Use of Multiple Metabolic and Genetic Markers to Improve the Prediction of Type 2 Diabetes: the EPIC-Potsdam Study

Matthias B. Schulze, drph[1,2]
Cornelia Weikert, md[2,3]
Tobias Pischon, md[2]
Manuela M. Bergmann, phd[2]
Hadi Al-Hasani, phd[4]
Erwin Schleicher, phd[5]
Andreas Fritsche, md[5]
Hans-Ulrich Häring, md[5]
Heiner Boeing, phd[2]
Hans-Georg Joost, md[4]

OBJECTIVE — We investigated whether metabolic biomarkers and single nucleotide polymorphisms (SNPs) improve diabetes prediction beyond age, anthropometry, and lifestyle risk factors.

RESEARCH DESIGN AND METHODS — A case-cohort study within a prospective study was designed. We randomly selected a subcohort ($n = 2,500$) from 26,444 participants, of whom 1,962 were diabetes free at baseline. Of the 801 incident type 2 diabetes cases identified in the cohort during 7 years of follow-up, 579 remained for analyses after exclusions. Prediction models were compared by receiver operatoring characteristic (ROC) curve and integrated discrimination improvement.

RESULTS — Case-control discrimination by the lifestyle characteristics (ROC-AUC: 0.8465) improved with plasma glucose (ROC-AUC: 0.8672, $P < 0.001$) and A1C (ROC-AUC: 0.8859, $P < 0.001$). ROC-AUC further improved with HDL cholesterol, triglycerides, γ-glutamyltransferase, and alanine aminotransferase (0.9000, $P = 0.002$). Twenty SNPs did not improve discrimination beyond these characteristics ($P = 0.69$).

CONCLUSIONS — Metabolic markers, but not genotyping for 20 diabetogenic SNPs, improve discrimination of incident type 2 diabetes beyond lifestyle risk factors.

Diabetes Care 32:2116–2119, 2009

Accurate identification of individuals who are at increased risk for type 2 diabetes is a requirement for a targeted prevention. We therefore tested whether metabolic and genetic markers add substantial prognostic information to age, anthropometry, and lifestyle characteristics.

RESEARCH DESIGN AND METHODS

The European Prospective Investigation into Cancer and Nutrition (EPIC)-Potsdam study involves 27,548 participants (16,644 women, mainly aged 35–65 years, and 10,904 men, mainly aged 40–65 years) recruited from the general population in Potsdam, Germany, between 1994 and 1998. Follow-up questionnaires were sent out every 2–3 years to identify incident cases of type 2 diabetes (response rates 93–97%), and self-reports were verified by questionnaires mailed to physicians. Informed consent was obtained from participants; approval was given by the ethics committee of the State of Brandenburg, Germany.

A prospective case-cohort study was designed (1) (supplemental Fig. 1, which can be found in the online appendix [available at http://care.diabetesjournals.org/cgi/content/full/dc09-0197/DC1]). Of 2,500 individuals randomly selected from 26,444 participants with collected blood, 1,962 remained after exclusion of prevalent diabetes, self-reported but unverified diabetes during follow-up, missing biomarker data, abnormal plasma glucose, or more than four missing genotypes. Of 801 incident cases identified in the full cohort with blood samples (mean follow-up 7.1 years), 579 remained for analyses after exclusions.

We used baseline information on age, waist circumference, height, history of hypertension, physical activity, smoking, and consumption of red meat, whole-grain bread, coffee, and alcohol to compute the German Diabetes Risk Score (DRS), a prediction model previously described (2). Measurement of glucose, HDL cholesterol, triglycerides, γ-glutamyltransferase, alanine aminotransferase, high-sensitivity C-reactive protein (hs-CRP), and A1C followed standard procedures (1). Total adiponectin was measured with an ELISA (LINCO Research, St. Charles, MO). Genotyping of 20 single nucleotide polymorphisms (SNPs) related to diabetes risk (3–6) (supplemental Tables 1–2) was performed with TaqMan technology (Applied Biosystems, Foster City, CA). The genotyping error was ≤0.5%, and genotype distributions were in Hardy-Weinberg equilibrium ($P > 0.05$). We computed an unweighted count genetic score assuming an additive genetic model for each SNP, applying a linear weighting of 0, 1, and 2 to genotypes containing 0, 1, or 2 risk alleles, respectively. Scores for individuals with missing genotypes were standardized to those for individuals with complete data (7).

