**Title**: Preventive interventions and type 2 diabetes risk prediction

Byron Jaeger, Ramon Casanova, Yitbarek Demesie, Jeanette Stafford, Brian Wells, Michael P. Bancks

Wake Forest University School of Medicine

**Introduction:** Few diabetes risk prediction models offer the option to predict individualized risk conditional on initiating different preventive interventions.

**Objective:** Develop and externally validate an individualized diabetes risk prediction model with preventive intervention effects.

**Methods**: The derivation cohort included participants in the Diabetes Prevention Program (DPP) trial randomized to placebo, metformin, or intensive lifestyle intervention (N=2640). A risk prediction model for incident diabetes was developed using Cox proportional hazards regression using clinically available predictors: sex, glycated hemoglobin, fasting glucose (FG), body mass index (BMI), triglycerides, and intervention. The model was individualized by including pairwise interactions between intervention and age, FG, and BMI. The discrimination, calibration, and net benefit of the model’s 3-year predictions for incident diabetes were internally validated within the DPP using 10-fold cross validation, and externally validated among participants with prediabetes in the Multi-Ethnic Study of Atherosclerosis (MESA; N=2104).

**Results**: In DPP and MESA, mean (standard deviation) age was 51 years (11) and 64 (10) and 67% and 50% of participants were women, respectively. The mean C-statistic was 0.71 (95% confidence interval [CI]: 0.68, 0.74) in DPP and 0.86 (95% CI: 0.83, 0.88) in MESA. The optimal intervention (lowest 3yr risk) was lifestyle for 86% and 97% of DPP and MESA participants, respectively, and metformin for the remaining. If assigned to lifestyle, 3-year risk was 10% for those where lifestyle was optimal and 21% when metformin was optimal. When metformin was optimal, it was associated with 44% lower T2D risk than placebo and 29% lower than lifestyle.

**Conclusion**: Individualized predictions that forecast risk of incident diabetes after initiating a preventive intervention may improve clinical decision-making and prevention.

**Introduction**

High variation in who will progress from prediabetes to develop type 2 diabetes mellitus poses challenges for clinical decisions related to preventive interventions. The Diabetes Prevention Program (DPP) trial was designed to assess prevention of diabetes from randomization to intensive lifestyle intervention (ILI) or metformin therapy, compared to placebo.1 Compared to placebo, diabetes risk was 58% lower for the ILI arm and 31% lower for the metformin arm over 3 years.2 Not everyone in each intervention arm benefitted equally; ILI was more effective for older adults and metformin was more effective with higher baseline glucose and body mass index. Diabetes risk prediction models developed in the DPP are restricted to include only one of the active intervention arms (metformin) or do not include intervention arm as a model risk predictor and do not enable for estimation of an individualized intervention effect and predicted risk.3,4 Other diabetes risk prediction models in US populations are based on observational study data and cannot quantify difference in diabetes risk across preventive interventions.5-11 Given high variability in benefit for standard interventions, there is a clinical need for diabetes risk prediction models capable of informing individuals how initiation of a preventive intervention may impact their individualized predicted risk of diabetes progression.12,13 The objective of this study was to develop and validate a model for individualized diabetes risk prediction that incorporates expected benefits for multiple first-line diabetes prevention strategies.

**Methods**

This study was reviewed by the institutional review board (IRB) of Wake Forest University School of Medicine and approved for exempt status (Exempt Protocol: IRB00091104). IRB approval was obtained for protocols for the Diabetes Prevention Program (DPP) and Multi-Ethnic Study of Atherosclerosis (MESA) and participants in each study provided written informed consent. Data for DPP are publicly available (reduced data set) and were accessed via <https://repository.niddk.nih.gov/studies/dpp/>. Data for MESA were obtained via a data sharing agreement from the coordinating center and are publicly available <https://biolincc.nhlbi.nih.gov/studies/mesa/>.

Study Populations and Methods

DPP: The derivation sample included data from the Diabetes Prevention Program (DPP) randomized clinical trial. Study design, methods, and clinical exam procedures have been reported in detail.1,14 From 1996-1999, 3819 individuals ages 25-85 years were enrolled at 27 clinical centers across the US.1 A four-step process consisted of consent, screening, and recruitment of participants who were at high risk for type 2 diabetes based on weight and glucose status. This included an initial fasting glucose (FPG) level assessment and subsequent 75-gram oral glucose tolerance test (OGTT). Major eligibility criteria included age ≥25 years, body mass index ≥24 kilograms per meter squared (kg/m2; ≥22 kg/m2 for Asian individuals), FPG of 95-125 mg/dl (altered from 95-140 mg/dl in 1997), and 2-hour OGTT glucose of 140-199 mg/dl. Major exclusions included occurrence in the prior 6 months of myocardial infarction, symptoms of coronary heart disease, serious illness, or use of medications known to impair glucose tolerance.1 Of those enrolled, 3665 (66% women) gave consent for their de-identified data to be shared with the public. Race/ethnicity was self-reported based on the 1990 census questionnaire and collapsed into four categories for the public use data: White (58%), African American (20%), Hispanic (17%), and All Other (5%).

DPP participants were randomized (stratified by clinical site) to one of four arms, described previously: intensive lifestyle intervention (ILI), standard care plus metformin, troglitazone, or standard care plus placebo tablet.1,14 Troglitazone was discontinued in 1998 due to concerns of liver toxicity and participants for this group were not included for analysis.1 Briefly, the protocol for the three active groups were as follows. ILI was designed to achieve and maintain 7% weight loss through healthy low-calorie and low-fat diet, maintaining moderate-intensity physical activity of ≥150 minutes weekly, and a 16-session behavioral change curriculum over 24 weeks designed to help set and achieve dietary and physical activity goals (monthly afterward). Metformin regimen was titration to 850 mg twice daily (or manageable dose) with standard lifestyle recommendations and annual individual lifestyle session with case manager. Placebo was one tablet daily with standard lifestyle recommendations and annual individual lifestyle session with case manager.

