# Appendix to Identifying Risk for Type 2 Diabetes in Different Age Cohorts: Does One Size Fit All?

A-1.: Graphical comparisons of odd ratios across individual’s characteristics

Note: the vertical dotted line represents the weighted average of a coefficient across all datasets. The size of the square represents the sample size, and the horizontal lines crossing the squares represent the 95% confidence intervals.





















A-2: Diagnostic characteristics in the testing sample of quintiles predicting risk of diabetes

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Models and rules** | **% >Threshold** | **Sensitivity** | **Specificity** | **Positive predicted value** | **Negative predicted value** |
| **CARDIA Simple Model** |   |   |   |   |   |
| Pr(DM) | AUROC = 0.73 (0.68– 0.79) |
| >=0.017 | 80 | 97 (92–100) | 21 (20–21) | 5 (4–6) | 99 (98–100) |
| >=0.025 | 60 | 89 (82–96) | 41 (41–42) | 6 (5–8) | 99 (98–100) |
| >=0.038 | 40 | 73 (61–84) | 61 (61–62) | 8 (6–10) | 98 (97–99) |
| >=0.056 | 20 | 46 (34–60) | 81 (81–82) | 10 (7–13) | 97 (96–98) |
| **CARDIA Enhanced Model** |  |  |  |  |  |
| Pr(DM) | AUROC = 0.76 (0.71– 0.82) |
| >=0.013 | 80 | 97 (92–100) | 21 (20–21) | 5 (4–6) | 99 (98–100) |
| >=0.022 | 60 | 90 (82–97) | 41 (41–42) | 6 (5–8) | 99 (98–100) |
| >=0.035 | 40 | 76 (65–86) | 62 (61–62) | 8 (6–10) | 98 (97–99) |
| >=0.058 | 20 | 53 (41–66) | 81 (81–82) | 11 (8–15) | 98 (97–98) |
| **CARDIA-10 Simple Model** |  |  |  |  |  |
| Pr(DM) | AUROC = 0.80 (0.75 - 0.85) |
| >=0.017 | 80 | 96 (91–100) | 21 (21–22) | 8 (6–10) | 99 (97–100) |
| >=0.029 | 60 | 90 (83–96) | 42 (41–43) | 10 (8–12) | 98 (97–99) |
| >=0.048 | 40 | 82 (73–90) | 63 (62–64) | 14 (11–17) | 98 (97–99) |
| >=0.092 | 20 | 64 (53–75) | 83 (82–84) | 21 (16–27) | 97 (96–98) |
| **CARDIA-10 Enhanced Model** |  |  |  |  |  |
| Pr(DM) | AUROC = 0.86 (0.81 - 0.90) |
| >=0.009 | 80 | 98 (94–100) | 21 (21–22) | 8 (7–10) | 99 (98–100) |
| >=0.018 | 60 | 95 (89–100) | 43 (42–43) | 11 (8–13) | 99 (98–100) |
| >=0.035 | 40 | 87 (80–94) | 63 (63–64) | 15 (12–18) | 99 (98–99) |
| >=0.084 | 20 | 72 (62–82) | 84 (83–85) | 24 (18–30) | 98 (97–99) |
| **ARIC Simple Model** |  |  |  |  |  |
| Pr(DM) | AUROC = 0.75 (0.73 - 0.78) |
| >=0.037 | 80 | 97 (95–99) | 22 (22–22) | 11 (10–13) | 99 (98–99) |
| >=0.029 | 60 | 88 (85–91) | 43 (42–43) | 14 (12–15) | 97 (96–98) |
| >=0.048 | 40 | 74 (69–79) | 64 (63–64) | 17 (16–19) | 96 (95–97) |
| >=0.092 | 20 | 49 (44–54) | 83 (82–84) | 23 (20–26) | 94 (93–95) |
| **ARIC Enhanced Model** |  |  |  |  |  |
| Pr(DM) | AUROC = 0.85 (0.83 - 0.87) |
| >=0.012 | 80 | 99 (97–100) | 22 (22–22) | 12 (10–13) | 99 (99–100) |
| >=0.018 | 60 | 96 (93–98) | 44 (43–44) | 15 (14–17) | 99 (98–99) |
| >=0.035 | 40 | 88 (84–91) | 65 (64–66) | 21 (19–23) | 98 (97–99) |
| >=0.084 | 20 | 68 (63–73) | 85 (84–86) | 32 (29–36) | 96 (96–97) |
| **CHS Simple Model** |  |  |  |  |  |
| Pr(DM) | AUROC = 0.71 (0.65 - 0.77) |
| >=0.019 | 80 | 95 (89–100) | 21 (20–21) | 6 (4–7) | 99 (97–100) |
| >=0.029 | 60 | 86 (77–95) | 41 (41–42) | 7 (5–9) | 98 (97–99) |
| >=0.048 | 40 | 70 (58–81) | 62 (61–62) | 8 (6–11) | 98 (96–99) |
| >=0.092 | 20 | 44 (33–56) | 81 (81–82) | 11 (7–15) | 97 (95–98) |
| **CHS Enhanced Model** |  |  |  |  |  |
| Pr(DM) | AUROC = 0.82 (0.76 - 0.88) |
| >=0.007 | 80 | 96 (90–100) | 21 (20–21) | 6 (4–7) | 99 (97–100) |
| >=0.018 | 60 | 90 (82–97) | 42 (41–42) | 7 (5–9) | 99 (98–100) |
| >=0.035 | 40 | 84 (74–92) | 62 (62–63) | 10 (8–13) | 99 (98–99) |
| >=0.084 | 20 | 70 (59–81) | 83 (82–83) | 17 (12–22) | 98 (97–99) |