We evaluated different prediction models through receiver operatoring characteristics (ROCs) based on logistic regression models comparing the area under the curve (AUC) of the fuller model with that of the sparser model (8). Model calibration was tested by Hosmer-

From the [1]Public Health Nutrition Unit, Technische Universität München, Freising, Germany; the [2]Department of Epidemiology, German Institute of Human Nutrition Potsdam-Rehbruecke, Nuthetal, Germany; the [3]Institute for Social Medicine, Epidemiology, and Health Economics, Charité University Medicine, Berlin, Germany; the [4]Department of Pharmacology, German Institute of Human Nutrition Potsdam-Rehbruecke, Nuthetal, Germany; and the [5]Division of Endocrinology, Diabetology, Nephrology, Vascular Disease and Clinical Chemistry, the Department of Internal Medicine, University of Tübingen, Tübingen, Germany.

Corresponding author: Matthias B. Schulze, matthias.schulze@wzw.tum.de.

Received 3 February 2009 and accepted 12 July 2009. Published ahead of print at http://care.diabetesjournals.org on 31 August 2009. DOI: 10.2337/dc09-0197.

Table 1—*Relative contribution of the German DRS and biochemical and genetic markers to prediction of type 2 diabetes risk*

	ROC*		IDI†	
	C statistic (95% CI)	*P*	Absolute IDI (95% CI)	Relative IDI (%)
DRS only‡	0.8465 (0.8299–0.8630)	Ref.	Ref.	Ref.
DRS and A1C	0.8859 (0.8716–0.9003)	<0.0001	0.0974 (0.0792–0.1155)	34.11
DRS and glucose	0.8672 (0.8515–0.8830)	<0.0001	0.0553 (0.0407–0.0699)	19.37
DRS and A1C	0.8859 (0.8716–0.9003)	Ref.	Ref.	Ref.
DRS, A1C, and glucose	0.8926 (0.8785–0.9067)	0.0040	0.0230 (0.0135–0.0325)	6.01
DRS and glucose	0.8672 (0.8515–0.8830)	Ref.	Ref.	Ref.
DRS, glucose, and A1C	0.8926 (0.8785–0.9067)	<0.0001	0.0651 (0.0506–0.0797)	19.11
DRS, glucose, and A1C	0.8926 (0.8785–0.9067)	Ref.	Ref.	Ref.
DRS, glucose, A1C, triglycerides, HDL cholesterol, γ-glutamyltransferase, and alanine aminotransferase	0.9000 (0.8862–0.9137)	0.0022	0.0223 (0.0142–0.0304)	5.50
DRS, glucose, A1C, and genetic markers§	0.8928 (0.8787–0.9070)	0.7361	0.0014 (−0.0010–0.0039)	0.36
DRS, glucose, A1C, triglycerides, HDL cholesterol, γ-glutamyltransferase, and alanine aminotransferase	0.9000 (0.8862–0.9137)	Ref.	Ref.	Ref.
DRS, glucose, A1C, triglycerides, HDL cholesterol, γ-glutamyltransferase, alanine aminotransferase, and adiponectin	0.9023 (0.8887–0.9158)	0.0471	0.0064 (0.0022–0.0107)	1.50
DRS, glucose, A1C, triglycerides, HDL cholesterol, γ-glutamyltransferase, alanine aminotransferase, and hs-CRP	0.9016 (0.8880–0.9151)	0.1523	0.0029 (−0.0007–0.0066)	0.69
DRS, glucose, A1C, triglycerides, HDL cholesterol, γ-glutamyltransferase, alanine aminotransferase, and genetic markers	0.9002 (0.8865–0.9140)	0.6868	0.0015 (−0.0010–0.0039)	0.34