MESA: The external validation sample included participants with prediabetes from the Multi-Ethnic Study of Atherosclerosis (MESA) observational cohort. In 2000-02, 6814 participants (53% women) were recruited at six field centers located in Baltimore, MD; Chicago, IL; Forsyth County, NC; Los Angeles, CA; New York, NY; and Saint Paul, MN.15 Participants were age 45-84 years and free of clinical CVD at enrollment from four racial/ethnic groups (self-identified): non-Hispanic white (38%), African American (28%), Hispanic (22%), and Chinese American (12%). MESA has completed five follow-up exams (2002-2004, 2004-2005, 2005-2007, 2010-2011, 2016-2018) all with standardized collection of demographic, socioeconomic, behavioral, clinical, and vascular imaging markers, and current medications information. Normal fasting glucose (FPG<100 mg/dl), prediabetes (FPG 100-125 mg/dl), and diabetes (FPG≥126 mg/dL or use of diabetes medications) status were determined at each MESA exam. At MESA exam 2, HbA1c was assessed and used to supplement these classifications. OGTT was not performed in MESA.

Study Population

Clinical Risk Predictors

Variables considered for inclusion in development of the risk prediction model were chosen based on a balance of availability for collection in a clinical or exam setting, biological, known risk prediction of or causal for developing type 2 diabetes, and DPP intervention response effects.2-10 A total of 20 variables were selected to be included in the risk prediction model: fasting plasma glucose (FPG) glycated hemoglobin (HbA1c), body mass index, lipids, age, sex, and DPP randomization arm. Between cohorts, continuous variables were harmonized to similar units and categorical variables to similar categorical definitions.

Primary Outcome: Diabetes Ascertainment

The primary outcome was incident diabetes, assumed to primarily be type 2 diabetes. In the DPP, diagnosis of type 2 diabetes was ascertained and defined by semi-annual measurement of FPG ≥126 mg/dl and annual post 75-gram OGTT 2-hour glucose ≥200 mg/dl.1 In MESA, diabetes was defined at each study exam as new use of insulin or oral hypoglycemic medications or FPG ≥126 mg/dl.16,17

Statistical Analysis

Participant characteristics were summarized overall and among DPP and MESA participants, separately. Continuous variables were summarized using mean (standard deviation) or median (25th, 75th percentile) and categorical variables were summarized using percentage.

Cox regression was applied to develop a risk prediction model for incident type 2 diabetes among DPP participants. The model included main effects for FPG, HbA1c, body mass index, lipids, age, sex, and DPP randomization arm, with additional pairwise interactions between randomization arm and age, FPG, and BMI. To evaluate the utility of individualizing risk predictions, a ‘non-individualized’ model was developed using the same main effects as the individualized model but without the pairwise interactions. Both model specifications were internally validated using 10-fold cross validation among DPP participants. Briefly, each DPP participant was assigned to 1 ‘fold’. Then, each fold was set aside as a temporary validation set, while the other 9 folds were used to train the models specified above. After fitting the models to 9 folds, their 3-year risk predictions for DPP participants in the temporary validation set were evaluated. After internal validation, both model specifications were fitted to the entire DPP sample, and 3-year risk predictions from these “final” prediction models were externally validated among MESA participants. For both internal and external evaluation, predictions were evaluated in terms of discrimination, calibration, and net benefit.

The final individualized prediction model was used to calculate counterfactual risk (i.e. “what if” scenarios) for participants in DPP and MESA to assess 3-year predicted risk of incident diabetes conditional on the participant initiating lifestyle, metformin, or no intervention. This model was also deployed in a freely available web application that calculates and summarizes an individual’s estimated risk for diabetes under each intervention scenario. Analyses were conducted using SAS version 9.4 (SAS Institute, Cary, North Carolina, USA) and R version 4.4.0 (The R Foundation) and statistical tests were 2-sided with alpha=0.05.

**Results**

Baseline characteristics including demographics and clinical risk factors for type 2 diabetes of the DPP (derivation) and MESA (validation) analytic samples are presented in Table 1. The DPP sample mean ± standard deviation (SD) age was 51 years ± 11, and was 67% female, 62% non-Hispanic White, 16% non-Hispanic Black, 17% Hispanic, and 5% Other race/ethnicity. Over three-quarters of DPP participants had greater than a high school education. Compared to DPP, the MESA sample was older and had a greater proportion of men, non-Hispanic Black, Hispanic, and Chinese race, and lower educational attainment. Mean ± SD BMI was 34 kg/m2 ± 7 in DPP and 30 kg/m2 ± 6 in MESA. DPP participants had higher mean values for blood glucose, insulin resistance, and beta-cell function than MESA participants. Characteristics of DPP and MESA participants who were excluded from analysis are shown in Supplemental Table 1.

At three years of follow-up, there were 386 incident cases of type 2 diabetes (17% cumulative incidence; 95% CI: 16, 19) in DPP and 202 incident cases in MESA (9.8% cumulative incidence; 95% CI: 8.5, 11).

