Published in final edited form as:

Am J Prev Med. 2017 August ; 53(2): 201–209. doi:10.1016/j.amepre.2017.04.019.

Individual- and Area-level SES in Diabetes Risk Prediction: The Multi-Ethnic Study of Atherosclerosis

Paul J. Christine, MPH¹, Rebekah Young, PhD², Sara D. Adar, ScD, MHS¹, Alain G. Bertoni, MD, MPH³, Michele Heisler, MD^{4,5}, Mercedes R. Carnethon, PhD⁶, Rodney A. Hayward, MD^{4,5}, and Ana V. Diez Roux, MD, PhD⁷

¹Department of Epidemiology, University of Michigan School of Public Health, Ann Arbor, Michigan

²Department of Biostatistics, University of Washington School of Public Health, Seattle, Washington

³Department of Epidemiology and Prevention, Wake Forest School of Medicine, Winston-Salem, North Carolina

⁴Veterans Affairs Center for Clinical Management Research, Ann Arbor, Michigan

⁵Department of Internal Medicine, University of Michigan, Ann Arbor, Michigan

⁶Department of Preventive Medicine, Northwestern University Feinberg School of Medicine, Chicago, Illinois

⁷Department of Epidemiology and Biostatistics, Drexel University School of Public Health, Philadelphia, Pennsylvania

Abstract

Introduction—The purpose of this study was to evaluate if adding SES to risk prediction models based upon traditional risk factors improves the prediction of diabetes.

Methods—Risk prediction models without and with individual- and area-level SES predictors were compared using the prospective Multi-Ethnic Study of Atherosclerosis. Cox proportional hazards models were utilized to estimate hazard ratios for SES predictors and to generate 10-year predicted risks for 5,021 individuals without diabetes at baseline followed from 2000 to 2012. C-statistics were used to compare model discrimination, and the proportion of individuals reclassified into higher or lower risk categories with the addition of SES predictors was calculated. The accuracy of risk prediction by SES was assessed by comparing observed and predicted risks across tertiles of the SES variables. Statistical analyses were performed in 2015–2016.

Address correspondence to: Paul J. Christine, MPH, Center for Social Epidemiology and Population Health, Department of Epidemiology, 1415 Washington Heights, 3rd Floor Tower, University of Michigan School of Public Health, Ann Arbor MI 48109-2029. pjchris@umich.edu.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Results—Over a median of 9.2 years of follow-up, 615 individuals developed diabetes. Individual- and area-level SES variables did not significantly improve model discrimination or reclassify substantial numbers of individuals across risk categories. Models without SES predictors generally underestimated risk for low-SES individuals or individuals residing in low-SES areas (underestimates ranging from 0.31% to 1.07%) and overestimated risk for high-SES individuals or individuals residing in high-SES areas (overestimates ranging from 0.70% to 1.30%), and the addition of SES variables largely mitigated these differences.

Conclusions—Standard diabetes risk models may underestimate risk for low-SES individuals and overestimate risk for those of high SES. Adding SES predictors helps correct this systematic misestimation, but may not improve model discrimination.

INTRODUCTION

In the U.S., an estimated 18.8 million adults have diagnosed Type 2 diabetes mellitus (T2DM).¹ If current trends of increasing prevalence and incidence persist, as many as one in three could have T2DM by 2050.² Thus, there is a need for cost-effective prevention strategies, including the identification of individuals at heightened risk for developing disease.^{3,4}

In light of evidence that nearly two-thirds of T2DM cases in high-risk individuals can be prevented through behavioral and pharmacologic interventions,⁵ risk prediction models are increasingly recommended for clinical practice and public health planning.^{3,4,6} Using clinical information, these models complement traditional approaches for identifying high-risk individuals (e.g., hemoglobin A1c) and may help with earlier identification.⁶

With the growth of risk prediction modeling to inform preventive interventions,⁴ much research has focused upon improving prediction models. For T2DM, this has often involved including novel biological information (e.g., genes).^{7,8} Relatively little research has explored the added value of using social and environmental information (e.g., individual- and area-level SES) in diabetes prediction models, with the exception of several models outside the U.S.^{9–11} This is surprising given pronounced disparities in T2DM by SES and geographic area,¹² and recognition of the importance of social and area-level factors in the development of diabetes.^{13,14} SES may capture aspects of diabetes risk, including psychosocial stress and health behaviors, which are not accounted for by traditional risk factors (Appendix Figure 1).^{15,16} Furthermore, ignoring SES in cardiovascular risk models has been shown to underestimate risk for low-SES individuals, yet few studies investigating this question exist for T2DM.^{17,18} Given the growth of electronic medical records and their ability to link to environmental data,¹⁹ evaluations of the utility of including multilevel social information in risk assessment are warranted.

Using a large, multi-ethnic, prospective cohort, this study investigates if adding individual- and area-level SES to traditional risk prediction models: (1) aids in the discrimination of people who developed diabetes from those who remained diabetes free; (2) improves prediction accuracy (i.e., model calibration) across the SES distribution; and (3) results in meaningful risk reclassification.

