Bridging the Gap: Biologic, Behavioral, and Environmental Contributions to the Development of Type 2 Diabetes

by

Paul Christine

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Epidemiological Science)
in the University of Michigan
2015

Doctoral Committee:

Professor Ana V. Diez Roux, Drexel University, Co-Chair
Assistant Professor Sara D. Adar, Co-Chair
Associate Professor Brisa N. Sánchez
Associate Professor Mary Ellen Michele Heisler
Assistant Research Scientist Jennifer Ann Smith

Dedication

To my parents, and my parents' parents…without your sacrifices,
none of this would have been possible

Acknowledgements

My most sincere thanks to my advisor, Ana Diez Roux, who gave me the opportunity to study public health. Ana provided me freedom to explore, encouraged me to read widely, and knew when to push. I am grateful for her patience and unwavering support, and deeply respect her work and commitment to justice. Thanks to Sara Adar, who graciously offered to step in as a co-chair on a project outside of her expertise. Sara's passion for rigorous science and her attention to detail have benefitted my work tremendously, and I am grateful for the example she has set. Thanks also to the other members of my committee for their time, patience, and expertise. Brisa Sánchez offered valuable analytical insights into virtually every piece of this work. Jennifer Smith was heroically patient with my many questions about genetic epidemiology, and was instrumental in obtaining and combining social and genetic data. Michele Heisler challenged me to integrate my clinical and epidemiological thinking, and helped pull me out of the trees to find the forest. I am grateful to Kari Moore, whose assiduous data management made literally every part of this dissertation easier to perform. Wei Zaho and Erin Ware provided crucial support in assembling the genetic data. I received expert advice from multiple collaborators, especially Amy Auchincloss, Rebekah Young, and Rod Hayward. And of course, this work would not have been possible without the commitment of thousands of MESA participants, staff, and researchers: their endurance

through countless hours of questions, needle pokes, data cleaning, and meetings enabled the questions in this dissertation to be asked in the first place.

The Center for Social Epidemiology and Population Health has served as a great intellectual home for the past four years, and its faculty, staff, and students have been a ceaseless source of insight, help, and friendship. I am deeply grateful to Amanda Dudley, whose work ethic, patience, and selflessness have never ceased to amaze me. Thanks to Marcia DeBoer and Meredith McGee for making the office a welcoming and pleasant place to be. CSEPH is brimming with smart, funny, justice-driven graduate students that I am lucky to count as friends. Nicole Novak, Amanda Onwuka, Kristen Brown, and Grace Noppert pushed me to think critically about health disparities research. I am grateful to Jeff Wing for his patience and help on topics ranging from programming in SAS, to tiling my bathroom floor. Owais Gilani not only kept me in shape on the squash court, but was instrumental in completing the analyses for one of my chapters. Wei Perng offered helpful advice and friendship throughout the process. I am also thankful to the members of Epidocs 2011 for providing a fun and supportive environment during our graduate work.

I learned a great deal about epidemiology from both teaching it and working on projects outside of my dissertation. Thanks to Hal Morgenstern for providing me the opportunity to work with him as a GSI for two semesters. Rafael Meza, Eduardo Villamor, Tonatiuh Barrientos-Gutierrez, Sarah Cherng, and Jamie Tam provided wonderful opportunities for me to expand my interests and skills beyond my thesis topic, and I am grateful for their collaborations.

While an eight-year training program still sounds a bit crazy to me, the faculty, staff, and students of the MSTP have made the process not only tolerable, but enjoyable. A special thanks to Ron Koenig for giving me not one, but two chances at pursuing a "non-traditional" PhD. Ron not only believed me when I told him that social epidemiology was really just applied history, but has been endlessly supportive both of my endeavors in public health as well as the importance of training MD/PhDs in the social sciences and humanities. As the first MSTP to pursue doctoral work in epidemiology (at least in recent memory), there were many administrative bumps that would have been far worse without the help of Ellen Elkin, Hilkka Ketola, and Laurie Koivupalo. I want to thank Ken Langa for encouraging me to pursue research beyond the formal medical system, and for steadying me whenever I felt lost in the process. Thanks to my co-conspirators in public health, Anup Das, Zishaan Farooqui, and Adam Markovitz, for their companionship and insights as we grapple with how to make our training useful to the world. And a special thanks to Kristin Collier who kept me engaged clinically throughout my graduate training, and provided a tremendous model for being a compassionate physician.

I am extremely grateful for a group of friends that have made living in Ann Arbor a delight. Nicole Novak's heartfelt support and tasty meals kept my spirits and my waistline from shrinking during graduate school. Priyanka Rao and Amit Gupta were always at the ready with a listening ear, delicious food, and a plan for fun outings, and I am grateful for their love and support. The Camp Take Notice community provided a bevy of friendships and love, as well as a sense of place and purpose in the Ann Arbor community. And to my best friends and roommates, Lizz Huntley and Sadie-dog: I

simply could not have done this without you. Thank you for tolerating late walks and dinners, for infusing our home with joy and laughter, and for shepherding me through this process with such love and patience.

Finally, I am grateful to my family. Their patience and understanding through this process warrants a tattoo, not merely a mention in a set of acknowledgements. Thanks to my Nana and Papa for encouraging me to think big and for instilling in me a sense of humor to fall back on during hard times. My sister Angela and brother-in-law Matt have been supportive and loyal at all the right times. They also provided me with two beautiful, feisty, and loving nieces, Mila and Abrianna, whose frequent photos and video messages warmed me on even the coldest Michigan days. And most of all, I want to express my gratitude to my parents, Bev and Dennis, who listened to me talk about every iteration of this work, and who helped me to move forward when my own doubts and critiques have stood in the way. I am eternally grateful for the sacrifices that they have made, and for their love and support at every step of the process.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

ACS	American Community Survey
AHEI	Alternative Healthy Eating Index
ARIC	Atherosclerosis Risk in Communities Study
BMI	Body mass index
bp	Base pair
CI	Confidence interval
CM	Clinical Model
FV	Fruit and vegetable
GIS	Geographic information system
HR	Hazard ratio
IQR	Interquartile range
LM	Laboratory Model
MESA	Multi-Ethnic Study of Atherosclerosis
NETS	National Establishment Time Series
RAF	Risk allele frequency
RERI	Relative excess risk due to interaction
SD	Standard deviation
SES	Socioeconomic status
SIC	Standard Industrial Classification
SNP	Single nucleotide polymorphism
US	United States

ABSTRACT

Type 2 diabetes is an important cause of death and disability worldwide. Causes of the growing epidemic have been primarily attributed to obesity, unhealthy diets, and physical inactivity. Prevention of diabetes, therefore, has focused largely on individual behavioral modification. However, the recognition that health behaviors are structured by social conditions and environmental resources has highlighted the importance of thinking about the multi-level causes of diabetes. With continued increases in diabetes prevalence and incidence, population-based prevention strategies that account for both individual and environmental causes of disease are necessary. In this dissertation, we used a large, multi-ethnic, prospective cohort, to examine the social and environmental contributions to the development of diabetes. Our goal was to understand: (1) if neighborhood environments, including the availability of physical resources to support healthy diets and physical activity and social resources to promote safety and social cohesion, are related to diabetes incidence; (2) how neighborhood environments interact with and shape individual genetic susceptibility to diabetes; and (3) the utility of including individual and area-level social information in public health and clinical decision-making using risk prediction models. In the first study, we found that long-term exposure to neighborhoods with greater availability of healthy food and physical activity resources was associated with a lower incidence of diabetes over 10 years. Neighborhood social environments were largely unrelated to diabetes risk. Our second study found that individual genetic susceptibility to diabetes was modified by the availability of healthy food and physical

activity resources. High genetic risk was most harmful for individuals living in neighborhoods with few healthy food and physical activity resources, but was considerably less harmful for individuals living in neighborhoods with more health-promoting resources. In the third study, we found that including social information in risk prediction models helped correct the systematic misestimation of risk for individuals at high and low levels of social disadvantage. The results all support the notion that social and environmental factors play an important role in the development of diabetes, and that altering neighborhood environments may represent a viable, population-based approach to diabetes prevention.