*The ROC curve is a plot of sensitivity versus false-positive rate across all possible cut points for a continuous predictor or prediction model. The area under the ROC curve (C statistic) is a measure of discrimination between case patients and control participants based on ranks and reflects the probability that the predicted risk is higher for a case subject than for a control subject. It ranges from 0.5 (no predictive ability) to a theoretical maximum of 1 (perfect discrimination)—the latter achieved if the scores or predicted risks for all case subjects are higher than those for all control subjects. †IDI is the difference between two models in discrimination slopes, which reflect the mean difference in predicted risk between case and control subjects. Instead of the difference, relative IDI expresses the discrimination slope of the more extensive model (e.g., including a new marker) as proportional increase compared with the discrimination slope of the basic model. ‡The German DRS combines baseline information on several risk factors to estimate the risk of developing type 2 diabetes (ref. 2). It is computed as follows: German DRS = 7.4 × waist (cm) − 2.4 × height (cm) + 4.3 × age (years) + 46 × hypertension (self-report) + 49 × red meat (each 150 g/day) − 9 × whole-grain bread (each 50 g/day) − 4 × coffee (each 150 g/day) × 20 × moderate alcohol (between 10 and 40 g/day) × 2 × physical activity (h/week) + 24 × former smoker + 64 × current heavy smoker (≥20 cigarettes/day). §Unweighted count genetic score of 20 SNPs assuming an additive genetic model for each SNP and applying a linear weighting of 0, 1, and 2 to genotypes containing 0, 1, or 2 risk alleles. Participants were excluded if they had five or more genotypes missing. Scores for individuals with missing genotypes were standardized to those of individuals with complete data.

Lemeshow tests (9). Reclassification was evaluated by the integrated discrimination improvement (IDI) (10). Analyses were performed with SAS (version 9.1; SAS Institute, Cary, NC). *P* values are two-tailed; $P < 0.05$ was considered statistically significant.

RESULTS — Baseline characteristics of the random subcohort and incident cases are presented in supplemental Table 3. ROC-AUC increased significantly when A1C or glucose was incorporated into a model with the German DRS (Table 1), most notably for A1C (from 0.8464 to 0.8862). *P* values for Hosmer-Lemeshow tests indicated better model calibration when A1C ($P = 0.1181$) or glucose ($P = 0.3823$) was included compared with a model with the German DRS alone ($P = 0.0157$). Measuring both glucose and A1C improved case-control discrimination. Also, blood lipids, γ-glutamyltransferase, and alanine aminotransferase significantly improved discrimination beyond the German DRS, A1C, and glucose (ROC-AUC: 0.9000). Moreover, IDI, as a marker of improved risk classification, was significantly different from zero (Table 1). In contrast to hs-CRP, additional information on adiponectin improved ROC-AUC. However, relative IDI was rather small (1.5%).

Diabetes risk increased with increasing number of prevalent risk alleles (supplemental Fig. 2). When genetic information was included along with the German DRS and metabolic markers, improvements of ROC-AUC and IDI were small and nonsignificant (Table 1).

CONCLUSIONS — Numerous diabetes prediction models have been developed, but few studies have investigated whether metabolic markers improve prediction beyond conventional risk factors. In the Atherosclerosis Risk in Communities study, clinical variables combined with fasting plasma glucose discriminated

better compared with clinical variables only (11). Further improvement in discrimination was observed with HDL cholesterol and triglycerides. In the Framingham Offspring Study, a model involving additional information on hypertension, fasting plasma glucose, HDL cholesterol, and triglycerides performed substantially better than a model including age, sex, BMI, and parental history (12). Our results support and extend these findings by indicating that a comprehensive basic model including important lifestyle risk factors such as physical activity, smoking, alcohol consumption, and diet is significantly improved by glucose, A1C, HDL cholesterol, triglycerides, and liver enzymes but not hs-CRP, adiponectin, or genetic markers.

Our observation that multiple SNPs do not substantially improve discrimination beyond age, sex, and clinical markers confirms previous studies (7,13–15). We extended their results by using a comprehensive set of anthropometric, nutritional, and lifestyle variables in the basic model; by including additional biomarkers (adiponectin, hs-CRP, and A1C); and by using the full set of currently confirmed diabetogenic SNPs. It should be noted that the prospective design rendered it necessary to exclude prevalent diabetes cases at baseline. Thus, results reflect genetic prediction in middle-aged individuals but not prediction at birth. Furthermore, we did not consider gene-gene interactions.

In conclusion, our study indicates that both plasma glucose and A1C considerably improve discrimination of incident type 2 diabetes by age, anthropometry, and lifestyle characteristics (DRS). HDL cholesterol, triglycerides, γ-glutamyltransferase, and alanine aminotransferase, but not 20 diabetogenic SNPs, further improve discrimination.

Acknowledgments— The recruitment phase of the EPIC-Potsdam study was supported by the Federal Ministry of Science, Germany (01 EA 9401), and the European Union (SOC 95201408 05F02). The follow-up of the EPIC-Potsdam study was supported by the German Cancer Aid (70-2488-Ha I) and the European Community (SOC 98200769 05F02). The present study was supported by grants from the European Union (EUGENE2: LSHM-CT 204 512013) and the German Ministry of Science and Technology (NGFN2: 01GS0487).