METHODS

Study Population

The Multi-Ethnic Study of Atherosclerosis (MESA) is a prospective cohort of 6,814 non-institutionalized adults (aged 45–84 years) who self-identified as white, black, Hispanic, or Chinese. Participants were free of clinical cardiovascular disease at baseline and were recruited from six U.S. locations (New York, NY; Baltimore, MD; Forsyth County, NC; Chicago, IL; St. Paul, MN; and Los Angeles, CA). The first examination occurred in 2000–2002, and four follow-up exams were completed an average of 1.6, 3.1, 4.8, and 9.5 years later. Written informed consent was obtained from participants, and the study was approved by IRBs at each site.

The analyses excluded individuals with T2DM at baseline ($n=859$), those who did not agree to have their address geocoded ($n=500$), were missing data on diabetes risk factors ($n=404$), or were without follow-up information regarding diabetes status ($n=29$). The final analytic sample consisted of 5,021 adults.

Measures

Incident T2DM was determined at each follow-up exam using the American Diabetes Association's 2003 criteria²⁰: fasting plasma glucose level ≥ 126 mg/dL (7 mmol/L) or use of oral anti-hyperglycemic medications or insulin. Information on oral medication and insulin was obtained via self-report and inspection of medication bottles. Individuals were considered at-risk until diagnosis of diabetes, last follow-up, or administrative censoring at Exam 5. Time of onset was assigned to the midpoint between the last disease-free examination and the examination with diagnosed diabetes.

Non-laboratory risk factors measured at baseline included: age, sex, family history of T2DM, systolic blood pressure (SBP; mmHg), treatment for hypertension (yes/no), waist circumference (cm), waist to hip ratio, BMI (kg/m^2), height (m), and smoking status (yes/no). Self-identified race/ethnicity, utilized in many risk scores, was also available at baseline. For laboratory-based risk factors, 12-hour fasting blood samples were utilized to measure plasma glucose levels (mg/dL), triglycerides (mmol/L), and high-density lipoprotein cholesterol (mmol/L).

To ensure usefulness in clinical settings, only simple SES measures that could be feasibly incorporated into medical records or public health databases were employed. At the individual level, this included education level by grade/degree, annual household income, and annual household income per supported individual ("household income per capita"), all assessed via questionnaire at baseline. Because individual-level SES variables are not always available and area-level features are independent risk factors for disease,²¹ census tract-level median household income, percentage of adults aged ≥ 25 years with a bachelor's degree or higher, and percentage of people living below the federal poverty threshold were also evaluated and matched to participant addresses using the *U.S. Census 2000*. In MESA and other studies, these variables have been associated with other area-level risk factors for T2DM, such as the availability of healthy food and physical activity resources,^{22,23} perceived safety,²⁴ and air pollution levels.^{25,26}

Individual-level, area-level, and composite SES indices were also created. Following recommended methods, the indices were created by transforming variables to remove skewness, z-scoring the variables, and then summing the z-scores.²⁷ Based upon prior work,^{28–30} education and annual household income were included a priori at the individual level. At the census tract level, the index included a priori percentage of adults with a high school education, percentage of adults with a bachelor's degree, median household income, median home value, percentage of adults in a managerial occupation, and percentage of households with income from interest, dividends, or rental properties. The combined SES index was created by summing the internally standardized individual- and area-level indices. Indices were created so that an increasing score indicates greater SES. Most SES variables were weakly correlated with traditional risk factors (strongest Pearson correlation coefficient, $r = -0.22$).

Statistical Analysis

Participant sociodemographic characteristics and diabetes risk factors were assessed overall and by incident diabetes status. Two separate traditional prediction models were then developed using Cox proportional hazards models to estimate 10-year incident diabetes risk, one employing only non-laboratory variables available from a routine medical visit (hereafter, the “Clinical Model”), and one also using laboratory variables (hereafter, the “Laboratory Model”). Of note, the term “risk” refers to the predicted probability of developing T2DM within a 10-year period. The models included risk factors from the Atherosclerosis Risk in Communities (ARIC) diabetes risk model³¹ (age, family history of diabetes, race/ethnicity, SBP, height, waist circumference, fasting glucose, fasting triglycerides, high-density lipoprotein cholesterol) as well as risk factors shown to be predictive in at least three other diabetes risk scores (BMI, waist to hip ratio, smoking status).⁴ Although the ARIC model performs well in MESA and other cohorts,^{28,32} an internal prediction model was developed to ensure that any improvements in model performance with SES variables was not due to simply improving the fit of poorly modeled variables from an external prediction model. For parsimony, only predictors that were marginally significant ($p < 0.1$) were retained and the Akaike Information Criterion was used to decide between highly correlated variables. Penalized b-splines (2 df) were included for continuous predictors with evidence of non-linearity³³ and the proportional hazards assumption was confirmed for all models. The final Clinical Model included age (spline), race/ethnicity, family history of T2DM, SBP, waist circumference, and hypertension treatment, and the Laboratory Model added fasting glucose and triglycerides (log transformed, spline).