CHAPTER 1 :

INTRODUCTION

Type 2 diabetes is an etiologically complex disease that affects an estimated 18.8 million adults in the US.[1] Despite robust epidemiologic evidence demonstrating the preventability of type 2 diabetes through changes in individual health behaviors,[2-4] such behavioral changes have been limited thus far on a population level as evidenced by the continued increases in diabetes prevalence and incidence.[5-7] This failure to prevent type 2 diabetes may be partially attributable to the limited attention afforded to the multi-level causes of the disease. A growing body of research linking health behaviors[8] and chronic disease risk factors[9-11] to the social circumstances and environment in which individuals live has suggested that altering environments may foster behavioral changes and promote wellbeing.[12-15] Such an approach has been advocated for in both chronic and infectious disease prevention,[9,16] yet few studies have evaluated the potential utility of such approaches.[17] Furthermore, whether information on social and environmental circumstances can be used to guide public health and clinical decision making for type 2 diabetes prevention is currently unknown.[18] Considering environmental contributions to type 2 diabetes may be especially salient given the profound disparities in disease burden by race/ethnicity and socioeconomic status (SES), which are hypothesized to be driven in part by differences in residential environments.[19,20] Further research investigating the links between environments, individual behaviors, and diabetes risk, and the application of this multi-level information to guide prevention strategies, is thus needed. With this in

mind, this dissertation seeks to answer three questions: (1) are neighborhood environments related to the risk for developing type 2 diabetes?; (2) do neighborhood environments modify the effect of traditional risk factors for type 2 diabetes, including genetic susceptibility?; and (3) can ignoring individual and area-level social information bias clinical and public health decision making, particularly when it is based upon risk prediction models which incorporate only traditional biological risk factors?

Background

Neighborhoods, Health Behaviors, and Type 2 Diabetes

The literature documenting associations between neighborhood features and health outcomes is vast, and has expanded tremendously in the past 20 years.[21] Growing from the recognition that individual-level risk factors are insufficient to explain population patterns of disease, studies of neighborhood influences on health outcomes have become a mainstay of contemporary epidemiology.[22] Of particular recent interest has been the relationship between neighborhood physical and social environments and health behaviors. A recent systematic review of studies published from 1998-2005 concluded that there is generally a positive relationship between number of physical activity resources in a neighborhood, including parks and recreational centers, and the level of physical activity among its residents.[23] Similarly, many,[24] but not all,[25] studies have demonstrated a positive association between the presence of stores selling healthy food options and the quality of residents' diets. Features of the social environment, including levels of collective efficacy and safety, have also been linked to physical activity levels and cardiometabolic outcomes,[26,27] though results have been inconsistent.[28,29] Drawing on these relationships, a growing number of cross sectional

and longitudinal studies have begun to link neighborhood physical and social environments to BMI.[14] For instance, a recent longitudinal analysis using the MESA Neighborhood Study found that lower levels of healthy food availability was associated with increased risk of becoming obese over a 5 year period.[9]

Despite the extensive literature documenting the links between neighborhood environments and both health behaviors and obesity, few studies have proceeded to demonstrate a relationship between neighborhood environments and type 2 diabetes.[30] This is curious, as type 2 diabetes is a disease that can be both prevented and substantially controlled though behaviors like increased physical activity.[3] Previous work using three sites of the MESA Neighborhood Study demonstrated that better access to neighborhood physical activity and healthy food resources at baseline was associated with decreased levels of insulin resistance[11] and diabetes.[12] Other work from the British Women's Heart and Health Study found that area-level deprivation was cross-sectionally associated with higher odds of type 2 diabetes, independent of individual-level socioeconomic position.[31] The strongest evidence to date comes from a randomized study (the Moving to Opportunity [MTO] project) that relocated low-income families to low-poverty neighborhoods. After 10 years of follow-up, researchers found that individuals randomized to low-poverty neighborhoods had a decreased prevalence of obesity and lower levels of hemoglobin A1c.[32]

Though literature linking residential neighborhood environments to the risk of type 2 diabetes is growing, the extent to which the observed associations are causal remains unclear.[33,34] Causal inference from prior studies has been limited due to the cross-sectional nature of many of the associations,[30,31,35] and the lack of specificity of

mechanisms by which neighborhood environments (defined largely by SES) may influence diabetes risk.[32,36] The few longitudinal studies that exist have been unable to evaluate long term neighborhood exposures as they relate to incident diabetes, further limiting causal inference.[12,36] Furthermore, while providing important evidence that neighborhood relocation may lead to a reduced incidence of obesity and type 2 diabetes, the MTO study failed to answer the more policy-relevant question regarding how changes in the neighborhood environment where people continually live influence their risk of developing diabetes. It also gave few indications regarding which neighborhood features may be most important, stating, "The mechanisms underlying these associations remain unclear but warrant further investigation…".[32]Longitudinal studies that seek to identify the specific components of neighborhoods that affect diabetes development are thus warranted.

Gene-Environment Interactions in Type 2 Diabetes

While neighborhood environments are important, type 2 diabetes is likely caused by the interplay of both genetic and environmental factors. The concordance of type 2 diabetes between identical twins is 70-90%, and individuals with two diabetic parents have a 40% increased risk of developing the disease,[37] suggesting substantial genetic contributions to disease development (overall heritability is estimated to be 26%). Through genome-wide association studies (GWAS), over 70 single nucleotide polymorphisms (SNPs) that increase the risk of disease have been identified.[38,39] In total, these SNPs only explain approximately 10% of the overall heritability of type 2 diabetes,[40] though considerable efforts are now being directed towards identifying less-

common genetic variants that may have larger effects on diabetes susceptibility than those discovered so far.

Recognition of the genetic and environmental contributions to type 2 diabetes has spurred great interest in exploring their interactions. While the results from some studies have been called into question due to insufficient sample sizes and publication bias,[41] several studies have provided robust evidence of interaction effects. In early studies of biological candidate genes, the effects of several diabetes-associated variants were shown to be attenuated in individuals with higher physical activity levels and specific dietary patterns.[42-45] More recently, genetic risk scores that pool diabetes-associated genetic variants have become available. Gene-environment interaction studies using these risk scores have shown that an individuals' pooled genetic risk can be modified by dietary patterns.[42] For instance, in the Health Professionals' Follow-up Study, researchers found that a Western dietary pattern led to increased risk of diabetes in individuals with higher, but not lower, genetic risk scores.[46]

Though important in demonstrating the modifiability of genetic risk for diabetes, the prior gene-environment research is limited by restricted notions of what constitutes "environment". Most work to date has focused exclusively on individual-level health behaviors like smoking and diet, ignoring how such "environments" are shaped by larger factors like neighborhood disadvantage and the spatial patterning of health promoting resources and norms.[8,24,47] Despite this recognition, and calls in the literature for broader conceptualizations of "environment",[48,49] empirical examinations of the interaction between genetic risk and these larger environmental features remain rare, and no such studies have focused upon diabetes. To the extent that area-level factors both shape health

behaviors and represent policy-relevant realms for intervention, studies investigating the interaction between genetic risk for type 2 diabetes and broader neighborhood environmental features are needed.[48] Such investigations may be important to accurately describe the health effects of social context (e.g. potential heterogeneity according to genetic risk), and to understand the contingent nature of genetic susceptibility.[49]

Multilevel Frameworks to Guide Clinical and Public Health Decisions

Given the growing recognition of individual and environmental contributions to diabetes, both public health practitioners and clinicians are increasingly interested in employing multilevel frameworks to guide prevention and treatment. While the healthcare system does not typically focus on these social determinants, health care is part of the larger system that seeks to address them.[50] Discussions of population health strategies that address the social roots of disease are now common in the medical literature.[17,51] Whether in publications about "comprehensive primary care"[52] or in the application of community health workers,[53] clinicians are increasingly encouraged to view patients within their social and environmental context. In 2014, an Institute of Medicine report recommend the inclusion of individual and area-level social information into electronic medical records to help promote research and clinical decision making that deliberately focuses on the social determinants of health.[54] Yet, despite the recent enthusiasm for this multilevel framework, there have been few empirical demonstrations of how such information could be used to guide clinical and public health decision making, particularly with respect to type 2 diabetes.

One potential avenue for incorporating multilevel frameworks into public health and clinical practice is through the use of predictive risk scores.[18] While best known for

their role in cardiovascular event prediction, risk scores are now used widely in clinical and public health practice for a variety of conditions.[55] Such scores are used to stratify patients into different risk groups, with the goal of directing preventive or curative interventions to those who will benefit most. This is potentially important in diabetes, as intensive lifestyle modifications have been shown to prevent or delay disease onset in high-risk individuals.[3]

A range of diabetes risk scores currently exist.[56] Most of these scores include clinical and biological risk factors such as waist circumference and fasting plasma glucose, and considerable efforts have been made to improve risk scores with novel biological information including genetic risk.[56-60] However, virtually no risk scores include individual or area-level socioeconomic features that likely contribute to diabetes risk in ways not easily rendered by clinical biomarkers. Prior work examining the effect of including such socioeconomic information into cardiovascular risk scores is telling. For instance, researchers who added individual-level income and education to the Framingham Risk Score (FRS) discovered that the score systematically underestimated risk in low-SES individuals.[61,62] Others have found that an individual's FRS can change considerably when neighborhood SES is taken into account.[63] No research has examined whether similar patterns exist with respect to diabetes risk scores. Given the growth of electronic health records (EHRs) and the prospect of linking "non-medical" environmental data to medical records,[64] empirical assessments of the utility of including multi-level social information in risk assessment are warranted.

