduce estimation error. Ideally, individualized treatment effects would be estimated on databases that combine the advantages of these different data sources: unbiased effect estimates through randomization, meticulous outcome ascertainment, consistency of predictors across derivation and implementation populations, and large heterogeneous populations. Improving the quality of data collection in routine care and integrating randomized trials into routine care[39-41] may narrow the gap between trial and real-world data. That oral glucose tolerance testing was used both for entry criteria and end point ascertainment presumably in the DPP trial contributed to the higher risk in the DPP trial cohort compared with the OLDW cohort. Lorenzo et al[42] demonstrated that the sensitivity of HbA_{1c} and/or FG levels in the diagnosis of prediabetes and diabetes is relatively low (76% and 52%, respectively). Because the results of oral glucose tolerance testing are not generally available for most patients in routine care, we were unable to adjust for these differences. Conversely, incorporating the use of pharmacotherapy into our identification of the outcome may have caused some misclassification of patients without diabetes. However, we anticipate that this rate is very low. There were also other variables known to be predictors of diabetes onset (eg, waist to hip ratio) that are not well collected in routine care and so were not considered for our model. Finally, although the OLDW is representative of the commercially insured population, some caution is recommended in extrapolating the results beyond this.

Despite these limitations, we obtained qualitatively consistent risk-stratified results in the DPP regardless of which risk model was used: consistency of relative treatment effects of lifestyle modification across all levels of risk and heterogeneous relative treatment effects with metformin, with much stronger relative effects in higher-risk patients.

CONCLUSION

Although the number of people in the United States who have prediabetes and qualify for diabetes prevention programs could potentially overwhelm health care systems, these patients have substantial variation in their risk for developing diabetes and in their likelihood of benefiting from prevention therapies. Incorporation of a tool into the EHR to support automated risk stratification of patients in routine clinical care, by predicting individualized benefits, can support shared decision making and prioritize patients who are most likely to benefit, when capacity might be limited.

ACKNOWLEDGMENTS

Research reported in this publication was funded through a Patient-Centered Outcomes Research Institute award (DI-1604-35234). The statements in this publication are solely the responsibility of the authors and do not necessarily represent the views of the Patient-Centered Outcomes Research Institute, its Board of Governors, or Methodology Committee. A.G. Pittas is supported in part by generous donations to the Tupper Research Fund at Tufts Medical Center.

SUPPLEMENTAL ONLINE MATERIAL

Supplemental material can be found online at http://www.mayoclinicproceedings.org. Supplemental material attached to journal articles has not been edited, and the authors take responsibility for the accuracy of all data.

Abbreviations and Acronyms: **AA,** Black/African American; **ADA,** American Diabetes Association; **BMI,** body mass index; **DPP,** Diabetes Prevention Program; **EHR,** electronic health record; **FG,** fasting glucose; **HbA_{1c},** hemoglobin A_{1c}; **HDL,** high-density lipoprotein; **HR,** hazard ratio; **HTN,** hypertension; **IRB,** institutional review board; **OLDW,** OptumLabs Data Warehouse

Affiliations (Continued from the first page of this article.): (A.P.), Tufts Medical Center, Boston, MA; Allegheny Health Network, Pittsburgh, PA (F.C.); Mercy Health, St. Louis, MO (C.K.); Department of Public Health, Erasmus MC University Medical Center, Rotterdam, the Netherlands (D.v.K.); and American Medical Group Association, Alexandria, VA (E.C., J.C.).

Grant Support: This work was supported by a Patient-Centered Outcomes Research Institute contract (DI-1604-35234). The funder had no role in the conduct of this research or the decision to publish the results.

Potential Competing Interests: The authors report no competing interests.

Data Previously Presented: A preliminary report of this work was previously published on medRxiv on January 8, 2021. https://doi.org/10.1101/2021.01.06.21249334

Correspondence: Address to David M. Kent, MD, MS, Predictive Analytics and Comparative Effectiveness (PACE) Center, Institute for Clinical Research and Health Policy Studies, Tufts Medical Center, 800 Washington St, Box 63, Boston, MA 02111 (dkent1@tuftsmedicalcenter.org; Twitter: @Tufts_PACE).