Pr(DM): probability of developing diabetes, derived from the prediction model, using the 20th, 40th, 60th and 80th percentiles of the distribution as cut off points.

A-3: Graphical representation of predicted probabilities across datasets over the entire support

Figures A-3a and A-3b expand the information presented in Table 4 of the paper by going beyond the mean and showing the entire range of predicted probabilities. We compared the predicted probability (first without [Figure A-3a] and then with calibration [Figure A-3b]) of each risk equation to the predicted probability in the target data (45 degree line). Perfect predictions are on the 45 degree line. By construction, the coefficients derived from the target sample generate predictive probabilities that perfectly match the predicted probability of the target sample. Predicted probabilities below the 45 degree line show under-predictions, and predicted probabilities above the 45 degree line show over-predictions from using risk equations in different age cohorts.

The annual risk from constrained simple coefficients shows a wider deviation from the 45-degree line (the predicted probability in the target data) than the annual risk from constrained enhanced coefficients. The higher the spread in predicted probability in the graph axis, the better the discriminating power of the model.

Figure A-3a: Comparing 1-year predicted probabilities across datasets forcing all coefficients including the constant term to be the same as in the source equations

|  |  |
| --- | --- |
| **Panel a. Simple model** | **Panel b. Enhanced model** |
|  |  |
|  |  |
|  |  |
|  |  |

Note: The straight line originating from the (0,0) point represents the 45% line and has a slope=1.

Figure A-3b: Comparing 1-year predicted probabilities across datasets after calibrating the constant of one cohort into others.

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|

|  |  |
| --- | --- |
| **Panel a. Simple model** | **Panel b. Enhanced model** |
|  |  |
|  |  |
|  |  |
|  |  |

 |  |

Note: The straight line originating from the (0,0) point represents the 45% line and has a slope=1.