Summary of Dissertation Aims

In light of the background above, this dissertation uses longitudinal data from the Multi-Ethnic Study of Atherosclerosis (MESA) to assess the contribution of residential environments, and their interactions with individual risk factors, to the development of type 2 diabetes. In particular, aim 1 (Chapter 2) investigates the relationship between cumulative exposure to neighborhood physical and social environments and incident type 2 diabetes; aim 2 (Chapter 3) explores how genetic susceptibility interacts with neighborhood physical and socioeconomic environments to influence the development of type 2 diabetes; and aim 3 (Chapter 4) assesses the utility of incorporating individual and area-level socioeconomic information into a diabetes risk score.

Specific Aim 1

To examine if long-term exposures to neighborhood physical and social environments, including the availability of healthy food and physical activity resources and levels of social cohesion and safety, are associated with the development of type 2 diabetes.

Hypotheses

1. Individuals with greater cumulative exposure to neighborhoods with increased healthy food availability and physical activity resources will be at reduced risk of developing type 2 diabetes, relative to individuals residing in neighborhoods with fewer resources.
2. Individuals with greater cumulative exposure to neighborhoods with increased levels of social cohesion and safety will be at reduced risk of developing type 2 diabetes, relative to individuals residing in neighborhoods with lower levels of social cohesion and safety.

3. The association between neighborhood physical and social environments and type 2 diabetes incidence will be partially mediated by individual health behaviors, including intentional physical activity and healthy diet.

Specific Aim 2

To investigate if genetic risk for type 2 diabetes, summarized using a genetic risk score, interacts with physical and socioeconomic features of residential neighborhoods to influence the risk for type 2 diabetes.

Hypotheses

1. The effect of genetic risk for type 2 diabetes will be significantly stronger for individuals living in neighborhoods with lower healthy food and physical activity resource availability and lower SES, relative to individuals living in neighborhoods with more healthy promoting resources and higher SES.

Specific Aim 3

To evaluate the utility of including individual and neighborhood-level socioeconomic information in type 2 diabetes risk scores, and to quantify the changes in predictive capacity and accuracy of the score when including such variables.

Hypotheses

1. Inclusion of individual and neighborhood-level SES into a diabetes prediction model based on traditional diabetes risk factors will significantly aid in the discrimination of people who will develop diabetes from those who will remain diabetes-free.
2. Diabetes risk prediction models based upon traditional risk factors will underestimate risk in low-SES individuals or those living in low-SES

environments, and addition of SES information will improve prediction accuracy (i.e. model calibration) across the SES distribution.

3. Adding social information to a diabetes prediction model based upon traditional risk factors will result in risk reclassification such that individuals of low individual or area-level SES will be reclassified into higher risk categories, and individuals of high individual or area-level SES will be reclassified into lower risk categories.

Theoretical Framework

The theoretical framework underlying the dissertation aims is illustrated in the conceptual diagram shown in Figure 1.1. As illustrated in the figure, there are multiple pathways through which neighborhood environments may affect the development of type 2 diabetes. The physical environment is hypothesized to exert an influence on health behaviors including diet quality and physical activity, which in turn influence BMI. Similarly, the social environment, through collective notions of social cohesion and safety, is hypothesized to influence both individuals' psychological states and health behaviors. Psychological distress may directly influence metabolic processes, leading to increases in BMI and inflammation, and/or operate through behavioral mechanisms such as diet and physical activity. BMI is in turn causally related to the development of type 2 diabetes through inflammation (or "metaflammation" – a term used to distinguish inflammation caused by metabolic rather than infectious sources).[65] Genetic susceptibility to type 2 diabetes directly influences the risk for the disease, but its effect may be modified by environmental and/or behavioral factors. For instance, genetic susceptibility may only become manifest when individuals have certain behaviors that are shaped by

their environments (e.g. genetics is known to play a role in appetite, which may modify the effect of the food environment on diet quality[49]).

Of note, in this conceptual framework, both the social and physical environment can be thought of as specific examples and/or consequences of neighborhood poverty and disadvantage. The "sorting" of people into high poverty neighborhoods is itself strongly influenced by factors such as race and SES, as has been widely discussed in the literature on residential segregation.[66] It is thus important to keep in mind then that the physical and social environments are but small samples of a larger pattern of structural inequality that systematically places historically marginalized populations at greater risk for disease, and that broader, more fundamental, constructs like residential segregation and the "sorting" mechanism are worthy of study in their own right.[67,68] Nonetheless, this dissertation seeks to highlight specific links between area-level resources and diabetes outcomes with the ultimate goal of helping complicate and combat the notion that disparities in the burden of type 2 diabetes arise simply due to "lifestyle choices" or inherent biologic differences.

References:

1. Ali MK, Bullard KM, Saaddine JB, Cowie CC, Imperatore G, Gregg EW. Achievement of goals in U.S. diabetes care, 1999-2010. *N. Engl. J. Med.* Apr 25 2013;368(17):1613-1624.

2. Tuomilehto J, Lindstrom J, Eriksson JG, et al. Prevention of type 2 diabetes mellitus by changes in lifestyle among subjects with impaired glucose tolerance. *N. Engl. J. Med.* May 3 2001;344(18):1343-1350.

3. Knowler WC, Barrett-Connor E, Fowler SE, et al. Reduction in the incidence of type 2 diabetes with lifestyle intervention or metformin. *N. Engl. J. Med.* Feb 7 2002;346(6):393-403.

4. Danaei G, Pan A, Hu FB, Hernan MA. Hypothetical midlife interventions in women and risk of type 2 diabetes. *Epidemiology.* Jan 2013;24(1):122-128.

5. Danaei G, Finucane MM, Lu Y, et al. National, regional, and global trends in fasting plasma glucose and diabetes prevalence since 1980: systematic analysis of health examination surveys and epidemiological studies with 370 country-years and 2.7 million participants. *Lancet.* Jul 2 2011;378(9785):31-40.

6. Boyle JP, Thompson TJ, Gregg EW, Barker LE, Williamson DF. Projection of the year 2050 burden of diabetes in the US adult population: dynamic modeling of incidence, mortality, and prediabetes prevalence. *Popul Health Metr.* 2010;8:29.

7. Seaquist ER. ADdressing the burden of diabetes. *JAMA.* 2014;311(22):2267-2268.

8. Moore LV, Diez Roux AV, Nettleton JA, Jacobs DR, Jr. Associations of the local food environment with diet quality--a comparison of assessments based on surveys and geographic information systems: the multi-ethnic study of atherosclerosis. *Am. J. Epidemiol.* Apr 15 2008;167(8):917-924.

9. Auchincloss AH, Mujahid MS, Shen M, Michos ED, Whitt-Glover MC, Diez Roux AV. Neighborhood health-promoting resources and obesity risk (the multi-ethnic study of atherosclerosis). *Obesity (Silver Spring, Md.).* Mar 2013;21(3):621-628.

10. Mujahid MS, Diez Roux AV, Morenoff JD, et al. Neighborhood characteristics and hypertension. *Epidemiology.* Jul 2008;19(4):590-598.

11. Auchincloss AH, Diez Roux AV, Brown DG, Erdmann CA, Bertoni AG. Neighborhood resources for physical activity and healthy foods and their association with insulin resistance. *Epidemiology.* Jan 2008;19(1):146-157.

12. Auchincloss AH, Diez Roux AV, Mujahid MS, Shen M, Bertoni AG, Carnethon MR. Neighborhood resources for physical activity and healthy foods and

incidence of type 2 diabetes mellitus: the Multi-Ethnic study of Atherosclerosis. *Arch. Intern. Med.* Oct 12 2009;169(18):1698-1704.

13. Boone-Heinonen J, Gordon-Larsen P. Obesogenic Environments in Youth Concepts and Methods from a Longitudinal National Sample. *Am. J. Prev. Med.* May 2012;42(5):E37-E46.

14. Sallis JF, Floyd MF, Rodríguez DA, Saelens BE. Role of Built Environments in Physical Activity, Obesity, and Cardiovascular Disease. *Circulation.* February 7, 2012 2012;125(5):729-737.

15. Ludwig J, Duncan GJ, Gennetian LA, et al. Neighborhood Effects on the Long-Term Well-Being of Low-Income Adults. *Science.* September 21, 2012 2012;337(6101):1505-1510.

16. Remais JV, Eisenberg JNS. Balance between clinical and environmental responses to infectious diseases. *Lancet.* Apr 14 2012;379(9824):1457-1459.

17. Kindig D, Mullahy J. Comparative Effectiveness-of What? Evaluating Strategies to Improve Population Health. *Jama-Journal of the American Medical Association.* Aug 25 2010;304(8):901-902.