Model comparisons

Among DPP participants, the individualized model obtained an AUC of 70.7 while the non-individualized model had AUC of 69.8, and the individualized model obtained an overall net reclassification index of 0.025 (95% CI -0.0041, 0.053) compared to the non-individualized model (Table S2). The net reclassification index indicated improvement in down-classification of risk but not in up-classification of risk for the individualized versus standard model among DPP participants. Among MESA participants, the models obtained similar C-statistics and the net reclassification index of the individualized model was 0.0060 (95% CI: -0.0052, 0.021). While the individualized model did not show signs of mis-calibration in internal validation among DPP participants, it under- and over-predicted risk for MESA participants with low and high risk, respectively (Figure 1). The individualized model classified participants into the same risk category as the standard model for 81% and 97% of DPP and MESA participants, respectively (Table S3) for 90% and 88% of men and women, respectively (Table S4), and for 86%, 90%, 90%, and 92% of Non-Hispanic White, Non-Hispanic Black, Hispanic, and Other race/ethnicity groups, respectively. Using the individualized model among DPP and MESA at a 20% risk threshold was estimated to result in 4 and 3 true positive cases identified per 100 screened, respectively (Figure S1).

Risk estimates

The associations between predictors and incident diabetes for the individualized intervention effect (interaction) model with instructions to compute predicted risk are reported in Table 2. For 86% of DPP participants and 97% of MESA participants, assignment to the intensive lifestyle intervention was optimal and resulted in the lowest 3-year mean predicted risk for diabetes (Table 3). For the remaining participants in each respective sample, assignment to metformin was the optimal individual intervention strategy. For those in DPP where lifestyle was the optimal intervention, mean 3-year risk for diabetes was 10.0% if assigned to lifestyle, 17.0% if assigned to metformin, and 22.0% if assigned to placebo. For those in DPP where metformin was the optimal intervention, mean 3-year risk for diabetes was 20.0% if assigned to lifestyle, 15.0% if assigned to metformin, and 27.0% if assigned to placebo. Supplemental Figure 2 displays individualized intervention effect 3-year predicted risk if assigned to each of the interventions according to rank order (low to high) of predicted risk for each respective intervention arm. Along the spectrum of predicted risk when assigned to placebo, respective individual predicted risk estimates when assigned to metformin or lifestyle were consistently lower. Along the spectrum of predicted risk when assigned to lifestyle, respective individual predicted risk estimates when assigned to placebo or metformin were consistently higher. Median and interquartile range of individualized 3-year predicted risk for diabetes under each intervention by race/ethnicity and sex are presented in Supplemental Figure 3. Patterns were similar across demographic groups, with median risk highest for placebo, then metformin, and lowest for lifestyle. The online diabetes risk calculator is published here: <https://bcjaeger.shinyapps.io/PreDm2Calc/>.

**Discussion**

This study used data from a large, randomized trial for type 2 diabetes prevention with multiple intervention arms to develop an individualized intervention effect risk prediction model and equation for developing diabetes over 3 years. Among participantsversuswith similar In MESA, an external sample without intervention information, the individualized model was similar to the standard model in performance metrics. This risk prediction model provides an important advancement in diabetes risk prediction and prevention. No current diabetes risk prediction tool can calculate and summarize an individual’s estimated risk for developing diabetes after starting a preventive intervention. This model includes two different preventive interventions, intensive lifestyle intervention and metformin therapy, and estimates an individual’s risk for diabetes should they start either preventive intervention or not start an intervention (placebo). Further, the model incorporates the differential preventive effects of the intensive lifestyle and metformin interventions observed in the Diabetes Prevention Program. Thus, it provides an estimate of diabetes risk for each intervention effect that is sensitive to an individual’s clinical risk profile rather than subtracting the mean risk reduction effect of each intervention group from placebo risk.

The risk predictor variables included in the model were six clinical and biological characteristics and intervention arm. Race was not included as a model predictor variable because race is not a biological trait. Both analytic samples that contributed to the derivation and validation of the model included racially and ethnically diverse study populations. Metrics of model fairness across race/ethnicity, namely positive predictive capability, were strong (>80%) in the derivation cohort and support the use of a race-free diabetes risk prediction model. Data on genetic ancestry or genetic risk for diabetes were not included in the publicly available data for DPP, would not be available in a clinical setting, and were not included in the model.

Multiple diabetes risk prediction models, some using DPP data, already exist for US populations and some of these models have better internal discrimination than the model presented here.3-11

Prior models using the DPP trial data show absolute risk reduction ranging from 0 to over 20% across quartiles of estimated T2D risk at baseline.3,4 Limitations of these models include lack of external validation, use of race as a predictor variable, restricting risk prediction to only the metformin arm, not directly modeling intervention arm, or not modeling heterogeneity of the intervention by individual clinical factors and rather using quartile groups of estimated T2D risk, which may be an oversimplification of a heterogenous variable.3,4 These models do not provide a comparison of diabetes predicted risk across each of the placebo, metformin, and lifestyle interventions for an individual and their specific clinical profile. In addition to sharing some of these limitations, the existing risk prediction models using observational cohort data can only quantify risk for diabetes under a placebo scenario and lack potential for inference from empirical metformin and lifestyle intervention effects.5-11

The inclusion of intervention arm and interaction terms between intervention arm with weight and fasting glucose are critically important. These parameters can help people understand their risk for developing diabetes on their current progress without any intervention and if they started either lifestyle or metformin, the current recommended first-line treatments for T2D.18 An important advantage of this parsimonious diabetes risk prediction model is that it can be implemented using data available in an electronic health record (EHR) and clinical setting. This risk prediction equation and accompanying online risk calculator can help inform precision medicine-based efforts for diabetes prevention and guide clinician-patient discussions on patient-specific expected benefit by providing a comparison of diabetes risk across potential intervention strategies. For most people, the intensive lifestyle intervention will be optimal and provide individuals with the lowest risk for developing diabetes over the next 3 years. For a small subset of individuals, metformin therapy might provide the lowest 3-year risk of diabetes. Where metformin is the optimal intervention (lowest 3-year risk), individuals should accompany metformin therapy with lifestyle modification of diet and physical activity for weight loss18, albeit possibly not as intensive as the caloric restriction and physical activity rigor and duration goals in the DPP. Metformin therapy in combination with intensive lifestyle intervention was not assessed in DPP.

