Individualized risk prediction for type 2 diabetes

**Notes**

**Table 1**: some variables listed in the table shells were not in the data. I am happy to include these but I’d need some help finding the variables to add.

**Table 4**: NRI and the fairness metrics required a cut-point to be computed. I used a cut-point of 20%. It’s easy to change this if you’d like to - just let me know.

**Table 5**: It isn’t feasible to stratify by incident cases here. This is because:

* an incident case would be someone who had diabetes diagnosis before 3 years had passed.
* a non incident case would be someone who did not have diabetes before 3 years passed
* a number of participants didn’t fit into either category because they were censored before 3 years passed.

Also, I had to modify the risk categories for table 5. The original risk categories went up to 10% but a large majority of predicted risk values were > 10% so I modified the categories to make the tables more balanced. Easy to change if needed.

**Figures**: I grouped some figures together as I thought it would be helpful to see them side by side. Easy to change this, just let me know if you’d prefer to keep them separate.

Table 0: Participants included.

| **Inclusion criteria** | **DPP** | **MESA** |
| --- | --- | --- |
| Study participants | 3,665 | 6,814 |
| Pre-diabetic at baseline1 | 3,152 | 2,104 |
| Randomized to placebo, metformin, or lifestyle | 2,640 | -- |
| *Abbreviations: DPP = Diabetes Prevention Program; and MESA = Multi-Ethnic Study of Atherosclerosis* | | |
| *1Baseline for Multi-Ethnic Study of Atherosclerosis participants was visit 1 or 2 dependent on pre-diabetes status* | | |

Table 1: Descriptive table of Diabetes Prevention Program and Multi-Ethnic Study of Atherosclerosis 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 | 1,352 (51%) | 196 (23%) |
| High School Graduate | 574 (22%) | 168 (20%) |
| Some College or College Graduate | 714 (27%) | 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: 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 | | |

Table 3. Cumulative incidence of diabetes over follow-up in the Diabetes Prevention Program and Multi-Ethnic Study of Atherosclerosis

|  | **DPP** | **MESA** |
| --- | --- | --- |
| Over all follow-up time | | |
| Number at risk | 2,640 | 2,103 |
| Total follow-up time, person-years | 7,305 | 18,026 |
| Median (95% CI) follow-up time, years | 3.0 (3.0, 3.0) | 13 (9.6, 13) |
| Up to three years after baseline | | |
| Number of incident cases | 386 | 202 |
| Incidence rate, per 100 years | 5.9 | 3.4 |
| Cumulative incidence (95% CI), % | 17 (16, 19) | 9.8 (8.5, 11) |

Table 4. Comparison of the individualized model with standard model.

| **Evaluation statistic1** | **Internal Validation2** | | **External Validation3** | |
| --- | --- | --- | --- | --- |
| **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. | | | | |
| 5Equal opportunity is satisfied when a model's predictions have the same true positive and false negative rates across protected groups. | | | | |
| 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. | | | | |

Table 5A: Reclassification matrix comparing individualized versus standard risk classification among participants in the Diabetes Prevention Program (internal) and the Multi-Ethnic Study of Atherosclerosis (external)

| **Standard risk categories** | **Individualized risk categories** | | |
| --- | --- | --- | --- |
| **0 to < 10%** | **10% to < 20%** | **≥ 20%** |
| Internal data | | | |
| 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%) |
| External data | | | |
| 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%) |

Table 5B: Reclassification matrix comparing individualized versus standard risk classification among men and women.

| **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%) |

Table 5C: Reclassification matrix comparing individualized versus standard risk classification among men and women.

| **Standard risk categories** | **Individualized risk categories** | | |
| --- | --- | --- | --- |
| **0 to < 10%** | **10% to < 20%** | **≥ 20%** |
| Caucasian | | | |
| 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%) |
| African American | | | |
| 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%) |

Table 6: Optimal preventive intervention and 3-year counterfactual risk for type 2 diabetes from an individualized risk prediction model in the Diabetes Prevention Program

| **Optimal intervention** | **N (%)** | **Counterfactual 3-year T2D Predicted Risk, Mean (SD)** | | |
| --- | --- | --- | --- | --- |
| **Lifestyle** | **Metformin** | **Placebo** |
| Internal | | | | |
| Lifestyle | 2,267 (86%) | 10 (6.5) | 17 (9.4) | 22 (16) |
| Metformin | 373 (14%) | 20 (12) | 15 (7.6) | 27 (17) |
| External | | | | |
| Lifestyle | 2,035 (97%) | 6.5 (5.2) | 14 (9.7) | 15 (15) |
| Metformin | 69 (3.3%) | 16 (11) | 13 (9.0) | 23 (20) |

Table 7: Summary of and instructions for using the individualized 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 | |
| Lifestyle | 0.95 (0.78, 1.16) |
| Metformin | 1.02 (0.86, 1.23) |
| Placebo | 0.97 (0.84, 1.13) |
| Body mass index | |
| Lifestyle | 1.37 (1.16, 1.60) |
| Metformin | 0.89 (0.74, 1.06) |
| Placebo | 1.04 (0.91, 1.19) |
| Fasting glucose | |
| 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 apriori 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: Plug LP into the risk formula:   Risk = 1 - exp(-0.155833 \* exp(LP - 11.216613)) | |

Figure 1: Decision curve analysis of the individualized and standard model in internal and external validation.