18. Gottlieb L, Sandel M, Adler NE. Collecting and applying data on social determinants of health in health care settings. *JAMA internal medicine.* Jun 10 2013;173(11):1017-1020.

19. Williams DR, Collins C. Racial residential segregation: a fundamental cause of racial disparities in health. *Public Health Rep.* Sep-Oct 2001;116(5):404-416.

20. Golden SH, Brown A, Cauley JA, et al. Health disparities in endocrine disorders: biological, clinical, and nonclinical factors--an Endocrine Society scientific statement. *J. Clin. Endocrinol. Metab.* Sep 2012;97(9):E1579-1639.

21. Macintyre S, Ellaway A, Cummins S. Place effects on health: how can we conceptualise, operationalise and measure them? *Soc. Sci. Med.* Jul 2002;55(1):125-139.

22. Kawachi I, Berkman LF. *Neighborhoods and Health*: Oxford University Press, USA; 2003.

23. Kaczynski AT, Henderson KA. Parks and recreation settings and active living: a review of associations with physical activity function and intensity. *J Phys Act Health.* Jul 2008;5(4):619-632.

24. Diez Roux AV, Mair C. Neighborhoods and health. *Ann. N. Y. Acad. Sci.* Feb 2010;1186:125-145.

25. Boone-Heinonen J, Gordon-Larsen P, Kiefe CI, Shikany JM, Lewis CE, Popkin BM. Fast food restaurants and food stores: longitudinal associations with diet in young to middle-aged adults: the CARDIA study. *Arch. Intern. Med.* Jul 11 2011;171(13):1162-1170.

26. 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.* Aug 2013;91:15-24.

27. Cohen DA, Finch BK, Bower A, Sastry N. Collective efficacy and obesity: the potential influence of social factors on health. *Soc. Sci. Med.* Feb 2006;62(3):769-778.

28. Mujahid MS, Diez Roux AV, Shen M, et al. Relation between neighborhood environments and obesity in the Multi-Ethnic Study of Atherosclerosis. *Am. J. Epidemiol.* Jun 1 2008;167(11):1349-1357.

29. Foster S, Giles-Corti B. The built environment, neighborhood crime and constrained physical activity: an exploration of inconsistent findings. *Prev. Med.* Sep 2008;47(3):241-251.

30. Leal C, Chaix B. The influence of geographic life environments on cardiometabolic risk factors: a systematic review, a methodological assessment and a research agenda. *Obesity Reviews.* 2011;12(3):217-230.

31. Andersen AF, Carson C, Watt HC, Lawlor DA, Avlund K, Ebrahim S. Life-course socio-economic position, area deprivation and Type 2 diabetes: findings from the British Women's Heart and Health Study. *Diabet. Med.* 2008;25(12):1462-1468.

32. Ludwig J, Sanbonmatsu L, Gennetian L, et al. Neighborhoods, obesity, and diabetes--a randomized social experiment. *N. Engl. J. Med.* Oct 20 2011;365(16):1509-1519.

33. Oakes JM. The (mis)estimation of neighborhood effects: causal inference for a practicable social epidemiology. *Soc. Sci. Med.* May 2004;58(10):1929-1952.

34. Oakes JM. Invited commentary: repeated measures, selection bias, and effect identification in neighborhood effect studies. *Am. J. Epidemiol.* Oct 15 2014;180(8):785-787.

35. Muller G, Kluttig A, Greiser KH, et al. Regional and Neighborhood Disparities in the Odds of Type 2 Diabetes: Results From 5 Population-Based Studies in Germany (DIAB-CORE Consortium). *Am. J. Epidemiol.* Jul 15 2013;178(2):221-230.

36. Schootman M, Andresen EM, Wolinsky FD, et al. The Effect of Adverse Housing and Neighborhood Conditions on the Development of Diabetes Mellitus among

Middle-aged African Americans. *Am. J. Epidemiol.* August 15, 2007 2007;166(4):379-387.

37. Powers AC. *Chapter 344. diabetes mellitus. In: Longo DL, Fauci AS, Kasper DL, Hauser SL, Jameson J, Loscalzo J. eds. Harrison's Priciples of Internal Medicine, 18th ed.*: McGraw-Hill; 2012.

38. Hu FB. Globalization of diabetes: the role of diet, lifestyle, and genes. *Diabetes Care.* Jun 2011;34(6):1249-1257.

39. Genome-wide trans-ancestry meta-analysis provides insight into the genetic architecture of type 2 diabetes susceptibility. *Nat. Genet.* Mar 2014;46(3):234-244.

40. McCarthy MI. Genomics, type 2 diabetes, and obesity. *N. Engl. J. Med.* Dec 9 2010;363(24):2339-2350.

41. Franks PW, Pearson E, Florez JC. Gene-environment and gene-treatment interactions in type 2 diabetes: progress, pitfalls, and prospects. *Diabetes Care.* May 2013;36(5):1413-1421.

42. Cornelis MC, Hu FB. Gene-environment interactions in the development of type 2 diabetes: recent progress and continuing challenges. *Annu. Rev. Nutr.* Aug 21 2012;32:245-259.

43. Cornelis MC, Qi L, Kraft P, Hu FB. TCF7L2, dietary carbohydrate, and risk of type 2 diabetes in US women. *Am. J. Clin. Nutr.* Apr 2009;89(4):1256-1262.

44. Fisher E, Boeing H, Fritsche A, Doering F, Joost HG, Schulze MB. Whole-grain consumption and transcription factor-7-like 2 (TCF7L2) rs7903146: gene-diet interaction in modulating type 2 diabetes risk. *Br. J. Nutr.* Feb 2009;101(4):478-481.

45. Florez JC, Jablonski KA, Bayley N, et al. TCF7L2 polymorphisms and progression to diabetes in the Diabetes Prevention Program. *N. Engl. J. Med.* Jul 20 2006;355(3):241-250.

46. Qi L, Cornelis MC, Zhang C, van Dam RM, Hu FB. Genetic predisposition, Western dietary pattern, and the risk of type 2 diabetes in men. *Am. J. Clin. Nutr.* May 2009;89(5):1453-1458.

47. Franco M, Diez-Roux AV, Nettleton JA, et al. Availability of healthy foods and dietary patterns: the Multi-Ethnic Study of Atherosclerosis. *Am. J. Clin. Nutr.* Mar 2009;89(3):897-904.

48. Boardman JD, Daw J, Freese J. Defining the environment in gene-environment research: lessons from social epidemiology. *Am. J. Public Health.* Oct 2013;103 Suppl 1:S64-72.

49. Glymour MM, Osypuk TL, Rehkopf DH. Invited commentary: Off-roading with social epidemiology--exploration, causation, translation. *Am. J. Epidemiol.* Sep 15 2013;178(6):858-863.

50. Gruen RL, Pearson SD, Brennan TA. Physician-citizens--public roles and professional obligations. *JAMA.* Jan 7 2004;291(1):94-98.

51. Sox HC. Resolving the tension between population health and individual health care. *JAMA.* Nov 13 2013;310(18):1933-1934.

52. Magnussen L, Ehiri J, Jolly P. Comprehensive versus selective primary health care: Lessons for global health policy. *Health Aff. (Millwood).* May-Jun 2004;23(3):167-176.

53. Spencer MS, Rosland AM, Kieffer EC, et al. Effectiveness of a Community Health Worker Intervention Among African American and Latino Adults With Type 2 Diabetes: A Randomized Controlled Trial. *Am. J. Public Health.* Dec 2011;101(12):2253-2260.

54. Institute of Medicine Committee on the Recommended Social and Behavioral Domains and Measures for Electronic Health Records. *Capturing Social and Behavioral Domains and Measures in Electronic Health Records: Phase 2*. Washington (DC): The National Academies Press; 2015.

55. Noble D, Mathur R, Dent T, Meads C, Greenhalgh T. Risk models and scores for type 2 diabetes: systematic review. *BMJ.* 2011;343:d7163.

56. Buijsse B, Simmons RK, Griffin SJ, Schulze MB. Risk assessment tools for identifying individuals at risk of developing type 2 diabetes. *Epidemiol. Rev.* Jul 2011;33(1):46-62.

57. Vassy JL, Durant NH, Kabagambe EK, et al. A genotype risk score predicts type 2 diabetes from young adulthood: the CARDIA study. *Diabetologia.* Oct 2012;55(10):2604-2612.

58. Vassy JL, Hivert MF, Porneala B, et al. Polygenic type 2 diabetes prediction at the limit of common variant detection. *Diabetes.* Feb 11 2014.

59. Meigs JB, Shrader P, Sullivan LM, et al. Genotype score in addition to common risk factors for prediction of type 2 diabetes. *N. Engl. J. Med.* Nov 20 2008;359(21):2208-2219.