The value of adding each SES variable to both the Clinical and Laboratory Models was evaluated. Model discrimination was assessed using likelihood ratio tests of nested models, C-statistics, and receiver operating characteristic curves for survival models.^{34–36} Collectively, these statistics reflect the ability of the models to differentiate who will and will not develop T2DM.³⁷ Model calibration across categories of SES variables was investigated by calculating mean differences between predicted risks generated from proportional hazards models and observed risks calculated using Kaplan–Meier analyses (Appendix Text 1 provides details).^{38,39} Predicted and observed risks were compared across

tertiles of SES because they matched natural categories for variables like education, but quintiles were also used in sensitivity analyses. For brevity, the calibration of models with and without SES predictors were compared for only the two individual- and area-level SES variables most predictive of diabetes. Finally, risk reclassification in models with and without SES predictors was assessed using a priori 10-year risk categories of <10%, 10%–20%, and >20%, as has been done for cardiovascular disease. The number of individuals classified in each category was calculated, and models with and without the two most predictive individual- and area-level SES variables were compared.

Several sensitivity analyses were performed, including an evaluation of models with both individual- and area-level SES variables; interactions between SES predictors and age, sex, and race/ethnicity; models without race/ethnicity; and use of the original ARIC³¹ and Framingham Offspring Study⁴⁰ risk scores to evaluate calibration across SES categories with and without the addition of SES variables.

All analyses were conducted in R, version 3.1.2 and STATA, version 12.1 and completed in 2015–2016 (Appendix Text 2 includes abbreviated code).

RESULTS

Of 5,021 individuals without diabetes at baseline, 615 developed diabetes over a median of 9.2 years of follow-up. Compared with individuals who did not develop diabetes, individuals who developed diabetes were more likely to be African American and Hispanic, have a family history of diabetes, and be obese (Table 1). Individuals developing diabetes also had higher baseline values of traditional diabetes risk factors, including fasting plasma glucose, triglycerides, waist circumference, and SBP. With respect to SES, those developing diabetes were less educated and had lower annual household incomes than those remaining diabetes-free. At the area level, individuals with incident diabetes resided in neighborhoods with lower education levels, lower median household incomes, and greater proportions of people living below the poverty line. SES indices at the individual and area level were lower (indicating lower SES) for individuals who developed diabetes.

Table 2 displays the association of each SES predictor with diabetes as well as model discrimination for multivariable Clinical and Laboratory prediction models. All SES variables were highly predictive of incident diabetes in single predictor models (all $p < 0.01$, data not shown), but the associations attenuated considerably in multivariable models. The strongest SES predictors in the Clinical Model were the individual-level SES index (hazard ratio [HR] per SD increase=0.91, 95% CI=0.82, 1.00) and the area-level percentage of adults with a bachelor's degree (HR per SD increase=0.91, 95% CI=0.83, 1.01). In the Laboratory Model, the strongest individual-level predictor was categorical household income (HR comparing highest to lowest category=0.75, 95% CI=0.58, 0.95), though continuous household income, and household income per capita were also predictive. At the area level, the SES index was most predictive (HR per SD increase=0.91, 95% CI=0.83, 1.01). In both the Clinical and Laboratory Models, the SES index combining individual- and area-level measures was predictive, but no more so than individual SES measures. The addition of SES predictors did not alter associations of traditional risk factors with diabetes, except for race/

ethnicity, which was modestly attenuated after the addition of SES predictors (HRs for all predictors are listed in Appendix Tables 1 and 2, and Appendix Figure 2).

Despite the significance of several SES predictors in the multivariable models, none of them significantly altered the model's discrimination (Table 2). The C-statistic went unchanged with the addition of SES predictors, and the receiver operating characteristic plots were overlapping (Appendix Figure 3).

The calibration of the Clinical and Laboratory Models across tertiles of the most consistent SES predictors is shown in Figures 1 and 2, respectively. For the Clinical Model excluding SES predictors, observed risks were generally higher than predicted risks for the lowest SES tertile, whereas the reverse was true for the highest SES tertile. For instance, those in neighborhoods with the lowest educational attainment had observed risks that were on average 1.06% (95% CI=0.54, 1.57) higher than the predicted risk, whereas those in neighborhoods with the highest educational attainment had observed risks 1.20% lower than predicted (95% CI=1.61, 0.78). The addition of area-level education improved model calibration across tertiles of area-level education, with mean differences between observed and predicted risks narrowing for each group. Results for the Laboratory Model were similar. Using quintiles of SES variables yielded similar results (data not shown).

Most SES predictors reclassified 2%–3% of individuals in a given risk category (Appendix Tables 3–6). In all models, net risk reclassification worsened for low-risk (0%–10%) individuals with the addition of SES predictors, and improved for middle- and high-risk individuals, typically by reclassifying those who did not develop diabetes to a lower risk category.

Though descriptive analyses suggested synergies between individual- and area-level SES measures (Appendix Table 7), analyses adding the variables to the same models did not improve discrimination or calibration. Adding interactions of SES predictors with each other and with age, sex, and race/ethnicity also failed to improve discrimination or calibration (data not shown). Removing race/ethnicity from models with SES predictors resulted in slightly stronger associations between SES and T2DM, but also decreased model discrimination (data not shown).

Sensitivity analyses evaluating the calibration of two established risk scores across tertiles of select SES predictors (Appendix Table 8) suggested an overestimation of risk for high-SES groups, but only the Framingham score underestimated risk for low-SES groups. The addition of SES predictors to the risk scores improved calibration across tertiles of the SES variables for the Framingham score, but results were mixed for the ARIC score, with the calibration slightly worsening for low-SES tertiles in several instances.