A-4: Literature Review of Diabetes-Type Two Risk Factor Analyses

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Author (year)** | **Data Used** | **Size** | **Duration(years)** | **Max. variables used(Simple or Enhanced Eq.)** | **AUROC** |
| **Studies measuring future risk** |
| [Vassy et al. (2012](#_ENREF_20)) [1] | CARDIA | 2,439  | 23.9  | 9 (enhanced) | 0.82\* |
| [Kahn et al. (2009](#_ENREF_12)) [2] | ARIC | 12,729  | 14.9  | 13 (enhanced) | 0.79 |
| [Mozaffarian et al. (2009](#_ENREF_14)) [3] | CHS | 4,883  | 7.1  | 6 (simple) | N/A |
| [Schmidt et al. (2005](#_ENREF_17)) [4] | ARIC | 7,915  | 9  | 9 (simple and enhanced) | 0.80 |
| [Chen et al. (2010](#_ENREF_5)) [5] | Ausdiab | 6,060  | 5  | 13 (enhanced) | 0.78 |
| [Hippisley-Cox et al. (2009](#_ENREF_11)) [6] | QResearch | 2,540,753  | 6.5  | 9 (enhanced) | 0.85 for women, 0.83 for men |
| [Lindstrom and Tuomilehto (2003](#_ENREF_13)) [7] | FINRISK | 4,435  | 10  | 7 (simple) | 0.86 |
| [Stern, Williams, and Haffner (2002](#_ENREF_19)) [8] | San Antonio Heart Study | 2,803  | 7.5  | 13 (enhanced) | 0.84 |
| [Rahman et al. (2008](#_ENREF_15)) [9] | EPIC-Norfolk | 25,639  | 5  | 7 (simple) | 0.75 |
| [Wilson et al. (2007](#_ENREF_21)) [10] | Framingham Offspring Study | 3,140  | 7  | 9 (enhanced) | 0.85 |
| [Schulze et al. (2007](#_ENREF_18)) [11] | EPIC-Potsdam | 25,167  | 7.0  | 11 (enhanced) | 0.84 |
| **Studies measuring current prevalence** |
| [Bang et al. (2009](#_ENREF_2)) [12] | NHANES | 5,258  | cross-sectional  | 6 (simple) | 0.79 |
| [Cabrera de León et al. (2008](#_ENREF_4)) [13] | Canary Islands Survey | 3,119  | cross-sectional  | 5 (simple) | 0.84 for men, 0.87 for women |
| [Bindraban et al. (2008](#_ENREF_3)) [14] | SUNSET Study | 1,434  | cross-sectional  | 8 (simple) | 0.77 |
| [Heikes et al. (2008](#_ENREF_10)) [15] | NHANES | 6,009  | cross-sectional  | 8 (simple) | 0.85 |
| [Al-Lawati and Tuomilehto (2007](#_ENREF_1)) [16] | Oman National Diabetes Survey | 4,881  | cross-sectional  | 5 (simple) | 0.83 |
| [Ramachandran et al. (2005](#_ENREF_16)) [17] | National Urban Diabetes Survey | 4,993  | cross-sectional  | 7 (simple) | 0.80 |
| [Glümer et al. (2004](#_ENREF_8)) [18] | Inter99 | 3,250  | cross-sectional  | 6 (simple) | 0.80 |
| [Griffin et al. (2000](#_ENREF_9)) [19] | Ely and Wessex Studies | 1,274  | cross-sectional  | 7 (simple) | 0.80 |

\* C-statistic

A-5a: Log odds for type two diabetes over T years—Simple model

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Self-report + FPG** | **CARDIA (10 yrs)** | **CARDIA-10 (10 yrs)** | **ARIC (9 yrs)** | **CHS (7 yrs)** |
| Age groupƗ | 0.295\* | 0.217 | 0.076 | -0.164 |
| Black (=1) | -0.055 | 0.342\* | 0.280\*\*\* | 0.235 |
| Male (=1) | -0.958\*\*\* | 0.322\* | 0.454\*\*\* | 0.414\*\* |
| BMI  | 0.083\*\*\* | 0.134\*\*\* | 0.130\*\*\* | 0.135\*\*\* |
| Parental history (=1) | 0.507\*\*\* | 0.857\*\*\* | 0.626\*\*\* | 0.281 |
| Smoker  | -0.130 | -0.013 | 0.305\*\*\* | 0.181 |
| High SBP (>140 mm Hg=1) | 1.347\*\* | 0.090 | 0.386\*\*\* | 0.635\*\*\* |
| High cholesterol | 0.431 | 0.328\* | 0.002 | -0.054 |
| Constant | -5.171\*\*\* | -7.516\*\*\* | -6.662\*\*\* | -7.405\*\*\* |
| Observations | 4,039 | 2,813 | 8,875 | 3,094 |

Ɨ For CARDIA Age (25-30) ==1, For CARDIA-10 Age (35-40) ==1, for ARIC Age (55-64) ==1, for CHS Age (75+) ==1

PPV = p\*Sensitivity/(p\*Sensitivity+(1-p)(1-Specificity)) where p is the prevalence

NPV=Specificity\*(1-p)/(p\*(1-Sensitivity)+(1-p)\*Specificity)