60. 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.* Jan 2013;36(1):70-76.

61. Fiscella K, Tancredi D. Socioeconomic status and coronary heart disease risk prediction. *JAMA.* Dec 10 2008;300(22):2666-2668.

62. Fiscella K, Tancredi D, Franks P. Adding socioeconomic status to Framingham scoring to reduce disparities in coronary risk assessment. *Am. Heart J.* Jun 2009;157(6):988-994.

63. Pollack CE, Slaughter ME, Griffin BA, Dubowitz T, Bird CE. Neighborhood socioeconomic status and coronary heart disease risk prediction in a nationally representative sample. *Public Health.* Oct 2012;126(10):827-835.

64. Weber GM, Mandl KD, Kohane IS. Finding the Missing Link for Big Biomedical Data. *JAMA.* May 22 2014.

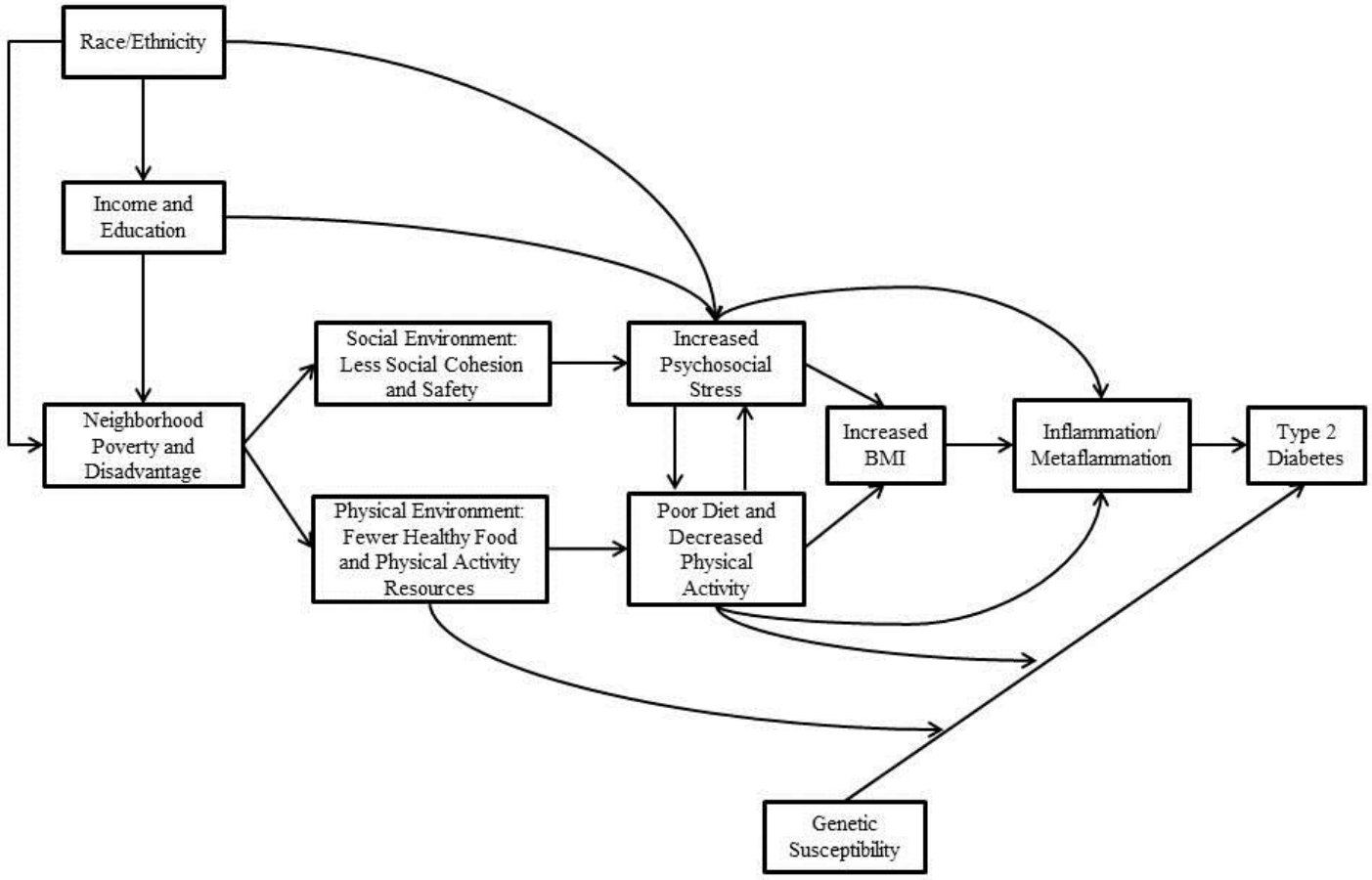
65. Hotamisligil GS. Inflammation and metabolic disorders. *Nature.* 2006;444(7121):860-867.

66. Osypuk TL. Invited commentary: integrating a life-course perspective and social theory to advance research on residential segregation and health. *Am. J. Epidemiol.* Feb 15 2013;177(4):310-315.

67. Sampson RJ, Sharkey P. Neighborhood selection and the social reproduction of concentrated racial inequality. *Demography.* Feb 2008;45(1):1-29.

68. Sampson RJ, Wilson WJ. *Great American City: Chicago and the Enduring Neighborhood Effect*: University of Chicago Press; 2012.

Figure 1.1 Conceptual Diagram of Pathways Linking Individual Attributes and Neighborhood Exposures to the Development of Type 2 Diabetes



CHAPTER 2 :

LONGITUDINAL RELATIONSHIPS BETWEEN NEIGHBORHOOD PHYSICAL AND SOCIAL ENVIRONMENTS AND INCIDENT TYPE 2 DIABETES MELLITUS: THE MULTI-ETHNIC STUDY OF ATHEROSCLEROSIS (MESA)

Introduction:

Diabetes is an important cause of death and disability worldwide.[1] Causes of the growing epidemic have been attributed to obesity, specific dietary patterns (e.g. diets with high glycemic load), physical inactivity, and to a lesser extent, smoking, alcohol use, and stress.[2-6] Prevention of diabetes, therefore, has focused largely on behavioral modification.[3,7-9] However, the extent to which individual behavioral modifications will succeed in unsupportive environments remains unknown.

A growing body of research linking health behaviors[10] and chronic disease risk factors[11-13] to environmental features has suggested that altering environments may foster behavioral changes.[14] Neighborhood physical environments, including access to healthy food and physical activity (PA) resources, may influence individual diet and exercise levels.[15,16] Similarly, local social norms and concerns about neighborhood safety might affect behaviors and stress.[17,18] Modifying environmental resources to support healthy diets, PA, and lower stress levels may therefore aid in diabetes prevention.

Most prior research linking environmental features to diabetes has been cross-sectional, limiting causal conclusions.[14,19-21] The few longitudinal studies that exist have been unable to evaluate long term neighborhood exposures as they relate to incident diabetes, further limiting causal inference.[22,23] One randomized study (Moving to Opportunity [MTO]) that relocated low-income families from high-poverty to low-poverty neighborhoods showed that changing neighborhood environments led to reduced prevalence of obesity and diabetes.[24] However, the MTO study did not answer the equally policy-relevant question regarding how the environment where people continually live, rather than residential relocation, influences their risk of developing diabetes, nor did it indicate which neighborhood features may be most important.[24] Longitudinal studies that seek to identify the specific components of neighborhoods that influence diabetes development are thus warranted.

No study, to our knowledge, has prospectively evaluated whether cumulative exposures to specific neighborhood features are related to incident diabetes in a large, multiethnic, geographically distributed sample. To that end, we investigated whether long-term exposures to neighborhood physical and social environments, including the availability of healthy food and PA resources and levels of social cohesion and safety, are associated with the development of type 2 diabetes over a 10-year period.

Methods:

Study population and analytic sample:

Beginning in 2000, the Multi Ethnic Study of Atherosclerosis (MESA) recruited non-institutionalized adults (45-84 years) who self-identified as white, black, Hispanic, or Chinese from 6 locations (New York, New York; Baltimore, Maryland; Forsyth County,

North Carolina; Chicago, Illinois; St. Paul, Minnesota; and Los Angeles, California).[25] People with clinical cardiovascular disease were excluded. The first examination took place between 2000 and 2002, and 4 follow-up exams occurred an average of 1.6, 3.1, 4.8, and 9.5 years later. Retention rates were 92%, 89%, 87%, and 76%, respectively. Written informed consent was obtained from participants, and the study was approved by institutional review boards at each site.

For this analysis of incident diabetes, we utilized data from an ancillary study, the MESA Neighborhood Study. [26] Of the 6814 individuals enrolled at baseline, 6191 agreed to participate in the Neighborhood Study. We excluded individuals with prevalent diabetes at baseline (n=736) and those with missing exposure, outcome, or covariate data (n=331), leaving 5124 individuals available for analyses.