DISCUSSION

In this large, multi-ethnic cohort of U.S. adults, the inclusion of individual- and area-level socioeconomic characteristics did not improve diabetes risk prediction model discrimination, but did help model calibration. Even though several socioeconomic characteristics were predictive of T2DM independent of traditional risk factors, no SES variable altered the

overall ability of the models to discriminate between those who would and would not develop diabetes. Adding SES predictors also failed to reclassify most individuals across selected risk categories. Models without socioeconomic predictors, however, generally underestimated risk for individuals of low SES (or residing in low-SES neighborhoods) and overestimated risk for those of high SES. Adding SES predictors, particularly area-level education, generally minimized differences between predicted and observed risk across SES groups, though results varied in established risk models.

The magnitude of underestimation and overestimation across SES categories varied according to the SES predictor and model, but never exceeded a few percentage points. Although differences between observed and predicted risks in the different SES groups may not be very meaningful on the individual level, there are more sizable implications at a population or health system level, as even a small systematic underestimation of risk could result in a substantial number of individuals who would not receive the preventive measures they need. This may become more important with the increasing use of prediction models and the development of thresholds for potential interventions, which are not currently well defined for T2DM.

Several explanations likely exist for the lack of improvement in model discrimination when adding SES predictors. Much of the association between SES and diabetes may be captured by “traditional” risk factors already included in the models, many of which are intermediates on the SES–T2DM pathway (Appendix Figure 1).⁴¹ For example, the addition of SES predictors generally attenuated the association between race/ethnicity and diabetes, and removing race/ethnicity slightly strengthened the associations for SES predictors, indicating that these variables may influence diabetes risk through partially overlapping pathways. Nonetheless, the improvement in prediction accuracy across SES groups with the addition of socioeconomic variables may be due to unmeasured elements of the SES–T2DM relationship not captured by traditional risk factors.

Although there are no other studies directly investigating the benefit of adding socioeconomic variables to T2DM prediction models in the U.S., the results are consistent with the cardiovascular event prediction literature. In studies using the Framingham Risk Score, researchers documented little improvement in discrimination with the addition of SES predictors, but did find that the inclusion of SES eliminated systematic underestimation of risk in low-SES individuals seen with the traditional model.^{18,42} This implies that the failure to include SES could lead to undertreatment of low-SES individuals with effective therapies (e.g. statins), thereby increasing social disparities in cardiovascular disease.¹⁷ If future T2DM management includes risk prediction models,⁴³ then similar effects might be expected based on traditional prediction models.

This study has several strengths. MESA is a diverse, longitudinal cohort with excellent data on diabetes risk factors and socioeconomic variables. Using SES data at the individual and area level enabled comparisons of multiple SES measures, which may be informative for the design of data capture in electronic medical records. Such data also allowed for testing interactions between individual- and area-level SES for predicting diabetes.

Limitations

The study also has several limitations. First, the MESA cohort includes urban middle-aged and older adults and may not reflect the experiences of rural areas or younger individuals. Second, SES predictors were chosen a priori based on associations with chronic disease development,^{44,45} yet other SES predictors (including property and other assets relevant for older adults) and predictors at different spatial resolution (e.g., census block groups) may yield different results. Similarly, other area-level data (e.g., community resources, such as parks or gyms) may be important. The methods used to create the SES indices were purposefully simple to allow for replication, but may not capture all relevant aspects of SES compared with other methods. Furthermore, the SES index combining both individual- and area-level attributes may over-measure certain variables (e.g., education). Finally, though the MESA data are technically interval censored, this study employed proportional hazards models using the midpoint assumption, which can lead to small misestimation of probabilities.⁴⁶

CONCLUSIONS

Diabetes risk prediction models without SES predictors tended to underestimate risk among low-SES individuals and overestimate risk for high-SES individuals. Adding SES predictors largely mitigated these differences. Even though no SES predictor improved model discrimination, it remains an open question whether such predictors should be included to improve prediction accuracy across SES strata. At present, clinicians and public health practitioners should be aware that prediction models that do not account for social information may misestimate risk differentially depending upon SES. Given the increasing use of risk prediction models for clinical and public health decision making,⁴³ future research and implementation efforts should evaluate risk prediction accuracy across the SES distribution to ensure that prevention programs and therapies are effective in reaching those who need them most.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This research was supported by contracts HHSN268201500003I, N01-HC-95159, N01-HC-95160, N01-HC-95161, N01-HC-95162, N01-HC-95163, N01-HC-95164, N01-HC-95165, N01-HC-95166, N01-HC-95167, N01-HC-95168, N01-HC-95169, and R01 HL071759 from the National Heart, Lung, and Blood Institute and by grants UL1-TR-000040, UL1-TR-001079, UL1-RR-024156 and UL1-RR-025005 from National Center for Research Resources. The authors thank the other investigators, the staff, and the participants of the Multi-Ethnic Study of Atherosclerosis for their valuable contributions. A full list of participating Multi-Ethnic Study of Atherosclerosis investigators and institutions can be found at www.mesa-nhlbi.org. The content is solely the responsibility of the authors and does not necessarily represent the office views of the NIH or National Center for Research Resources.