Several limitations merit consideration when interpreting these results. The study populations used to develop and validate the risk prediction model were restricted to individuals with prediabetes. Model inference and application to individuals with normal glucose is not advised. Validation was performed in an observational cohort where no individuals were randomized to receive intensive lifestyle or metformin for diabetes prevention. The preventive interventions, metformin and intensive lifestyle modification, were assumed to follow the protocol employed in the DPP trial and modeling does not account for cross-over or non-adherence to intervention. Newer pharmacotherapies for weight loss have been developed since the origin of DPP and are used for individuals without diabetes and with overweight or obesity. This model does not include these newer medications as a potential preventive intervention.

In conclusion, this study provides the first diabetes risk prediction model with individualized preventive intervention effects based on data from a randomized control trial of metformin therapy or intensive lifestyle intervention. The race-free diabetes risk prediction model and equation were developed and validated in racially and ethnically diverse study populations. Parsimonious and commonly available clinical risk predictors coupled with the development of a free and publicly available online diabetes risk calculator enables high potential clinical utility of this research on individualized diabetes prevention.

**Acknowledgments and Funding:** The authors have no conflicts to disclose. The views expressed in this manuscript are those of the authors and do not necessarily represent the views of the American Diabetes Association, National Heart, Lung, and Blood Institute; the National Eye Institute; the National Institutes of Health; or the U.S. Department of Health and Human Services. This specific project was funded by the American Diabetes Association (ADA) Grant 11-22-ICTSPM-18 (PI: Bancks). MESA was funded by contracts 75N92020D00001, HHSN268201500003I, N01-HC-95159, 75N92020D00005, N01-HC-95160, 75N92020D00002, N01-HC-95161, 75N92020D00003, N01-HC-95162, 75N92020D00006, N01-HC-95163, 75N92020D00004, N01-HC-95164, 75N92020D00007, N01-HC-95165, N01-HC-95166, N01-HC-95167, N01-HC-95168 and N01-HC-95169 from the National Heart, Lung, and Blood Institute, grants UL1-TR-000040, UL1-TR-001079, and UL1-TR-001420 from the National Center for Advancing Translational Sciences (NCATS). The original DPP trial was supported by the National Institutes of Health through the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK), the Office of Research on Minority Health, the National Institute of Child Health and Human Development, the National Institute on Aging, the Indian Health Service, the Centers for Disease Control and Prevention, the American Diabetes Association, and Bristol-Myers Squibb and Parke-Davis.

**References:**

1. Diabetes Prevention Program Research Group. The Diabetes Prevention Program. Design and Methods for a Clinical Trial in the Prevention of Type 2 Diabetes. *Diabetes Care.* 1999;22(4):623-634.

2. Diabetes Prevention Program Research Group. Reduction in the Incidence of Type 2 Diabetes with Lifestyle Intervention or Metformin. *The New England Journal of Medicine.* 2002;346(6):393-403.

3. Parast L, Mathews M, Friedberg MW. Dynamic Risk Prediction for Diabetes Using Biomarker Change Measurements. *BMC Medical Research Methodology.* 2019;19(1):175.

4. 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.

5. Ayensa-Vazquez JA, Leiva A, Tauler P, López-González AA, Aguiló A, Tomás-Salvá M, Bennasar-Veny M. Agreement between Type 2 Diabetes Risk Scales in a Caucasian Population: A Systematic Review and Report. *J Clin Med.* 2020;9(5).

6. Stern MP, Williams K, Haffner SM. Identification of Persons at High Risk for Type 2 Diabetes Mellitus: Do We Need the Oral Glucose Tolerance Test? *Annals of Internal Medicine.* 2002;136(8):575-581.

7. Schmidt MIs, Duncan BB, Bang H, Pankow JS, Ballantyne CM, Golden SH, Folsom AR, Chambless LE, for the Atherosclerosis Risk in Communities I. Identifying Individuals at High Risk for Diabetes: The Atherosclerosis Risk in Communities Study. *Diabetes Care.* 2005;28(8):2013-2018.

8. Wilson PW, Meigs JB, Sullivan L, Fox CS, Nathan DM, D'Agostino RB, Sr. Prediction of Incident Diabetes Mellitus in Middle-Aged Adults: The Framingham Offspring Study. *Archives of Internal Medicine.* 2007;167(10):1068-1074.

9. Kahn HS, Cheng YJ, Thompson TJ, Imperatore G, Gregg EW. Two Risk-Scoring Systems for Predicting Incident Diabetes Mellitus in U.S. Adults Age 45 to 64 Years. *Annals of Internal Medicine.* 2009;150(11):741-751.

10. Casanova R, Saldana S, Simpson SL, Lacy ME, Subauste AR, Blackshear C, Wagenknecht L, Bertoni AG. Prediction of Incident Diabetes in the Jackson Heart Study Using High-Dimensional Machine Learning. *PloS One.* 2016;11(10):e0163942.

11. Bang H, Edwards AM, Bomback AS, Ballantyne CM, Brillon D, Callahan MA, Teutsch SM, Mushlin AI, Kern LM. Development and Validation of a Patient Self-Assessment Score for Diabetes Risk. *Annals of Internal Medicine.* 2009;151(11):775-783.

12. American Diabetes Association Professional Practice Committee. Standards of Medical Care in Diabetes - 2022. *Diabetes Care.* 2022;45(Supplement 1):S1-264.