Type 2 diabetes:

Incident type 2 diabetes was determined at each exam according to the American Diabetes Association 2003 criteria[27]: fasting plasma glucose level ≥126 mg/dL (7 mmol/L), or use of oral hypoglycemic medications or insulin. Glucose levels were obtained from blood samples taken after a 12-hour fast as previously described.[28] The use of oral hypoglycemic medications and insulin was assessed by visual inspection of medications or self-report on the study questionnaire.

Neighborhood physical and social environments:

Assessment of neighborhood healthy food and PA resources was done in two ways using methods consistent with prior studies.[10,26,29-31] First, we constructed Geographic Information System (GIS)-based measures of access to food stores more likely to sell healthier foods (supermarkets and fruit and vegetable markets) and

commercial recreational establishments (facilities for indoor conditioning, dance, bowling, golf, team and racquet sports, and water activities) using annual information from the National Establishment Time Series (NETS) database for years 2000-2012 (See Table 1 and Text 2.A1 for details).[32] For simplicity, these measures will be referred to as "GIS-based supermarkets/FV markets" and "GIS-based commercial recreational establishments". Simple densities per square mile were created for 1-mile buffers around each participant's residence using ArcGIS, version 9.3 (Esri, Redlands, California). Densities were matched to participants annually such that changes over time occurred whenever neighborhood resources changed or a participant moved. One-mile densities were chosen as proxies for neighborhoods based on an area in which most individuals could reasonably walk and federal government definitions of access to services.[33]

As a complementary measure, we also used survey-based measures of neighborhood environments collected in 2003-2005 and 2010-2012 from both MESA participants, and from an independent, but co-located, sample of non-MESA participants recruited from the same census tracts via random-digit dialing or list-based sampling.[26] Respondents were asked to rate the area within 1 mile or a 20-minute walk of their home with respect to availability of healthy foods and walking environment. Social environment was also assessed using scales for safety and social cohesion (see Table 1 and Text 2.A1). Survey responses within 1-mile of each participant's residential address, excluding their own responses, were averaged to create neighborhood measures and assigned based on the closest survey time. A median of 78 responses were available within a 1-mile buffer (see Table 2.A1). All survey scales had good internal consistency

(Cronbach's alphas: 0.64-0.83) and ecometric properties (neighborhood reliabilities: 0.38-0.53).[26]

Because different measures (e.g. GIS- and survey-based for healthy food and PA environments, safety and social cohesion scales for social environment) may reflect different aspects of the same environmental construct, we also calculated summary measures by summing the standardized component measures for healthy food, PA, and social environments (see Text 2.A1). The summary measures had good internal consistency for PA and social environments (α=0.68 and 0.78, respectively) but internal consistency for the healthy food environment was lower (α=0.39).Pearson correlations between the GIS- and survey-based measures were r=0.30 for food environment, and r=0.57 for PA environment.

Covariates:

Covariates measured at baseline included age, sex, race/ethnicity, education, family history of diabetes, and the presence of chronic stress (>6 months of serious financial, health, job, or relationship problems). Time-varying information included household income per capita, alcohol use (no, moderate, or heavy use according to established guidelines),[34] and smoking status (current, former, or never). Potential mediators of the neighborhood resource-diabetes association, including body mass index (BMI, measured weight in kg/(height in m^2)), diet quality, and PA, were assessed via clinical exams (BMI) and questionnaires (see Text 2.A2). At the neighborhood-level, a time-varying socioeconomic index (neighborhood SES) was developed using principal components analysis of census tract data from the Census and American Community

Surveys and linked to each participant's address at their closest exam date (see Text 2.A3).

Statistical Analysis:

We performed descriptive analyses of individual-level variables by diabetes status and tertiles of the summary neighborhood exposures. Crude incidence rates across tertiles of each neighborhood exposure were calculated using Poisson regression. Cox proportional hazards models were used to estimate the hazard ratio (HR) of diabetes for each neighborhood exposure separately. Individuals were considered at risk until diagnosis of diabetes, last follow-up visit, or administrative censoring at exam 5, whichever occurred first. Incident diabetes cases were assigned to the midpoint between their previous diabetes-free and current exam dates. Because long-term neighborhood exposures are most relevant for slowly developing diseases like type 2 diabetes, we parameterized our exposures as time-varying cumulative averages, defined as the average across all months between the baseline and each follow-up exam. Though our outcome is interval censored, we elected to use Cox models because of our interest in time-varying exposures, which are not easily included in interval censored models.[35] Clustering within census tracts was accounted for by computing robust standard errors.

Potential confounders were defined *a priori*, and entered into models in stages. Our primary models adjusted for age, sex, family history of diabetes, per capita household income, education, race/ethnicity, smoking status, and alcohol consumption. Additional models were adjusted for neighborhood SES, though it is debatable whether it is a cause or consequence of some neighborhood exposures (e.g. safety).[36,37] To examine

whether BMI, diet, and/or PA mediate the association between neighborhood resources and diabetes, we compared HRs before and after adjustment for these measures.[38,39]

We evaluated the proportional hazard assumption by plotting Schoenfeld residuals against time, and no violations were found. There was limited evidence of nonlinearity for neighborhood exposures in adjusted Cox models, permitting their inclusion as continuous variables. To facilitate comparisons across exposures with different scales, we estimated HRs for an interquartile range (IQR) increase in the neighborhood exposure. This corresponded to increases of 2.2 supermarkets/FV markets and 3.2 commercial recreational establishments for GIS-based exposures, and between 0.3 and 0.7-unit increases for survey-based exposures.

Based upon prior literature, we evaluated effect modification of the summary measures by age at baseline, sex, and household income per capita using interaction terms.[13,14,23] Because residential environments are hypothesized to be especially salient for individuals with highly stressful lives,[40] we also evaluated effect modification by the presence of chronic stress.

We performed several sensitivity analyses. First, we ran interval censored parametric survival models with a Weibull distribution to assess sensitivity to our modeling approach. We also explored alternative exposure specifications using different geographic (3-mile buffer for GIS measures; census tracts for survey measures) and time (1-year lagged exposures for GIS measures; survey measures unavailable annually) scales. Because population density and regional norms may affect health behaviors independent of neighborhood resources,[29,42] we ran additional models controlling for population density and study site. To help control for unmeasured confounding at the

neighborhood-level, we ran shared frailty models with random intercepts for each census tract (see Table 2.A5).[43,44] Finally, though long-term neighborhood exposures are likely most relevant for diabetes risk, we examined baseline and change since baseline exposure measures to evaluate how these parameterizations were related to diabetes risk (see Text 2.A4 for details).

Results:

Over a median of 8.9 years (37,394 person-years), 616 participants developed type 2 diabetes (12.0%; crude incidence rate = 16.47/1000 person-years; 95% CI, 15.22, 17.83). Compared to participants who did not develop diabetes, incident cases were more likely to be black or Hispanic, had lower baseline household income, fewer years of education, less healthy diets, lower levels of moderate and vigorous PA, a higher BMI, and a family history of type 2 diabetes (Table 2). Participants developing diabetes also lived in poorer census tracts.

Neighborhood physical and social resources were highly patterned by race, diet, PA levels, BMI, and neighborhood SES, such that racial/ethnic minorities, and those with greater risk factor profiles were generally more likely to reside in neighborhoods with fewer resources (Table 3). Temporal changes in neighborhood exposures varied by exposure type, ranging from mean 10-year changes of 2.01 for GIS-based commercial recreational establishments to -0.20 for GIS-based supermarkets/FV markets (Table 2.A2). At baseline, the median duration of neighborhood residence was 15 years, and 32% of individuals moved during follow-up.

Higher baseline summary measures of neighborhood PA, social, and to a lesser extent, healthy food resources were associated with lower crude diabetes incidence rates

(Table 4). For instance, participants residing in neighborhoods in the bottom tertile of summary PA environment developed diabetes at nearly double the rate as those living in the top tertile (incidence rates = 20.5 and 11.8 per 1000 person-years, respectively). GIS-based supermarkets/FV markets and social cohesion were not related to diabetes incidence rates.