PJC and AVDR conceived of, designed, wrote, reviewed, and edited the manuscript. PJC performed all of the statistical analyses with assistance from RY. SDA, PJC, RY, SDA, AGB, MH, MRC, RAH, and AVDR all interpreted the data and critically edited and reviewed the manuscript for important intellectual content. AVDR and SDA supervised the study. PJC had full access to all the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

No financial disclosures were reported by the authors of this paper.

References

1. Ali MK, Bullard KM, Saaddine JB, et al. Achievement of goals in U.S. diabetes care, 1999–2010. *N Engl J Med*. 2013; 368(17):1613–1624. <https://doi.org/10.1056/NEJMsa1213829>. [PubMed: 23614587]
2. Boyle JP, Thompson TJ, Gregg EW, et al. Projection of the year 2050 burden of diabetes in the U.S. adult population: dynamic modeling of incidence, mortality, and prediabetes prevalence. *Popul Health Metr*. 2010; 8:29. <https://doi.org/10.1186/1478-7954-8-29>. [PubMed: 20969750]
3. Seaquist ER. Addressing the burden of diabetes. *JAMA*. 2014; 311(22):2267–2268. <https://doi.org/10.1001/jama.2014.6451>. [PubMed: 24915253]
4. Noble D, Mathur R, Dent T, et al. Risk models and scores for type 2 diabetes: systematic review. *BMJ*. 2011; 343:d7163. <https://doi.org/10.1136/bmj.d7163>. [PubMed: 22123912]
5. 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. <https://doi.org/10.1056/NEJMoa012512>. [PubMed: 11832527]
6. National Institute for Health and Care Excellence. [Accessed July 5, 2014] Preventing type 2 diabetes: Risk identification and interventions for individuals at high risk. 2012. www.nice.org.uk/guidance/ph38
7. Buijsse B, Simmons RK, Griffin SJ, et al. Risk assessment tools for identifying individuals at risk of developing type 2 diabetes. *Epidemiol Rev*. 2011; 33(1):46–62. <https://doi.org/10.1093/epirev/mxq019>. [PubMed: 21622851]
8. Raynor LA, Pankow JS, Duncan BB, et al. Novel risk factors and the prediction of type 2 diabetes in the Atherosclerosis Risk in Communities (ARIC) study. *Diabetes Care*. 2013; 36(1):70–76. <https://doi.org/10.2337/dc12-0609>. [PubMed: 22933437]
9. Rivera LA, Lebenbaum M, Rosella LC. The influence of socioeconomic status on future risk for developing Type 2 diabetes in the Canadian population between 2011 and 2022: differential associations by sex. *Int J Equity Health*. 2015; 14:101. <https://doi.org/10.1186/s12939-015-0245-0>. [PubMed: 26496768]
10. Sun F, Tao Q, Zhan S. An accurate risk score for estimation 5-year risk of type 2 diabetes based on a health screening population in Taiwan. *Diabetes Res Clin Pract*. 2009; 85(2):228–234. <https://doi.org/10.1016/j.diabres.2009.05.005>. [PubMed: 19500871]
11. Hippisley-Cox J, Coupland C, Robson J, et al. Predicting risk of type 2 diabetes in England and Wales: prospective derivation and validation of QDScore. *BMJ*. 2009; 338:b880. <https://doi.org/10.1136/bmj.b880>. [PubMed: 19297312]
12. Agardh E, Allebeck P, Hallqvist J, et al. Type 2 diabetes incidence and socio-economic position: a systematic review and meta-analysis. *Int J Epidemiol*. 2011; 40(3):804–818. <https://doi.org/10.1093/ije/dyr029>. [PubMed: 21335614]
13. Auchincloss AH, Diez Roux AV, Mujahid MS, et al. Neighborhood resources for physical activity and healthy foods and incidence of type 2 diabetes mellitus: the Multi-Ethnic study of Atherosclerosis. *Arch Intern Med*. 2009; 169(18):1698–1704. <https://doi.org/10.1001/archinternmed.2009.302>. [PubMed: 19822827]
14. Ludwig J, Sanbonmatsu L, Gennetian L, et al. Neighborhoods, obesity, and diabetes—a randomized social experiment. *N Engl J Med*. 2011; 365(16):1509–1519. <https://doi.org/10.1056/NEJMsa1103216>. [PubMed: 22010917]
15. Kelly SJ, Ismail M. Stress and Type 2 Diabetes: A Review of How Stress Contributes to the Development of Type 2 Diabetes. *Annu Rev Public Health*. 2015; 36(1):441–462. <https://doi.org/10.1146/annurev-publhealth-031914-122921>. [PubMed: 25581145]
16. Nandi A, Glymour MM, Subramanian SV. Association among socioeconomic status, health behaviors, and all-cause mortality in the United States. *Epidemiology*. 2014; 25(2):170–177. <https://doi.org/10.1097/EDE.0000000000000038>. [PubMed: 24487200]
17. Fiscella K, Tancredi D. Socioeconomic status and coronary heart disease risk prediction. *JAMA*. 2008; 300(22):2666–2668. <https://doi.org/10.1001/jama.2008.792>. [PubMed: 19066387]