13. Chung WK, Erion K, Florez JC, Hattersley AT, Hivert MF, Lee CG, McCarthy MI, Nolan JJ, Norris JM, Pearson ER, Philipson L, McElvaine AT, Cefalu WT, Rich SS, Franks PW. Precision Medicine in Diabetes: A Consensus Report from the American Diabetes Association (Ada) and the European Association for the Study of Diabetes (Easd). *Diabetes Care.* 2020;43(7):1617-1635.

14. Diabetes Prevention Program Research Group. The Diabetes Prevention Program: Baseline Characteristics of the Randomized Cohort. The Diabetes Prevention Program Research Group. *Diabetes Care.* 2000;23(11):1619-1629.

15. Bild DE, Bluemke DA, Burke GL, Detrano R, Diez Roux AV, Folsom AR, Greenland P, Jacob DR, Jr., Kronmal R, Liu K, Nelson JC, O'Leary D, Saad MF, Shea S, Szklo M, Tracy RP. Multi-Ethnic Study of Atherosclerosis: Objectives and Design. *American Journal of Epidemiology.* 2002;156(9):871-881.

16. Bancks M, Bielinski S, Decker P, Hanson N, Larson N, Sicotte H, Wassel C, Pankow J. Circulating Level of Hepatocyte Growth Factor Predicts Incidence of Type 2 Diabetes Mellitus: The Multi-Ethnic Study of Atherosclerosis (Mesa). *Metabolism Clinical and Experimental.* 2016;65(3):64-72.

17. Bild DE, Detrano R, Peterson D, Guerci A, Liu K, Shahar E, Ouyang P, Jackson S, Saad MF. Ethnic Differences in Coronary Calcification: The Multi-Ethnic Study of Atherosclerosis (Mesa). *Circulation.* 2005;111(10):1313-1320.

18. American Diabetes Association Professional Practice Committee. Standards of Care in Diabetes—2024. *Diabetes Care.* 2024;47(S1):S1-S321.

**Tables and Figures**

**Table 1.** Baseline characteristics of Diabetes Prevention Program (derivation) and Multi-Ethnic Study of Atherosclerosis (validation) participants included in the current analysis.

| **Characteristic** | **DPP, N = 2,640** | **MESA, N = 2,104** |
| --- | --- | --- |
| Age, years | 51 (11) | 64 (10) |
| Sex |  |  |
| Male | 864 (33%) | 1,058 (50%) |
| Female | 1,776 (67%) | 1,046 (50%) |
| Race/ethnicity |  |  |
| Non-Hispanic White | 1,626 (62%) | 682 (32%) |
| Non-Hispanic Black | 423 (16%) | 558 (27%) |
| Hispanic | 448 (17%) | 510 (24%) |
| Other/Chinese | 143 (5.4%) | 354 (17%) |
| Educational attainment |  |  |
| < High School | 175 (7%) | 196 (23%) |
| High School Graduate | 467 (18%) | 168 (20%) |
| Some College or College Graduate | 1998 (76%) | 496 (58%) |
| Fasting glucose, mg/dl | 106 (7) | 101 (10) |
| Glycated Hemoglobin, % | 5.78 (0.40) | 5.75 (0.30) |
| HOMA-Insulin Resistance | 6.0 (4.2, 8.6) | --- |
| HOMA-Beta cell function | 199 (136, 272) | --- |
| Body mass index, kg/m2 | 34 (7) | 30 (6) |
| Triglycerides, mg/dl | 144 (101, 205) | 118 (84, 167) |
| Low-density lipoprotein cholesterol, mg/dl | 107 (27) | --- |
| High-density lipoprotein cholesterol, mg/dl | 46 (12) | --- |
| Abbreviations: DPP = Diabetes Prevention Program; HOMA = Homeostatic Model Assessment; and MESA = Multi-Ethnic Study of Atherosclerosis | | |