BMI= Body Mass Index=(weight in kg)/ (height in meters)2

\*\*\* p<0.01, \*\* p<0.05, \* p<0.1

A.5b: Log odds for type two diabetes over T years—Enhanced model

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Self-report + FPG** | **CARDIA (10 yrs)** | **CARDIA-10 (10 yrs)** | **ARIC (9 yrs)** | **CHS (7 yrs)** |
| Age groupƗ | 0.278 | -0.019 | 0.108\*\*\* | -0.093 |
| Black (=1) | 0.160 | 0.441\*\* | 0.641\*\*\* | 0.572\*\* |
| Male (=1) | -1.232\*\*\* | -0.431\*\* | -0.229\*\*\* | 0.099 |
| Parental history (=1) | 0.551\*\*\* | 0.797\*\*\* | 0.513\*\*\* | 0.265 |
| SBP (mm Hg) | -0.005 | 0.020\*\*\* | 0.006\*\*\* | 0.005 |
| Smoker (=1) | -0.086 | -0.059 | 0.207 | 0.157 |
| BMI  | 0.057\*\*\* | 0.076\*\*\* | 0.087\*\*\* | 0.054\*\*\* |
| FPG (mg/dL) | 0.033\*\*\* | 0.077\*\*\* | 1.978\*\*\* | 0.108\*\*\* |
| HDL (mg/dL) | -0.010 | -0.021\*\* | -0.026\*\*\* | -0.014 |
| Triglycerides (mg/dL) | 0.005\*\*\* | 0.002\* | 0.002\*\*\* | 0.004\*\*\* |
| Constant | -6.571\*\*\* | -13.579\*\*\* | -15.997\*\*\* | -16.396\*\*\* |
| Observations | 4,039 | 2,813 | 8,875 | 3,094 |

Ɨ For CARDIA Age (25-30) ==1, For CARDIA-10 Age (35-40) ==1, for ARIC Age (55-64) ==1, for CHS Age (75+) ==1

PPV = p\*Sensitivity/(p\*Sensitivity+(1-p)(1-Specificity)) where p is the prevalence

NPV=Specificity\*(1-p)/(p\*(1-Sensitivity)+(1-p)\*Specificity)

BMI= Body Mass Index=(weight in kg)/ (height in meters)2

\*\*\* p<0.01, \*\* p<0.05, \* p<0.1

A-6: Producing annual probabilities from Logit models

Using logistic regression models:

The cumulative probability of diabetes is:



For each individual we estimate the predicted probability of developing diabetes by year t:

using the baseline values of X for individuals without diabetes, we assume a constant hazard to obtain the annual incidence of diabetes for each one of the age/risk cohorts:

For each one of the models we compute the annual incidence rate as follows:

From this, we calculate an average of across all individuals [20].