18. Fiscella K, Tancredi D, Franks P. Adding socioeconomic status to Framingham scoring to reduce disparities in coronary risk assessment. *Am Heart J*. 2009; 157(6):988–994. <https://doi.org/10.1016/j.ahj.2009.03.019>. [PubMed: 19464408]
19. Weber GM, Mandl KD, Kohane IS. Finding the Missing Link for Big Biomedical Data. *JAMA*. 2014; 311(24):2479–2480. <https://doi.org/10.1001/jama.2014.4228>. [PubMed: 24854141]
20. Genuth S, Alberti KG, Bennett P, et al. Follow-up report on the diagnosis of diabetes mellitus. *Diabetes Care*. 2003; 26(11):3160–3167. <https://doi.org/10.2337/diacare.26.11.3160>. [PubMed: 14578255]
21. Diez Roux AV, Mair C. Neighborhoods and health. *Ann N Y Acad Sci*. 2010; 1186:125–145. <https://doi.org/10.1111/j.1749-6632.2009.05333.x>. [PubMed: 20201871]
22. Moore LV, Diez Roux AV. Associations of neighborhood characteristics with the location and type of food stores. *Am J Public Health*. 2006; 96(2):325–331. <https://doi.org/10.2105/AJPH.2004.058040>. [PubMed: 16380567]
23. Duncan DT, Kawachi I, White K, et al. The geography of recreational open space: influence of neighborhood racial composition and neighborhood poverty. *J Urban Health*. 2013; 90(4):618–631. <https://doi.org/10.1007/s11524-012-9770-y>. [PubMed: 23099625]
24. Mason P, Kearns A, Livingston M. “Safe Going”: the influence of crime rates and perceived crime and safety on walking in deprived neighbourhoods. *Soc Sci Med*. 2013; 91:15–24. <https://doi.org/10.1016/j.socscimed.2013.04.011>. [PubMed: 23849234]
25. Hajat A, Diez-Roux AV, Adar SD, et al. Air pollution and individual and neighborhood socioeconomic status: evidence from the Multi-Ethnic Study of Atherosclerosis (MESA). *Environ Health Perspect*. 2013; 121(11–12):1325–1333. <https://doi.org/10.1289/ehp.1206337>. [PubMed: 24076625]
26. Havard S, Deguen S, Zmirou-Navier D, et al. Traffic-Related Air Pollution and Socioeconomic Status: A Spatial Autocorrelation Study to Assess Environmental Equity on a Small-Area Scale. *Epidemiology*. 2009; 20(2):223–230. <https://doi.org/10.1097/EDE.0b013e31819464e1>. [PubMed: 19142163]
27. Githorpe MS. The importance of normalisation in the construction of deprivation indices. *J Epidemiol Community Health*. 1995; 49(Suppl 2):S45–50. https://doi.org/10.1136/jech.49.Suppl_2.S45. [PubMed: 8594134]
28. Abbasi A, Peelen LM, Corpeleijn E, et al. Prediction models for risk of developing type 2 diabetes: systematic literature search and independent external validation study. *BMJ*. 2012; 345:e5900. <https://doi.org/10.1136/bmj.e5900>. [PubMed: 22990994]
29. Diez Roux AV, Merkin SS, Arnett D, et al. Neighborhood of residence and incidence of coronary heart disease. *N Engl J Med*. 2001; 345(2):99–106. <https://doi.org/10.1056/NEJM200107123450205>. [PubMed: 11450679]
30. Krishnan S, Cozier YC, Rosenberg L, et al. Socioeconomic Status and Incidence of Type 2 Diabetes: Results From the Black Women’s Health Study. *Am J Epidemiol*. 2010; 171(5):564–570. <https://doi.org/10.1093/aje/kwp443>. [PubMed: 20133518]
31. Schmidt MI, Duncan BB, Bang H, et al. Identifying individuals at high risk for diabetes: The Atherosclerosis Risk in Communities study. *Diabetes Care*. 2005; 28(8):2013–2018. <https://doi.org/10.2337/diacare.28.8.2013>. [PubMed: 16043747]
32. Mann DM, Bertoni AG, Shimbo D, et al. Comparative Validity of 3 Diabetes Mellitus Risk Prediction Scoring Models in a Multiethnic U.S. Cohort. *Am J Epidemiol*. 2010; 171(9):980–988. <https://doi.org/10.1093/aje/kwq030>. [PubMed: 20375194]
33. Peng, RD., Dominici, F. *Statistical Methods for Environmental Epidemiology with R: A Case Study in Air Pollution and Health*. Springer; 2008.
34. Pencina MJ, D’Agostino RB. Overall C as a measure of discrimination in survival analysis: model specific population value and confidence interval estimation. *Stat Med*. 2004; 23(13):2109–2123. <https://doi.org/10.1002/sim.1802>. [PubMed: 15211606]
35. Harrell FE Jr, Lee KL, Califf RM, et al. Regression modelling strategies for improved prognostic prediction. *Stat Med*. 1984; 3(2):143–152. <https://doi.org/10.1002/sim.4780030207>. [PubMed: 6463451]