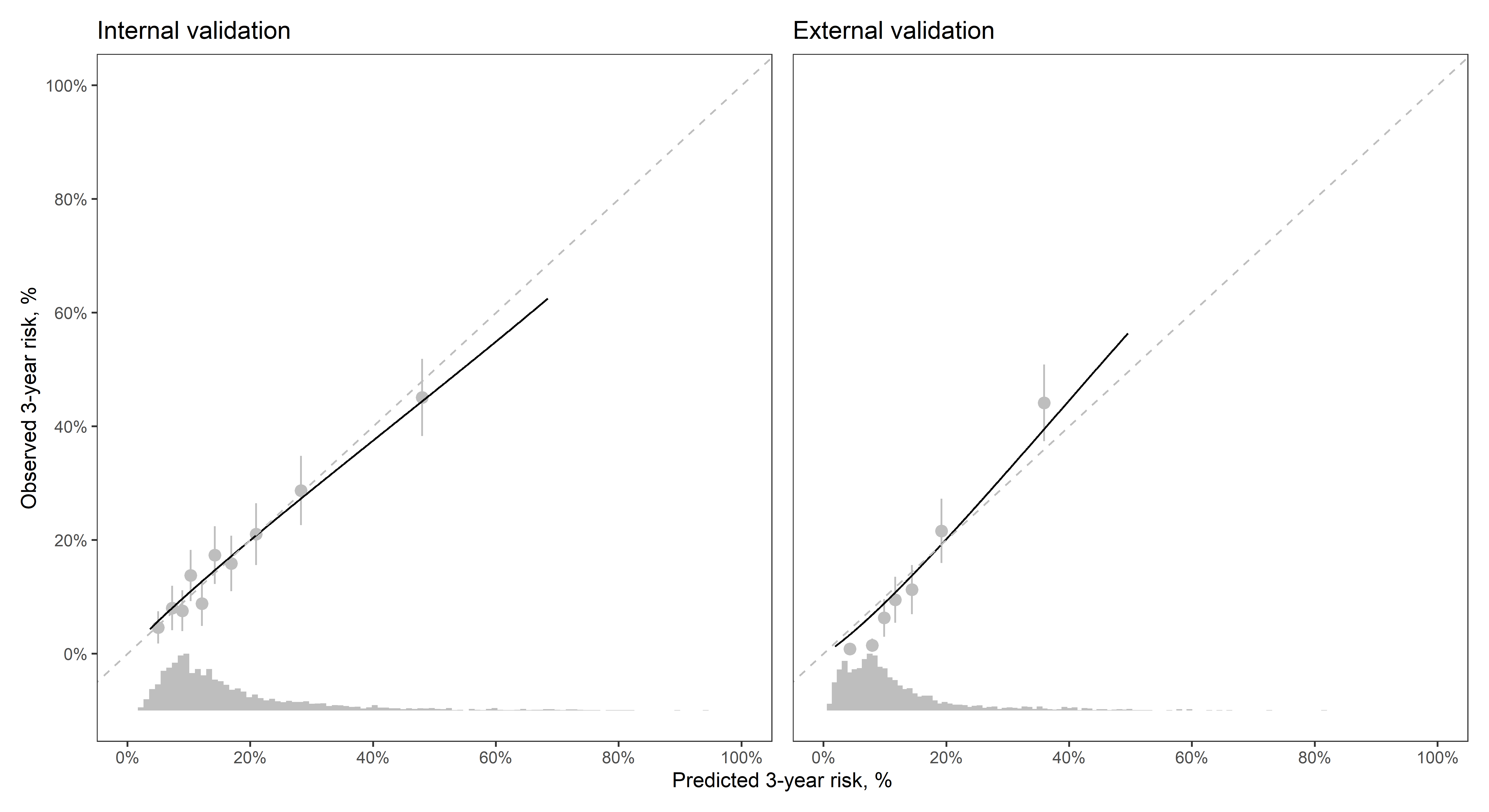
**Table 2**. Summary of and instructions for using the individualized intervention effect risk prediction model for type 2 diabetes.

| **Variable1** | **Hazard ratio (95% CI)2** |
| --- | --- |
| Glycated Hemoglobin | 1.21 (1.09, 1.34) |
| Triglycerides | 1.24 (1.15, 1.35) |
| Age, per 10.6 years, conditional on treatment | |
| Lifestyle | 0.95 (0.78, 1.16) |
| Metformin | 1.02 (0.86, 1.23) |
| Placebo | 0.97 (0.84, 1.13) |
| Body mass index, per 6.6 kg/m2, conditional on treatment | |
| Lifestyle | 1.37 (1.16, 1.60) |
| Metformin | 0.89 (0.74, 1.06) |
| Placebo | 1.04 (0.91, 1.19) |
| Fasting glucose, per 6.7 mg/dL, conditional on treatment | |
| Lifestyle | 1.47 (1.23, 1.76) |
| Metformin | 1.53 (1.31, 1.78) |
| Placebo | 1.90 (1.67, 2.17) |
| Sex | |
| Female | 1.16 (0.95, 1.43) |
| Male | 1.00 (Reference) |
| 1Predictor variables included in the table were selected *a priori* based on clinical availability and known associations. | |
| 2Hazard ratios are adjusted for all variables listed in the table. Hazard ratios for continuous variables correspond to a one standard deviation change in the variable.  Instructions for computing 3-year predicted risk:  **Step 1: Scale predictors:**  - divide age by 10.58244063 - divide fasting glucose by 6.67887385 - divide glycated hemoglobin by 0.39663737 - divide triglycerides by 95.62000286 - divide body mass index by 6.55160219  **Step 2: Compute linear predictor (LP):**  LP = (sex = female) \* 0.15268123 + (treatment group = metformin) \* 1.73746560 + (treatment group = placebo) \* -2.06808662 + (glycated hemoglobin) \* 0.19224983 + (triglycerides) \* 0.21835857 age \* [-0.04914900 + (treatment group = metformin) \* 0.07302687) + (treatment group = placebo) \* 0.02231801)] body mass index \* [ 0.31117408 + (treatment group = metformin) \* -0.42922833) + (treatment group = placebo) \* -0.26864075)] fasting glucose \* [0.38697814 + (treatment group = metformin) \* 0.03748913) + (treatment group = placebo) \* 0.25646930)]  **Step 3: Enter linear predictor into the risk formula:** Risk = 1 - exp(-0.155833 \* exp(LP - 11.216613)) | |

**Table 3.** Optimal preventive intervention and mean 3-year predicted risk (counterfactual risk) for type 2 diabetes from an individualized intervention effect risk prediction model in the Diabetes Prevention Program (derivation) and Multi-Ethnic Study of Atherosclerosis (validation).

| **Optimal intervention** | **N (% sample)** | **Mean (SD) 3-year Predicted Risk (Counterfactual Risk)** | | |
| --- | --- | --- | --- | --- |
| **Lifestyle** | **Metformin** | **Placebo** |
| Diabetes Prevention Program (derivation) | | | | |
| Lifestyle | 2,267 (86%) | 10.2 (6.5) | 17.3 (9.4) | 22.2 (15.8) |
| Metformin | 373 (14%) | 20.5 (11.7) | 14.7 (7.6) | 26.7 (17.4) |
| Multi-Ethnic Study of Atherosclerosis (validation) | | | | |
| Lifestyle | 2,035 (97%) | 6.5 (5.2) | 14 (9.7) | 15 (15) |
| Metformin | 69 (3%) | 16 (11) | 13 (9) | 23 (20) |
| \*Optimal intervention for an individual is the intervention arm with the lowest 3-year predicted risk for diabetes for each respective individual in the DPP. Placebo is never the optimal intervention to prevent diabetes for any individual. | | | | |

**Figure 1.** Calibration with histogram of the individualized intervention effect risk prediction model in the Diabetes Prevention Program (derivation, left pane) and Multi-Ethnic Study of Atherosclerosis (validation, right pane).