# Appendix REFERENCES

1. Vassy, J. L., N. H. Durant, E. K. Kabagambe, M. R. Carnethon, L. J. Rasmussen-Torvik, M. Fornage, C. E. Lewis, D. S. Siscovick, and J. B. Meigs. 2012. “A genotype risk score predicts type 2 diabetes from young adulthood: the CARDIA study.” Diabetologia 55(10): 2604-12.
2. Kahn, H. S., Y. J. Cheng, T. J. Thompson, G. Imperatore, and E. W. Gregg. 2009. “Two risk-scoring systems for predicting incident diabetes mellitus in U.S. adults age 45 to 64 years.” Ann Intern Med 150(11): 741-51
3. Mozaffarian, D., A. Kamineni, M. Carnethon, L. Djoussé, K. J. Mukamal, and D. Siscovick. 2009. “Lifestyle risk factors and new-onset diabetes mellitus in older adults: The cardiovascular health study.” Archives of Internal Medicine 169(8): 798-807.
4. Schmidt, M. I., B. B. Duncan, H. Bang, J. S. Pankow, C. M. Ballantyne, S. H. Golden, A. R. Folsom, and L. E. Chambless. 2005. “Identifying individuals at high risk for diabetes: The Atherosclerosis Risk in Communities study.” Diabetes Care 28(8): 2013-8.
5. Chen, L., D. J. Magliano, B. Balkau, S. Colagiuri, P. Z. Zimmet, A. M. Tonkin, P. Mitchell, P. J. Phillips, and J. E. Shaw. 2010. “AUSDRISK: an Australian Type 2 Diabetes Risk Assessment Tool based on demographic, lifestyle and simple anthropometric measures.” Med J Aust 192(4): 197-202.
6. Hippisley-Cox, J., C. Coupland, J. Robson, A. Sheikh, and P. Brindle. 2009. “Predicting risk of type 2 diabetes in England and Wales: prospective derivation and validation of QDScore.” BMJ 338: b880.
7. Lindstrom, J. and J. Tuomilehto. 2003. “The diabetes risk score: a practical tool to predict type 2 diabetes risk.” Diabetes Care 26(3): 725-31.
8. Stern, M. P., K. Williams, and S. M. Haffner. 2002. “Identification of persons at high risk for type 2 diabetes mellitus: do we need the oral glucose tolerance test?” Ann Intern Med 136(8): 575-81.
9. Rahman, M., R. K. Simmons, A. H. Harding, N. J. Wareham, and S. J. Griffin. 2008. “A simple risk score identifies individuals at high risk of developing Type 2 diabetes: a prospective cohort study.” Fam Pract 25(3): 191-6.
10. Wilson, P. W., J. B. Meigs, L. Sullivan, C. S. Fox, D. M. Nathan, and R. B. D'Agostino, Sr. 2007. “Prediction of incident diabetes mellitus in middle-aged adults: the Framingham Offspring Study.” Arch Intern Med 167(10): 1068-74.
11. Schulze, M. B., K. Hoffmann, H. Boeing, J. Linseisen, S. Rohrmann, M. Möhlig, A. F. Pfeiffer, J. Spranger, C. Thamer, and H.-U. Häring. 2007. “An accurate risk score based on anthropometric, dietary, and lifestyle factors to predict the development of type 2 diabetes.” Diabetes Care 30(3): 510-15.
12. Bang, H., A. M. Edwards, A. S. Bomback, C. M. Ballantyne, D. Brillon, M. A. Callahan, S. M. Teutsch, A. I. Mushlin, and L. M. Kern. 2009. “Development and validation of a patient self-assessment score for diabetes risk.” Annals of internal medicine 151(11): 775-83.
13. Cabrera de León A, Coello SD, Rodríguez Pérez Mdel C, Medina MB, Almeida González D, Díaz BB, de Fuentes MM, Aguirre-Jaime A. 2008. “A simple clinical score for type 2 diabetes mellitus screening in the Canary Islands”.Diabetes Res Clin Pract. 80(1):128-33.
14. Bindraban, N. R., I. G. van Valkengoed, G. Mairuhu, F. Holleman, J. B. Hoekstra, B. P. Michels, R. P. Koopmans, and K. Stronks. 2008. “Prevalence of diabetes mellitus and the performance of a risk score among Hindustani Surinamese, African Surinamese and ethnic Dutch: a cross-sectional population-based study.” BMC Public Health 8(1): 271.
15. Heikes, K. E., D. M. Eddy, B. Arondekar, and L. Schlessinger. 2008. “Diabetes Risk Calculator A simple tool for detecting undiagnosed diabetes and pre-diabetes.” Diabetes Care 31(5): 1040-45.
16. Al-Lawati, J. and J. Tuomilehto. 2007. “Diabetes risk score in Oman: a tool to identify prevalent type 2 diabetes among Arabs of the Middle East.” Diabetes Research and Clinical Practice 77(3): 438-44.
17. Ramachandran, A., C. Snehalatha, V. Vijay, N. J. Wareham, and S. Colagiuri. 2005. “Derivation and validation of diabetes risk score for urban Asian Indians.” Diabetes Research and Clinical Practice 70(1): 63-70.
18. Glümer, C., B. Carstensen, A. Sandbæk, T. Lauritzen, T. Jørgensen, and K. Borch-Johnsen. 2004. “A Danish Diabetes Risk Score for Targeted Screening: The Inter99 study.” Diabetes Care 27(3): 727-33.
19. Griffin, S. J., P. S. Little, C. N. Hales, A. L. Kinmonth, and N. J. Wareham. 2000. “Diabetes risk score: towards earlier detection of Type 2 diabetes in general practice.” Diabetes/Metabolism Research and Reviews 16(3): 164-71.
20. Briggs A., Claxton K., Sculpher M., Decision Modelling for Health Economic Evaluation 2006, Published by Oxford University Press