No potential conflicts of interest relevant to this article were reported.

Parts of this study were accepted for presentation in abstract form at the 45th Annual Meeting of the European Association for the Study of Diabetes, Vienna, Austria, 29 September–2 October 2009.

We gratefully acknowledge the excellent technical assistance of Anna Bury, Sabina Herbert, Albrecht Pfäfflin, and Ingo Besenthal, who were involved in the biochemical analyses, and of Frank Döring, who was involved in the genetic analyses. We also thank Kay Behling, Kathrein Kühn, Birgit Czullay, Birgit Schmidtchen, Ellen Kohlsdorf, and Wolfgang Bernigau for their efforts in data collection and management, as well as Janine Kröger for computational assistance.

References

1. Stefan N, Fritsche A, Weikert C, Boeing H, Joost HG, Haring HU, Schulze MB. Plasma fetuin-A levels and the risk of type 2 diabetes. Diabetes 2008;57:2762–2767
2. Schulze MB, Hoffmann K, Boeing H, Linseisen J, Rohrmann S, Mohlig M, Pfeiffer AF, Spranger J, Thamer C, Haring HU, Fritsche A, Joost HG. An accurate risk score based on anthropometric, dietary, and lifestyle factors to predict the development of type 2 diabetes. Diabetes Care 2007;30:510–515
3. Gudmundsson J, Sulem P, Steinthorsdottir V, Bergthorsson JT, Thorleifsson G, Manolescu A, Rafnar T, Gudbjartsson D, Agnarsson BA, Baker A, Sigurdsson A, Benediktsdottir KR, Jakobsdottir M, Blondal T, Stacey SN, Helgason A, Gunnarsdottir S, Olafsdottir A, Kristinsson KT, Birgisdottir B, Ghosh S, Thorlacius S, Magnusdottir D, Stefansdottir G, Kristjansson K, Bagger Y, Wilensky RL, Reilly MP, Morris AD, Kimber CH, Adeyemo A, Chen Y, Zhou J, So WY, Tong PC, Ng MC, Hansen T, Andersen G, Borch-Johnsen K, Jorgensen T, Tres A, Fuertes F, Ruiz-Echarri M, Asin L, Saez B, van Boven E, Klaver S, Swinkels DW, Aben KK, Graif T, Cashy J, Suarez BK, van Vierssen Trip O, Frigge ML, Ober C, Hofker MH, Wijmenga C, Christiansen C, Rader DJ, Palmer CN, Rotimi C, Chan JC, Pedersen O, Sigurdsson G, Benediktsson R, Jonsson E, Einarsson GV, Mayordomo JI, Catalona WJ, Kiemeney LA, Barkardottir RB, Gulcher JR, Thorsteinsdottir U, Kong A, Stefansson K. Two variants on chromosome 17 confer prostate cancer risk, and the one in TCF2 protects against type 2 diabetes. Nat Genet 2007; 39:977–983
4. Saxena R, Voight BF, Lyssenko V, Burtt NP, de Bakker PI, Chen H, Roix JJ, Kathiresan S, Hirschhorn JN, Daly MJ, Hughes TE, Groop L, Altshuler D, Almgren P, Florez JC, Meyer J, Ardlie K, Bengtsson Bostrom K, Isomaa B, Lettre G, Lindblad U, Lyon HN, Melander O, Newton-Cheh C, Nilsson P, Orho-Melander M, Rastam L, Speliotes EK, Taskinen MR, Tuomi T, Guiducci C, Berglund A, Carlson J, Gianniny L, Hackett R, Hall L, Holmkvist J, Laurila E, Sjogren M, Sterner M, Surti A, Svensson M, Tewhey R, Blumenstiel B, Parkin M, Defelice M, Barry R, Brodeur W, Camarata J, Chia N, Fava M, Gibbons J, Handsaker B, Healy C, Nguyen K, Gates C, Sougnez C, Gage D, Nizzari M, Gabriel SB, Chirn GW, Ma Q, Parikh H, Richardson D, Ricke D, Purcell S. Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels. Science 2007;316: 1331–1336
5. Zeggini E, Weedon MN, Lindgren CM, Frayling TM, Elliott KS, Lango H, Timpson NJ, Perry JR, Rayner NW, Freathy RM, Barrett JC, Shields B, Morris AP, Ellard S, Groves CJ, Harries LW, Marchini JL, Owen KR, Knight B, Cardon LR, Walker M, Hitman GA, Morris AD, Doney AS, McCarthy MI, Hattersley AT. Replication of genome-wide association signals in UK samples reveals risk loci for type 2 diabetes. Science 2007;316:1336–1341
6. Zeggini E, Scott LJ, Saxena R, Voight BF, Marchini JL, Hu T, de Bakker PI, Abecasis GR, Almgren P, Andersen G, Ardlie K, Bostrom KB, Bergman RN, Bonnycastle LL, Borch-Johnsen K, Burtt NP, Chen H, Chines PS, Daly MJ, Deodhar P, Ding CJ, Doney AS, Duren WL, Elliott KS, Erdos MR, Frayling TM, Freathy RM, Gianniny L, Grallert H, Grarup N, Groves CJ, Guiducci C, Hansen T, Herder C, Hitman GA, Hughes TE, Isomaa B, Jackson AU, Jorgensen T, Kong A, Kubalanza K, Kuruvilla FG, Kuusisto J, Langenberg C, Lango H, Lauritzen T, Li Y, Lindgren CM, Lyssenko V, Marvelle AF, Meisinger C, Midthjell K, Mohlke KL, Morken MA, Morris AD, Narisu N, Nilsson P, Owen KR, Palmer CN, Payne F, Perry JR, Pettersen E, Platou C, Prokopenko I, Qi L, Qin L, Rayner NW, Rees M, Roix JJ, Sandbaek A, Shields B, Sjogren M, Steinthorsdottir V, Stringham HM, Swift AJ, Thorleifsson G, Thorsteinsdottir U, Timpson NJ, Tuomi T, Tuomilehto J, Walker M, Watanabe RM, Weedon MN, Willer CJ, Illig T, Hveem K, Hu FB, Laakso M, Stefansson K, Pedersen O, Wareham NJ, Barroso I, Hattersley AT, Collins FS, Groop L, McCarthy MI, Boehnke M, Altshuler D. Meta-analysis of genome-wide association data and large-scale replication identifies additional susceptibility loci for type 2 diabetes. Nat Genet 2008;40:638–645
7. Cornelis MC, Qi L, Zhang C, Kraft P, Manson J, Cai T, Hunter DJ, Hu FB. Joint effects of common genetic variants on the risk for type 2 diabetes in U.S. men and women of European ancestry. Ann Intern Med 2009;150:541–550
8. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating char-