36. Heagerty PJ, Lumley T, Pepe MS. Time-dependent ROC curves for censored survival data and a diagnostic marker. *Biometrics*. 2000; 56(2):337–344. <https://doi.org/10.1111/j.0006-341X.2000.00337.x>. [PubMed: 10877287]
37. Ioannidis JP, Tzoulaki I. What makes a good predictor?: the evidence applied to coronary artery calcium score. *JAMA*. 2010; 303(16):1646–1647. <https://doi.org/10.1001/jama.2010.503>. [PubMed: 20424257]
38. Cook NR, Ridker PM. Advances in Measuring the Effect of Individual Predictors of Cardiovascular Risk: The Role of Reclassification Measures. *Ann Intern Med*. 2009; 150(11):795–W143. <https://doi.org/10.7326/0003-4819-150-11-200906020-00007>. [PubMed: 19487714]
39. Royston P, Moons KG, Altman DG, et al. Prognosis and prognostic research: Developing a prognostic model. *BMJ*. 2009; 338:b604. <https://doi.org/10.1136/bmj.b604>. [PubMed: 19336487]
40. Wilson PW, Meigs JB, Sullivan L, et al. Prediction of incident diabetes mellitus in middle-aged adults: the Framingham Offspring Study. *Arch Intern Med*. 2007; 167(10):1068–1074. <https://doi.org/10.1001/archinte.167.10.1068>. [PubMed: 17533210]
41. Adler NE, Newman K. Socioeconomic Disparities In Health: Pathways And Policies. *Health Aff (Millwood)*. 2002; 21(2):60–76. <https://doi.org/10.1377/hlthaff.21.2.60>. [PubMed: 11900187]
42. Brindle PM, McConnachie A, Upton MN, et al. The accuracy of the Framingham risk-score in different socioeconomic groups: a prospective study. *Br J Gen Pract*. 2005; 55(520):838–845. [PubMed: 16281999]
43. Sussman JB, Kent DM, Nelson JP, et al. Improving diabetes prevention with benefit based tailored treatment: risk based reanalysis of Diabetes Prevention Program. *BMJ*. 2015; 350:h454. <https://doi.org/10.1136/bmj.h454>. [PubMed: 25697494]
44. Galobardes B, Lynch J, Smith GD. Measuring socioeconomic position in health research. *Br Med Bull*. 2007; 81–82(1):21–37. <https://doi.org/10.1093/bmb/ldm001>.
45. Krieger N, Chen JT, Waterman PD, et al. Painting a truer picture of U.S. socioeconomic and racial/ethnic health inequalities: the Public Health Disparities Geocoding Project. *Am J Public Health*. 2005; 95(2):312–323. <https://doi.org/10.2105/AJPH.2003.032482>. [PubMed: 15671470]
46. Kleinbaum, DG., Klein, M. *Survival Analysis: A Self-Learning Text*. Springer; New York: 2006.

Figure 1.

Calibration of clinical models with and without socioeconomic information by tertiles of individual household income per capita, individual SES index, area-level education, and area SES index, Multi-Ethnic Study of Atherosclerosis, 2000–2012.

Figure 2. Calibration of laboratory models with and without socioeconomic information by tertiles of individual income category, individual household income per capita, area education, and area SES index, Multi-Ethnic Study of Atherosclerosis, 2000–2012.

Table 1

Baseline Sociodemographic Characteristics and Diabetes Risk Factors, Multi-Ethnic Study of Atherosclerosis, 2000

Characteristics	Overall	Incident cases	Non-cases
Sample, n	5,021	615	4,406
Age, mean (SD), years	61.31 (10.16)	60.61 (9.48)	61.41 (10.25)
Female, n (%)	2,664 (53.06)	326 (53.01)	2,338 (53.06)
Race/Ethnicity, n (%)			
White	2,153 (42.88)	191 (31.06)	1,962 (44.53)
African-American	1,217 (24.24)	180 (29.27)	1,037 (23.54)
Hispanic	1,046 (20.83)	168 (27.32)	878 (19.93)
Chinese-American	605 (12.05)	76 (12.36)	529 (12.01)
Parent or sibling with diabetes, n (%)	1,756 (34.97)	302 (49.11)	1,454 (33.00)
Fasting plasma glucose, mean (SD), mg/dL	89.33 (10.40)	100.67 (12.75)	87.74 (8.96)
HDL cholesterol, mean (SD), mg/dL	51.65 (14.84)	46.60 (11.96)	52.35 (15.07)
Triglycerides, median (IQR), mg/dL	109.00 (80.00)	133.00 (94.00)	106.00 (78.00)
BMI, mean (SD), kg/m ²	27.98 (5.26)	31.09 (5.93)	27.53 (5.01)
25–29.9, n (%)	2,011 (40.05)	222 (36.10)	1,789 (40.60)
30, n (%)	1,473 (29.34)	314 (51.06)	1,159 (26.31)
Waist circumference, mean (SD), cm	97.04 (14.05)	104.99 (14.38)	95.93 (13.64)
Systolic blood pressure, mean (SD), mmHg	124.63 (20.66)	129.74 (20.34)	123.91 (20.61)
Hypertension, n (%) ^a	2,010 (40.03)	324 (52.68)	1,686 (32.86)
Taking anti-hypertensive medication, n (%)	1,616 (32.18)	272 (44.23)	1,344 (30.50)
Individual-level SES variables			
Education, mean (SD), y	13.50 (3.83)	12.99 (3.99)	13.57 (3.81)
High school or less, n (%)	1,602 (31.91)	226 (36.75)	1,376 (31.23)
Some college or associates/technical degree, n (%)	1,438 (28.64)	190 (30.89)	1,248 (28.33)
Bachelor's degree or higher, n (%)	1,981 (39.45)	199 (32.36)	1,782 (40.44)
Household income, mean (SD), per \$10,000	52.16 (34.46)	47.81 (32.24)	52.77 (34.72)
<\$25,000, n (%)	1,380 (27.48)	183 (29.76)	1,197 (27.17)
\$25,000–\$75,000, n (%)	2,356 (46.92)	315 (51.22)	2,041 (46.32)
>\$75,000, n (%)	1,285 (25.59)	117 (19.02)	1,168 (26.51)
Household income per capita, mean (SD), per \$10,000 ^b	2.77 (2.12)	2.42 (1.83)	2.82 (2.15)
SES index, mean (SD) ^c	0.00 (1.71)	–0.30 (1.62)	0.04 (1.71)
Area-level SES variables			
Percent of adults aged ≥25 years with a bachelor's degree, median (IQR)	24.39 (30.88)	21.27 (24.09)	25.01 (32.02)
Median household income, median (IQR), per \$10,000	41.71 (24.35)	40.39 (22.29)	41.78 (24.86)
Percent of persons living below the poverty line, median (IQR)	12.14 (15.27)	13.62 (17.24)	12.14 (14.97)
SES index, mean (SD) ^c	0.79 (5.66)	–0.25 (5.56)	0.94 (5.66)
Combined individual and area-level SES			
Combined SES index, mean (SD) ^c	0.00 (1.72)	–0.36 (1.64)	0.05 (1.73)