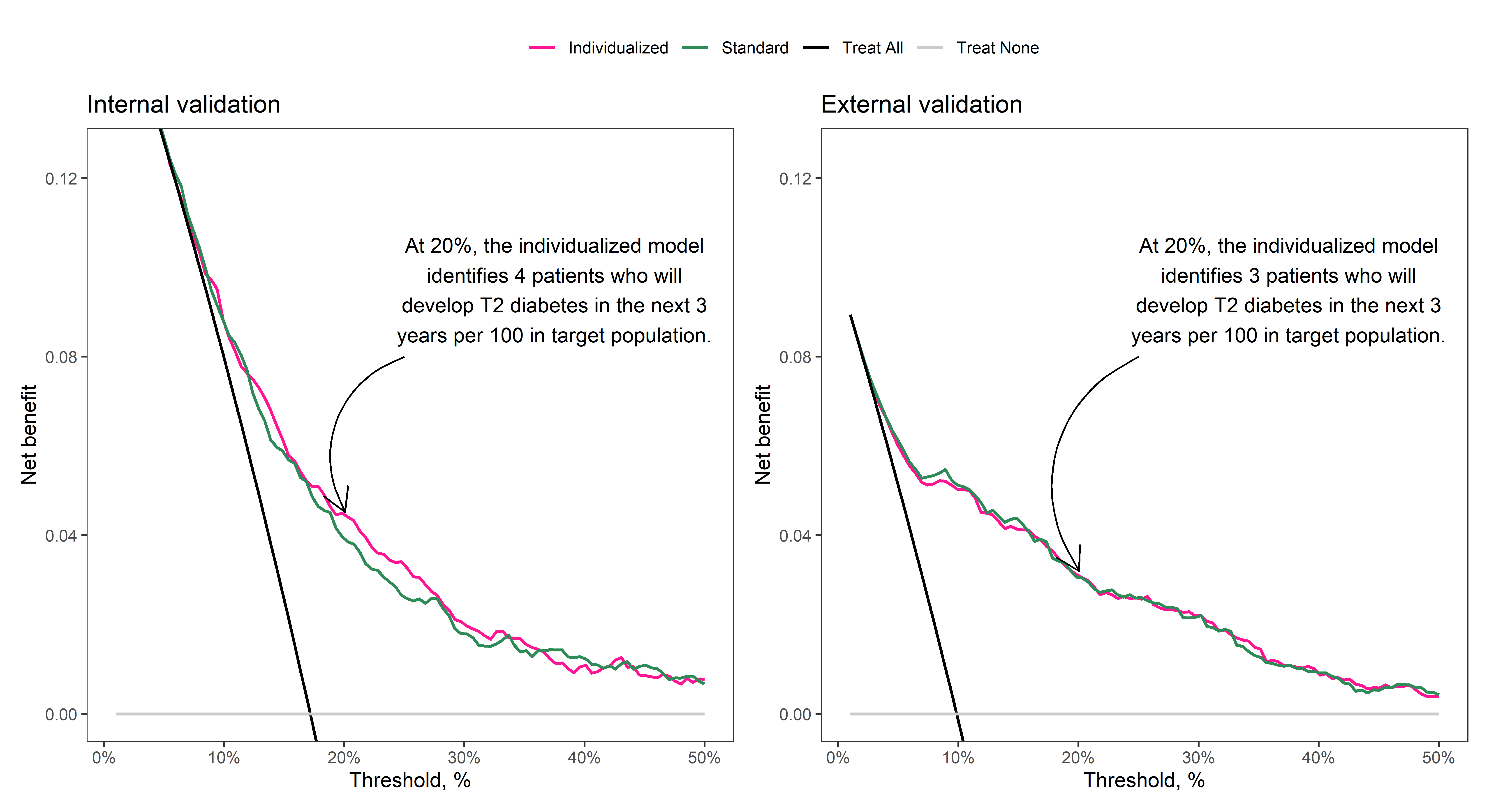


Figure 2: Calibration of the individualized model in internal and external validation.