**Supplemental Materials -** Preventive interventions and type 2 diabetes risk prediction

Byron Jaeger, Ramon Casanova, Yitbarek Demesie, Jeanette Stafford, Brian Wells, Michael P. Bancks, Wake Forest University School of Medicine

**Supplemental Table 1.** Descriptive table of Diabetes Prevention Program and Multi-Ethnic Study of Atherosclerosis participants excluded from the current analysis.

**Supplemental Table 2.** Model performance and fairness statistics overall for the individualized intervention effect and standard risk prediction models in the Diabetes Prevention Program and Multi-Ethnic Study of Atherosclerosis.

**Supplemental Table 3.** Reclassification matrix overall comparing the individualized intervention effect versus standard risk prediction model classification according to sample, the Diabetes Prevention Program (internal) and the Multi-Ethnic Study of Atherosclerosis (external).

**Supplemental Table 4.** Reclassification matrix comparing individualized intervention effect versus standard risk prediction model classification according to sex in the Diabetes Prevention Program (internal) and the Multi-Ethnic Study of Atherosclerosis (external) combined.

**Supplemental Table 5.** Reclassification matrix comparing individualized intervention effect versus standard risk prediction model classification according to race/ethnicity in the Diabetes Prevention Program (internal) and the Multi-Ethnic Study of Atherosclerosis (external) combined.

**Supplemental Figure 1.** Decision curve analysis of the individualized intervention effect and standard risk prediction model in the Diabetes Prevention Program (derivation, left pane) and Multi-Ethnic Study of Atherosclerosis (validation, right pane).

**Supplemental Figure 2.** Individualized intervention effect 3-year predicted risk for diabetes if assigned to each of the interventions according to rank order (low to high) of predicted risk for placebo (top panel), metformin (middle panel), and intensive lifestyle intervention (bottom panel) in the Diabetes Prevention Program.

**Supplemental Figure 3.** Median and interquartile range (25th and 75th percentile) for individualized intervention effect 3-year predicted risk for diabetes if assigned to each of the interventions according to race/ethnicity (left panel) and sex (right panel) in the Diabetes Prevention Program.

**Supplemental Table 1.** Descriptive table of Diabetes Prevention Program and Multi-Ethnic Study of Atherosclerosis participants excluded from the current analysis.

| **Characteristic** | **DPP, N = 1,025** | **MESA, N = 4,710** |
| --- | --- | --- |
| Age, years | 52 (10) | 62 (10) |
| Sex |  |  |
| Male | 364 (36%) | 2,155 (46%) |
| Female | 661 (64%) | 2,555 (54%) |
| Race/ethnicity |  |  |
| Non-Hispanic White | 491 (48%) | 1,941 (41%) |
| Non-Hispanic Black | 328 (32%) | 533 (11%) |
| Hispanic | 161 (16%) | 1,207 (26%) |
| Other/Chinese | 45 (4.4%) | 1,029 (22%) |
| Educational attainment |  |  |
| < High School | 550 (54%) | 821 (17%) |
| High School Graduate | 186 (18%) | 812 (17%) |
| Some College or College Graduate | 289 (28%) | 3,063 (65%) |
| Fasting glucose, mg/dl | 110 (9) | 98 (36) |
| Glycated Hemoglobin, % | 6.23 (0.60) | --- |
| HOMA-Insulin Resistance | 6.7 (4.5, 9.1) | --- |
| HOMA-Beta cell function | 188 (130, 263) | --- |
| Body mass index, kg/m2 | 34 (7) | 28 (5) |
| Triglycerides, mg/dl | 133 (96, 194) | 109 (76, 159) |
| Low-density lipoprotein cholesterol, mg/dl | 108 (28) | --- |
| High-density lipoprotein cholesterol, mg/dl | 45 (12) | --- |
| Abbreviations: DPP = Diabetes Prevention Program; HOMA = Homeostatic Model Assessment; and MESA = Multi-Ethnic Study of Atherosclerosis | | |

**Supplemental Table 2.** Model performance and fairness statistics overall for the individualized intervention effect and standard risk prediction models in the Diabetes Prevention Program and Multi-Ethnic Study of Atherosclerosis.

| **Evaluation statistic1** | **Internal Validation (DPP)2** | | **External Validation (MESA)3** | |
| --- | --- | --- | --- | --- |
| **Standard** | **Individualized** | **Standard** | **Individualized** |
| NRI, Overall | 0 (ref) | 2.5 (-0.41, 5.3) | 0 (ref) | 0.60 (-0.53, 2.1) |
| NRI, Positive | 0 (ref) | 0.15 (-2.6, 2.6) | 0 (ref) | 0.96 (0.00, 2.5) |
| NRI, Negative | 0 (ref) | 2.4 (1.3, 3.5) | 0 (ref) | -0.37 (-0.71, -0.10) |
| AUC | 69.8 (66.7 73.0) | 70.7 (67.6 73.9) | 85.8 (83.3 88.3) | 85.6 (83.0 88.1) |
| IPA | 9.9 | 10.4 | 18.8 | 18.6 |
| Race/ethnicity | | | | |
| Demographic parity4 | 90 | 87 | 50 | 51 |
| Equal opportunity5 | 89 | 74 | 93 | 87 |
| Equal odds6 | 89 | 74 | 93 | 87 |
| Sex | | | | |
| Demographic parity | 93 | 90 | 88 | 88 |
| Equal opportunity | 83 | 84 | 82 | 85 |
| Equal odds | 83 | 84 | 82 | 85 |
| Abbreviations: AUC = Area underneath the receiver-operator characteristic curve; IPA = Index of prediction accuracy; and NRI = Net reclassification index | | | | |
| 1Table values are scaled by a factor of 100 for ease of interpretation. | | | | |
| 2Internal validation results are based on 10-fold cross-validation in the Diabetes Prevention Program data | | | | |
| 3External validation results are based on application of models fitted to the Diabetes Prevention Program data to the Multi-Ethnic Study of Atherosclerosis data. | | | | |
| 4Demographic parity is satisfied when a model's predictions have the same predicted positive rate across groups. Range 0-100 (ideal). | | | | |
| 5Equal opportunity is satisfied when a model's predictions have the same true positive and false negative rates across protected groups. Range 0-100 (ideal). | | | | |
| 6Equal odds is satisfied when a model's predictions have the same false positive, true positive, false negative, and true negative rates across protected groups. Range 0-100 (ideal). | | | | |