acteristic curves: a nonparametric approach. Biometrics 1988;44:837–845
9. Hosmer DWJ, Lemeshow S. *Applied Logistic Regression.* 2nd ed., chapter 5. New York, John Wiley & Sons, 2000, p. 147–156
10. Pencina MJ, D'Agostino RB Sr, D'Agostino RB Jr, Vasan RS. Evaluating the added predictive ability of a new marker: from area under the ROC curve to reclassification and beyond (Letter). Stat Med 2008; 27:207–212
11. Schmidt MI, Duncan BB, Bang H, Pankow JS, Ballantyne CM, Golden SH, Folsom AR, Chambless LE. Identifying individuals at high risk for diabetes: the Atherosclerosis Risk in Communities study. Diabetes Care 2005;28:2013–2018
12. 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. Arch Intern Med 2007;167:1068–1074
13. Lyssenko V, Jonsson A, Almgren P, Pulizzi N, Isomaa B, Tuomi T, Berglund G, Altshuler D, Nilsson P, Groop L. Clinical risk factors, DNA variants, and the development of type 2 diabetes. N Engl J Med 2008;359:2220–2232
14. Meigs JB, Shrader P, Sullivan LM, McAteer JB, Fox CS, Dupuis J, Manning AK, Florez JC, Wilson PW, D'Agostino RB Sr, Cupples LA. Genotype score in addition to common risk factors for prediction of type 2 diabetes. N Engl J Med 2008;359: 2208–2219
15. van Hoek M, Dehghan A, Witteman JC, van Duijn CM, Uitterlinden AG, Oostra BA, Hofman A, Sijbrands EJ, Janssens AC. Predicting type 2 diabetes based on polymorphisms from genome-wide association studies: a population-based study. Diabetes 2008;57:3122–3128