After adjustment for baseline age, sex, income, education, race/ethnicity, and alcohol and smoking status, an IQR increase in cumulative exposure to survey-based healthy food resources was associated with a 16% lower diabetes risk (HR, 0.84; 95% CI, 0.76, 0.93), but no association was found using the GIS-based measure (HR, 0.99; 95% CI, 0.94, 1.04) (Figure 1, Model 1). An IQR increase in the summary healthy food environment measure was associated with a 12% lower risk for developing diabetes (HR, 0.88; 95% CI, 0.79, 0.98). Further adjustment for neighborhood SES attenuated the associations (Figure1, Model 2). For PA environments, greater cumulative exposure to neighborhoods with resources supporting PA was inversely associated with diabetes incidence; IQR increases in GIS-based, survey-based, and summary environmental measures were associated with 4% (HR, 0.96; 95% CI, 0.92, 0.99), 21% (HR, 0.79; 95% CI, 0.71, 0.88) and 21% (HR, 0.79; 95% CI, 0.69, 0.90) lower risk for diabetes, respectively. Adjusting for neighborhood SES attenuated the GIS-based association, but left the other associations virtually unchanged. Social cohesion, safety, and the summary measure for social environment were largely unassociated with risk for diabetes (HRs per IQR increase, 0.99; 95% CI, 0.88, 1.10; 0.92; 95% CI, 0.80, 1.05; 0.96; 95% CI, 0.86, 1.07, respectively). Further adjustment of models for BMI, diet, and PA as potential

mediators demonstrated minimal attenuation of most associations (≤25%, see Table 2.A3).

Neighborhood healthy food resources had a stronger inverse association with diabetes among participants who were younger, higher income, and reporting chronic stress burden (P-values for multiplicative and additive interaction≤ 0.06; Figure 2.A1). Similarly, the inverse association between neighborhood PA resources and diabetes was stronger in higher income participants (P-value for multiplicative and additive interaction, 0.07 and 0.04, respectively). Neighborhood social environment was inversely associated with diabetes in women but not men, and in low-income but not high-income participants (P-values for multiplicative and additive interaction all ≤ 0.07).

Sensitivity analyses demonstrated qualitatively similar findings when using interval censored survival methods, different exposure specifications, controls for population density and study site, shared frailty models, and adjustment for baseline risk factors for diabetes (Tables 2.A4-6). Alternative modeling strategies showed that baseline and change in neighborhood exposure levels were associated with incident diabetes in the expected (inverse) direction for survey-based measures, though results were imprecise (Table 2.A7). Baseline levels, but not change, were associated with diabetes for GIS-based commercial recreational establishments.

Discussion:

In this large, multiethnic cohort, long-term exposure to residential environments with greater resources to support PA, and to a lesser extent healthy diets, was associated with lower incidence of type 2 diabetes over 10 years. The associations were generally robust to adjustment for other risk factors and model specifications, though associations

were primarily found with survey-based, but not GIS-based, exposures. Inclusion of BMI, diet, and PA as hypothesized mediators only modestly attenuated the relationships. Neighborhood safety and social cohesion were largely unassociated with the development of diabetes.

Unlike previous studies of residential environments and diabetes,[19,24] we measured specific, time-varying features of participants' neighborhoods using complementary measures. Both geographic proximity to commercial recreational establishments and greater survey-based assessments of the walking environment were inversely associated with diabetes incidence. Previous work using the MESA cohort has demonstrated that an increase in commercial PA resources is associated with less age-related decline in PA.[45] Other studies have found that residential relocation to neighborhoods more supportive of PA is associated with increased levels of PA, independent of reasons for relocation.[46,47] Our study suggests that such neighborhood associations with PA behavior may translate to reduced diabetes risk.

We found that geographic proximity to supermarkets and stores selling fruits and vegetables had no association with diabetes incidence. This finding is consistent with recent observational and quasi-experimental evidence demonstrating that simply improving retail food infrastructure may not translate into healthier diets and decreased risk for chronic diseases.[48-50] On the other hand, survey-based measures of the local food environment were associated with diabetes, suggesting that such measures may take into account other factors like the affordability and quality of food that are known to influence diet and diabetes risk.[51-53]

Finally, though social features of residential environments have been hypothesized to be related to obesity and diabetes through their association with health behaviors and stress,[17,18] we find limited support for these relationships. Additional research with alternative exposure measures is needed to further clarify the role of the social environment.

While the use of multiple modalities for measuring neighborhood environments is a strength in our study, the difference in the associations for GIS-based and survey-based measures of the food and PA environments are noteworthy. The most likely explanation for the discrepancies is that the GIS counts and survey responses measure different aspects of the same construct.[10] For instance, our survey-based PA exposure assesses non-commercial neighborhood features related to walkability and aesthetics not captured in the GIS-based measures. Neighborhood residents also likely consider unmeasured attributes such as cost or quality that are not captured with simple counts from tax parcel data.[54] Differences between the GIS-based and survey-based associations could also be due reverse causation if individuals with less interest in healthy food or PA resources are less likely to perceive that such resources are available. We think this is unlikely for two reasons: the neighborhood survey assesses community ratings of the local environment (with a median of 78 residents in a 1-mile area whose survey responses were averaged), and we excluded an individual's survey response from their own exposure measure. Nonetheless, future research would benefit from including multiple measures of the same neighborhood environmental constructs to further understand the most relevant features for diabetes risk.

We observed differences in the associations between neighborhood features and diabetes according to individual characteristics, though given the multiple comparisons assessed, caution should be exercised in interpreting the results. Household income appeared to be a consistent effect modifier, such that increased healthy food and PA resources were more beneficial to high-income households than low-income households. For low-income households, growing evidence suggests factors like cost may trump geographic proximity to healthy food and PA resources.[55,56] Interestingly, the social environment demonstrated the opposite pattern: increasing safety and social cohesion was associated with lower diabetes risk in low-income but not high-income households. Community safety and social relationships have been associated with BMI and PA in several studies,[57-60] but further work is needed to understand if and why such associations may differ by income. The presence of chronic stressors also modified the association for healthy food environments such that increasing healthy food resources was associated with lower diabetes risk for those with chronic stressors. We are unaware of other studies evaluating this question, though our findings are consistent with literature suggesting that environmental resources may be especially salient for individuals experiencing chronic stress.[40]

Models adjusting for BMI as a mediator modestly attenuated the associations between residential healthy food and PA environments and diabetes incidence. Such modest attenuation is not surprising given the long-term nature of diabetes development,[61] and the difficulty in separating direct and indirect effects in standard regression analyses.[62,63] Diet and PA are also notoriously difficult to measure precisely, and measurement error can distort the magnitude of mediation observed.[64] Further work

focusing specifically on mediation is warranted to quantify the behavioral and biological pathways through which features of the neighborhood environment may influence diabetes risk.

The primary strength of our study is the longitudinal measurement of specific features of neighborhood environments and diabetes status over time in a multiethnic sample. Given that type 2 diabetes develops over a protracted period, such long-term exposure measures are more relevant than simple cross-sectional exposures. Furthermore, utilizing multiple measures for specific environmental features has several advantages. First, such measures can be used to evaluate which features may be most critical for mitigating diabetes risk, rather than focusing solely on neighborhood socioeconomic status, which may be a proxy for many interrelated neighborhood features.[65] Second, specific measures of neighborhood environments may be less susceptible to problems of endogeneity or reverse causation, wherein the characteristics of a neighborhood environment are simply the result of the individual attributes and preferences of residents.[65] Finally, prospective collection of covariate information allowed for updating of confounding variables.

As with all observational studies of neighborhood exposures, residential self-selection, wherein individuals with certain risk profiles select to live in certain neighborhoods, may bias the associations reported.[66] While we attempted to minimize such bias by including individual-level variables related to neighborhood selection,[67] there may be unobserved or mismeasured characteristics that influence both neighborhood exposure and the risk for diabetes. Further use of experimental, quasi-experimental, and observational data utilizing different methodologies may help to

increase our confidence in the associations observed. Other exposures, such as neighborhood traffic safety and availability of green spaces, or those encountered near work or during a commute (e.g. food stores), may also be relevant to diabetes risk.[14,68,69] Finally, 24% of eligible MESA participants were lost to follow-up by exam 5, raising the possibility of bias due to "informative censoring". Dropout was not highly patterned by neighborhood exposures however, making this bias less likely.

The prevalence of type 2 diabetes continues to increase in the US despite its preventability through behavioral modifications.[7,9] While individualized prevention and treatment approaches are necessary to decrease the burden of diabetes, environmental modifications that promote healthy behaviors represent a complementary, perhaps prerequisite, population health approach. Our results suggest that modifying specific features of neighborhood environments, including increasing the availability of healthy foods and PA resources, may help mitigate the risk of diabetes, though additional intervention studies with measures of multiple neighborhood features are needed. Such approaches may be especially important for addressing disparities in type 2 diabetes, given the concentration of low-income and minority populations in neighborhoods with fewer health-promoting resources.[70-72]

References:

1. Danaei G, Finucane MM, Lu Y, et al. National, regional, and global trends in fasting plasma glucose and diabetes prevalence since 1980: systematic analysis of health examination surveys and epidemiological studies with 370 country-years and 2.7 million participants. *Lancet.* Jul 2 2011;378(9785):31-40.