ORCID

David M. Kent: https://orcid.org/0000-0002-9205-5070

REFERENCES

1. Knowler WC, Barrett-Connor E, Fowler SE, et al; Diabetes Prevention Program Research Group. Reduction in the incidence of type 2 diabetes with lifestyle intervention or metformin. *N Engl J Med*. 2002;346(6):393-403.
2. Centers for Disease Control and Prevention (CDC). Prediabetes - your chance to prevent type 2 diabetes. https://www.cdc.gov/diabetes/basics/prediabetes.html. Published 2020. Accessed October 16, 2020.
3. Chen L, Magliano DJ, Zimmet PZ. The worldwide epidemiology of type 2 diabetes mellitus–present and future perspectives. *Nat Rev Endocrinol*. 2011;8(4):228-236.
4. Herman WH, Zimmet P. Type 2 diabetes: an epidemic requiring global attention and urgent action. *Diabetes Care*. 2012;35(5):943-944.
5. Yudkin JS, Montori VM. The epidemic of pre-diabetes: the medicine and the politics. *BMJ*. 2014;349:g4485.
6. Moin T, Li J, Duru OK, et al. Metformin prescription for insured adults with prediabetes from 2010 to 2012: a retrospective cohort study. *Ann Intern Med*. 2015;162(8):542-548.
7. Balk EM, Earley A, Raman D, Avendano EA, Pittas AG, Remington PL. Combined diet and physical activity promotion programs to prevent type 2 diabetes among persons at increased risk: a systematic review for the Community Preventive Services Task Force. *Ann Intern Med*. 2015;163(6):437-451.
8. Sussman JB, Kent DM, Nelson JP, Hayward RA. Improving diabetes prevention with benefit based tailored treatment: risk based reanalysis of Diabetes Prevention Program. *BMJ*. 2015;350:h454.
9. American Diabetes Association. Diagnosis and classification of diabetes mellitus. *Diabetes Care*. 2010;33(suppl 1):S62-S69.
10. Watson J, Hutyra CA, Clancy SM, et al. Overcoming barriers to the adoption and implementation of predictive modeling and machine learning in clinical care: what can we learn from US academic medical centers? *JAMIA Open*. 2020;3(2):167-172.
11. Wallace E, Johansen ME. Clinical prediction rules: challenges, barriers, and promise. *Ann Fam Med*. 2018;16(5):390-392.
12. Diabetes Prevention Program Research Group. Design and methods for a clinical trial in the prevention of type 2 diabetes. *Diabetes Care*. 1999;22(4):623-634.
13. American Diabetes Association. Standards of medical care in diabetes–2014. *Diabetes Care*. 2014;37(suppl 1):S14-S80.
14. McCoy RG, Nori VS, Smith SA, Hane CA. Development and Validation of HealthImpact: An Incident Diabetes Prediction Model Based on Administrative Data. *Health Serv Res*. 2016;51(5):1896-1918.
15. Collins GS, Mallett S, Omar O, Yu LM. Developing risk prediction models for type 2 diabetes: a systematic review of methodology and reporting. *BMC Med*. 2011;9:103.
16. Wells BJ, Chagin KM, Nowacki AS, Kattan MW. Strategies for handling missing data in electronic health record derived data. *EGEMS (Wash DC)*. 2013;1(3):1035.
17. Sisk R, Lin L, Sperrin M, et al. Informative presence and observation in routine health data: a review of methodology for clinical risk prediction. *J Am Med Inform Assoc*. 2021;28(1):155-166.
18. Groenwold RHH. Informative missingness in electronic health record systems: the curse of knowing. *Diagn Progn Res*. 2020;4:8.
19. Beck RW, Riddlesworth TD, Ruedy K, et al; DIAMOND Study Group. Continuous glucose monitoring versus usual care in patients with type 2 diabetes receiving multiple daily insulin injections: a randomized trial. *Ann Intern Med*. 2017;167(6):365-374.
20. Zhu Y, Sidell MA, Arterburn D, et al. Racial/ethnic disparities in the prevalence of diabetes and prediabetes by BMI: Patient Outcomes Research To Advance Learning (PORTAL) multisite cohort of adults in the U.S. *Diabetes Care*. 2019;42(12):2211-2219.
21. Harrell FE. *Regression Modeling Strategies: With Applications to Linear Models, Logistic Regression, and Survival Analysis*. New York, NY: Springer-Verlag; 2001.
22. Kent DM, Steyerberg EW, van Klaveren D. Personalized evidence based medicine: predictive approaches to heterogeneous treatment effects. *BMJ*. 2018;363:k4245.
23. Luijken K, Groenwold RHH, Van Calster B, Steyerberg EW, van Smeden M. Impact of predictor measurement heterogeneity across settings on the performance of prediction models: a measurement error perspective. *Stat Med*. 2019;38(18):3444-3459.
24. Kent DM, Paulus JK, van Klaveren D, et al. The Predictive Approaches to Treatment Effect Heterogeneity (PATH) statement. *Ann Intern Med*. 2020;172(1):35-45.
25. Kent DM, van Klaveren D, Paulus JK, et al. The Predictive Approaches to Treatment Effect Heterogeneity (PATH) statement: explanation and elaboration. *Ann Intern Med*. 2020;172(1):W1-W25.
26. Abadie A, Chingos M, West M. Endogenous stratification in randomized experiments. *Rev Econ Stat*. 2018;100(4):567-580.