**Supplemental Table 3.** Reclassification matrix overall comparing the individualized intervention effect versus standard risk prediction model classification according to sample, the Diabetes Prevention Program (derivation) and the Multi-Ethnic Study of Atherosclerosis (validation).

| **Standard risk categories** | **Individualized risk categories** | | |
| --- | --- | --- | --- |
| **0 to < 10%** | **10% to < 20%** | **≥ 20%** |
| Diabetes Prevention Program (derivation) | | | |
| 0 to < 10% | 708 (27%) | 157 (5.9%) | 3 (0.11%) |
| 10% to < 20% | 170 (6.4%) | 790 (30%) | 60 (2.3%) |
| ≥ 20% | 0 | 114 (4.3%) | 638 (24%) |
| Multi-Ethnic Study of Atherosclerosis (validation) | | | |
| 0 to < 10% | 1,174 (56%) | 37 (1.8%) | 0 |
| 10% to < 20% | 1 (0.05%) | 610 (29%) | 10 (0.48%) |
| ≥ 20% | 0 | 1 (0.05%) | 271 (13%) |

**Supplemental Table 4.** Reclassification matrix comparing individualized intervention effect versus standard risk prediction model classification according to sex in the Diabetes Prevention Program (internal) and the Multi-Ethnic Study of Atherosclerosis (external) combined.

| **Standard risk categories** | **Individualized risk categories** | | |
| --- | --- | --- | --- |
| **0 to < 10%** | **10% to < 20%** | **≥ 20%** |
| Men | | | |
| 0 to < 10% | 819 (43%) | 72 (3.7%) | 0 |
| 10% to < 20% | 60 (3.1%) | 559 (29%) | 25 (1.3%) |
| ≥ 20% | 0 | 34 (1.8%) | 353 (18%) |
| Women | | | |
| 0 to < 10% | 1,063 (38%) | 122 (4.3%) | 3 (0.11%) |
| 10% to < 20% | 111 (3.9%) | 841 (30%) | 45 (1.6%) |
| ≥ 20% | 0 | 81 (2.9%) | 556 (20%) |

**Supplemental Table 5.** Reclassification matrix comparing individualized intervention effect versus standard risk prediction model classification according to race/ethnicity in the Diabetes Prevention Program (internal) and the Multi-Ethnic Study of Atherosclerosis (external) combined.

| **Standard risk categories** | **Individualized risk categories** | | |
| --- | --- | --- | --- |
| **0 to < 10%** | **10% to < 20%** | **≥ 20%** |
| Non-Hispanic White | | | |
| 0 to < 10% | 790 (34%) | 109 (4.7%) | 2 (0.09%) |
| 10% to < 20% | 98 (4.2%) | 707 (31%) | 47 (2.0%) |
| ≥ 20% | 0 | 73 (3.2%) | 482 (21%) |
| Non-Hispanic Black | | | |
| 0 to < 10% | 498 (51%) | 37 (3.8%) | 1 (0.10%) |
| 10% to < 20% | 33 (3.4%) | 239 (24%) | 8 (0.82%) |
| ≥ 20% | 0 | 19 (1.9%) | 146 (15%) |
| Hispanic | | | |
| 0 to < 10% | 407 (42%) | 36 (3.8%) | 0 |
| 10% to < 20% | 26 (2.7%) | 289 (30%) | 7 (0.73%) |
| ≥ 20% | 0 | 17 (1.8%) | 176 (18%) |
| Other | | | |
| 0 to < 10% | 187 (38%) | 12 (2.4%) | 0 |
| 10% to < 20% | 14 (2.8%) | 165 (33%) | 8 (1.6%) |
| ≥ 20% | 0 | 6 (1.2%) | 105 (21%) |

**Supplemental Figure 1.** Decision curve analysis of the individualized intervention effect and standard risk prediction model in the Diabetes Prevention Program (derivation, left pane) and Multi-Ethnic Study of Atherosclerosis (validation, right pane).

A graph of a patient

Description automatically generated with medium confidence

**Supplemental Figure 2.** Individualized intervention effect 3-year predicted risk for diabetes if assigned to each of the interventions according to rank order (low to high) of predicted risk for placebo (top panel), metformin (middle panel), and intensive lifestyle intervention (bottom panel) in the Diabetes Prevention Program.

A graph of different levels of risk

Description automatically generated with medium confidence

**Supplemental Figure 3.** Median and interquartile range (25th and 75th percentile) for individualized intervention effect 3-year predicted risk for diabetes if assigned to each of the interventions according to race/ethnicity (left panel) and sex (right panel) in the Diabetes Prevention Program.

**A graph of different colored lines

Description automatically generated with medium confidence**