^aHypertension defined as blood pressure $\geq 140/90$ mmHg, or treatment with anti-hypertensive medication

^bDefined as annual household income divided by the number of persons supported by that income

^cHigher values indicate higher SES

HDL, high-density lipoprotein; IQR, inter-quartile range

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 2

Association of Individual and Area-level SES Variables With Incident Type 2 Diabetes in Prediction Models Containing Clinical and Laboratory Variables, Multi-Ethnic Study of Atherosclerosis, 2000–2012^a

SES predictors ^b	Multivariable clinical model ^c			Multivariable lab model ^d		
	HR (95% CI) ^e	p-value	C	HR (95% CI) ^e	p-value	C
None			0.72			0.83
Individual-level						
Education (continuous)	0.94 (0.86, 1.03)	0.17	0.72	1.00 (0.91, 1.10)	0.99	0.83
Education (categorical)						
High school or less	1.00 (Ref)	0.36 ^f	0.72	1.00 (Ref)	0.97 ^f	0.83
Some college or associate's/technical degree	0.92 (0.75, 1.13)	0.42		0.98 (0.79, 1.21)	0.82	
Bachelor's degree or higher	0.86 (0.70, 1.06)	0.16		0.97 (0.78, 1.22)	0.82	
Household income (continuous)	0.95 (0.86, 1.04)	0.28	0.72	0.91 (0.83, 1.01)	0.07	0.83
Household income (categorical)						
<\$25,000	1.00 (Ref)	0.18 ^f	0.72	1.00 (Ref)	0.02 ^f	0.83
\$25,000–\$74,999	0.99 (0.81, 1.21)	0.91		1.01 (0.83, 1.23)	0.94	
\$75,000	0.81 (0.63, 1.06)	0.12		0.75 (0.58, 0.97)	0.03	
Household income per capita (continuous)	0.91 (0.82, 1.01)	0.07	0.72	0.91 (0.82, 1.02)	0.09	0.83
SES index	0.91 (0.82, 1.00)	0.05	0.72	0.95 (0.86, 1.05)	0.33	0.83
Area-level ^g						
Percent of adults aged ≥25 years with a bachelor's degree	0.91 (0.83, 1.01)	0.07	0.72	0.92 (0.83, 1.02)	0.13	0.83
Median household income	0.97 (0.89, 1.07)	0.60	0.72	0.92 (0.83, 1.02)	0.12	0.83
Percent of persons living below the poverty line	1.02 (0.94, 1.12)	0.61	0.72	1.07 (0.98, 1.16)	0.14	0.83
SES index	0.93 (0.84, 1.02)	0.12	0.72	0.91 (0.83, 1.01)	0.07	0.83
Combined individual and area-level						
Overall SES index	0.90 (0.81, 1.00)	0.05	0.72	0.92 (0.83, 1.01)	0.09	0.83

^aNumber of incident diabetes cases=615.

^bSES predictors added individually to separate models.

^c Cox proportional hazards models that include the following predictors: age (spline), race, family history of type 2 diabetes, systolic blood pressure, waist circumference, and anti-hypertensive medication use.

^d Cox proportional hazards models that include all predictors from the clinical models, plus fasting plasma glucose, and log triglycerides (spline)

^e For continuous predictors, HRs are estimated per SD change in the predictor. SDs correspond to 3.8 years for individual education, \$34,460 for individual income, \$20,120 for individual income per capita, 1.71 units for individual SES index, 22% for percent of adults aged 25 years with a bachelor's degree, \$20,994 for median household income, 11% for percent of persons living below the poverty line, 5.66 units for area SES index, and 1.72 units for overall SES index.

^f *p*-value corresponds to a likelihood ratio test for all categories combined.

^g Area-level variables measured at the Census-tract level.

C, Harrell's C Statistic; HR, hazard ratio