27. Burke JF, Hayward RA, Nelson JP, Kent DM. Using internally developed risk models to assess heterogeneity in treatment effects in clinical trials. *Circ Cardiovasc Qual Outcomes*. 2014;7(1):163-169.
28. van Klaveren D, Balan TA, Steyerberg EW, Kent DM. Models with interactions overestimated heterogeneity of treatment effects and were prone to treatment mistargeting. *J Clin Epidemiol*. 2019;114:72-83.
29. Mandel JC, Kreda DA, Mandl KD, Kohane IS, Ramoni RB. SMART on FHIR: a standards-based, interoperable apps platform for electronic health records. *J Am Med Inform Assoc*. 2016;23(5):899-908.
30. Siu AL; U S Preventive Services Task Force. Screening for abnormal blood glucose and type 2 diabetes mellitus: U.S. Preventive Services Task Force Recommendation Statement. *Ann Intern Med*. 2015;163(11):861-868.
31. Genuth S, Alberti KG, Bennett P, et al. Expert Committee on the Diagnosis and Classification of Diabetes Mellitus. Follow-up report on the diagnosis of diabetes mellitus. *Diabetes Care*. 2003;26(11):3160-3170.
32. Vijan S, Sussman JB, Yudkin JS, Hayward RA. Effect of patients' risks and preferences on health gains with plasma glucose level lowering in type 2 diabetes mellitus. *JAMA Intern Med*. 2014;174(8):1227-1234.
33. Longford NT. Selection bias and treatment heterogeneity in clinical trials. *Stat Med*. 1999;18(12):1467-1474.
34. Droitcour J, Silberman G, Chelimsky E. Cross-design synthesis: a new form of meta-analysis for combining results from randomized clinical trials and medical-practice databases. *Int J Technol Assess Health Care*. 1993;9(3):440-449.
35. Kaizar EE. Estimating treatment effect via simple cross design synthesis. *Stat Med*. 2011;30(25):2986-3009.
36. Cole SR, Stuart EA. Generalizing evidence from randomized clinical trials to target populations. *Am J Epidemiol*. 2010;172(1):107-115.
37. Varadhan R, Henderson NC, Weiss CO. Cross-design synthesis for extending the applicability of trial evidence when treatment effect is heterogeneous: part I. Methodology. *Commun Stat Case Stud Data Anal Appl*. 2017;2(3-4):112-126.
38. Luijken K, Wynants L, van Smeden M, Van Calster B, Steyerberg EW, Groenwold RHH, Collaborators. Changing predictor measurement procedures affected the performance of prediction models in clinical examples. *J Clin Epidemiol*. 2020;119:7-18.
39. Vickers AJ, Scardino PT. The clinically-integrated randomized trial: proposed novel method for conducting large trials at low cost. *Trials*. 2009;10:14.
40. Simon KC, Tideman S, Hillman L, et al. Design and implementation of pragmatic clinical trials using the electronic medical record and an adaptive design. *JAMIA Open*. 2018;1(1):99-106.
41. van Staa TP, Dyson L, McCann G, et al. The opportunities and challenges of pragmatic point-of-care randomised trials using routinely collected electronic records: evaluations of two exemplar trials. *Health Technol Assess*. 2014;18(43):1-146.
42. Lorenzo C, Wagenknecht LE, Hanley AJ, Rewers MJ, Karter AJ, Haffner SM. A1C between 5.7 and 6.4% as a marker for identifying pre-diabetes, insulin sensitivity and secretion, and cardiovascular risk factors: the Insulin Resistance Atherosclerosis Study (IRAS). *Diabetes Care*. 2010;33(9):2104-2109.