2. Hu FB, Manson JE, Stampfer MJ, et al. Diet, lifestyle, and the risk of type 2 diabetes mellitus in women. *N. Engl. J. Med.* Sep 13 2001;345(11):790-797.

3. Mozaffarian D, Kamineni A, Carnethon M, Djousse L, Mukamal KJ, Siscovick D. Lifestyle risk factors and new-onset diabetes mellitus in older adults: the cardiovascular health study. *Arch. Intern. Med.* Apr 27 2009;169(8):798-807.

4. Hu FB. Globalization of diabetes: the role of diet, lifestyle, and genes. *Diabetes Care.* Jun 2011;34(6):1249-1257.

5. Mokdad AH, Bowman BA, Ford ES, Vinicor F, Marks JS, Koplan JP. THe continuing epidemics of obesity and diabetes in the united states. *JAMA.* 2001;286(10):1195-1200.

6. Schulze MB, Hu FB. Primary prevention of diabetes: what can be done and how much can be prevented? *Annu. Rev. Public Health.* 2005;26:445-467.

7. Knowler WC, Barrett-Connor E, Fowler SE, et al. Reduction in the incidence of type 2 diabetes with lifestyle intervention or metformin. *N. Engl. J. Med.* Feb 7 2002;346(6):393-403.

8. Tuomilehto J, Lindstrom J, Eriksson JG, et al. Prevention of type 2 diabetes mellitus by changes in lifestyle among subjects with impaired glucose tolerance. *N. Engl. J. Med.* May 3 2001;344(18):1343-1350.

9. Danaei G, Pan A, Hu FB, Hernan MA. Hypothetical midlife interventions in women and risk of type 2 diabetes. *Epidemiology.* Jan 2013;24(1):122-128.

10. Moore LV, Diez Roux AV, Nettleton JA, Jacobs DR, Jr. Associations of the local food environment with diet quality--a comparison of assessments based on surveys and geographic information systems: the multi-ethnic study of atherosclerosis. *Am. J. Epidemiol.* Apr 15 2008;167(8):917-924.

11. Auchincloss AH, Mujahid MS, Shen M, Michos ED, Whitt-Glover MC, Diez Roux AV. Neighborhood health-promoting resources and obesity risk (the multi-ethnic study of atherosclerosis). *Obesity (Silver Spring, Md.).* Mar 2013;21(3):621-628.

12. Mujahid MS, Diez Roux AV, Morenoff JD, et al. Neighborhood characteristics and hypertension. *Epidemiology.* Jul 2008;19(4):590-598.

13. Auchincloss AH, Diez Roux AV, Brown DG, Erdmann CA, Bertoni AG. Neighborhood resources for physical activity and healthy foods and their association with insulin resistance. *Epidemiology.* Jan 2008;19(1):146-157.

14. Sallis JF, Floyd MF, Rodríguez DA, Saelens BE. Role of Built Environments in Physical Activity, Obesity, and Cardiovascular Disease. *Circulation.* February 7, 2012 2012;125(5):729-737.

15. Papas MA, Alberg AJ, Ewing R, Helzlsouer KJ, Gary TL, Klassen AC. The built environment and obesity. *Epidemiol. Rev.* 2007;29:129-143.

16. Pearson TA, Palaniappan LP, Artinian NT, et al. American Heart Association Guide for Improving Cardiovascular Health at the Community Level, 2013 Update: A Scientific Statement for Public Health Practitioners, Healthcare Providers, and Health Policy Makers. *Circulation.* April 23, 2013 2013;127(16):1730-1753.

17. Fowler-Brown AG, Bennett GG, Goodman MS, Wee CC, Corbie-Smith GM, James SA. Psychosocial stress and 13-year BMI change among blacks: the Pitt County Study. *Obesity (Silver Spring, Md.).* Nov 2009;17(11):2106-2109.

18. Morenoff JD, Diez Roux AV, Hansen BB, Osypuk TL. Residential environments and obesity: what can we learn about policy interventions from observational studies? In: Schoeni RF, House JS, Kaplan GA, Pollack H, eds. *Making Americans Healthier: Social and Economic Policy as Health Policy*: Russell Sage Foundation; 2008:309-343.

19. Muller G, Kluttig A, Greiser KH, et al. Regional and Neighborhood Disparities in the Odds of Type 2 Diabetes: Results From 5 Population-Based Studies in Germany (DIAB-CORE Consortium). *Am. J. Epidemiol.* Jul 15 2013;178(2):221-230.

20. Andersen AF, Carson C, Watt HC, Lawlor DA, Avlund K, Ebrahim S. Life-course socio-economic position, area deprivation and Type 2 diabetes: findings from the British Women's Heart and Health Study. *Diabet. Med.* 2008;25(12):1462-1468.

21. Leal C, Chaix B. The influence of geographic life environments on cardiometabolic risk factors: a systematic review, a methodological assessment and a research agenda. *Obes Rev.* Mar 2011;12(3):217-230.

22. Schootman M, Andresen EM, Wolinsky FD, et al. The Effect of Adverse Housing and Neighborhood Conditions on the Development of Diabetes Mellitus among Middle-aged African Americans. *Am. J. Epidemiol.* August 15, 2007 2007;166(4):379-387.

23. Auchincloss AH, Diez Roux AV, Mujahid MS, Shen M, Bertoni AG, Carnethon MR. Neighborhood resources for physical activity and healthy foods and

incidence of type 2 diabetes mellitus: the Multi-Ethnic study of Atherosclerosis. *Arch. Intern. Med.* Oct 12 2009;169(18):1698-1704.

24. Ludwig J, Sanbonmatsu L, Gennetian L, et al. Neighborhoods, obesity, and diabetes--a randomized social experiment. *N. Engl. J. Med.* Oct 20 2011;365(16):1509-1519.

25. Bild DE, Bluemke DA, Burke GL, et al. Multi-ethnic study of atherosclerosis: objectives and design. *Am. J. Epidemiol.* Nov 1 2002;156(9):871-881.

26. Mujahid MS, Diez Roux AV, Morenoff JD, Raghunathan T. Assessing the measurement properties of neighborhood scales: from psychometrics to ecometrics. *Am. J. Epidemiol.* Apr 15 2007;165(8):858-867.

27. Genuth S, Alberti KG, Bennett P, et al. Follow-up report on the diagnosis of diabetes mellitus. *Diabetes Care.* Nov 2003;26(11):3160-3167.

28. Bertoni AG, Burke GL, Owusu JA, et al. Inflammation and the incidence of type 2 diabetes: the Multi-Ethnic Study of Atherosclerosis (MESA). *Diabetes Care.* Apr 2010;33(4):804-810.

29. Boone-Heinonen J, Diez-Roux AV, Goff DC, et al. The neighborhood energy balance equation: does neighborhood food retail environment + physical activity environment = obesity? The CARDIA study. *PLoS One.* 2013;8(12):e85141.

30. Auchincloss AH, Moore KA, Moore LV, Diez Roux AV. Improving retrospective characterization of the food environment for a large region in the United States during a historic time period. *Health Place.* Nov 2012;18(6):1341-1347.

31. Moore LV, Diez Roux AV. Associations of neighborhood characteristics with the location and type of food stores. *Am. J. Public Health.* Feb 2006;96(2):325-331.

32. Walls & Associates. National Establishment Time-Series (NETS) Database. 2010; Database Description: www.youreconomy.org/nets/NETSDatabaseDescription.pdf.

33. Ver PLoeg M, Bereneman V, Farrigan T, et al. *Access to affordable and nutritions food - measuring and understanding food deserts and their consequences: Report to congress.* Washington, DC: United States Department of Agriculture;2009.

34. Allen JP, Wilson V, eds. *Assessing alcohol probelms: A guide for clinicians and researchers.* 2nd ed. Bethesda, MD: National Institute on Alcohol Abuse and Alcoholism; 2003.

35. Allison PD. *Survival Analysis Using SAS: A Practical Guide*: SAS Institute; 2010.

36. Greenbaum RT, Tita GE. The Impact of Violence Surges on Neighbourhood Business Activity. *Urban Studies.* December 1, 2004 2004;41(13):2495-2514.

37. Chaix B, Leal C, Evans D. Neighborhood-level confounding in epidemiologic studies: unavoidable challenges, uncertain solutions. *Epidemiology.* Jan 2010;21(1):124-127.

38. Hafeman DM, Schwartz S. Opening the Black Box: a motivation for the assessment of mediation. *Int. J. Epidemiol.* Jun 2009;38(3):838-845.

39. Naimi AI, Kaufman JS, Howe CJ, Robinson WR. Mediation considerations: serum potassium and the racial disparity in diabetes risk. *Am. J. Clin. Nutr.* Aug 2011;94(2):614-616.

40. Mezuk B, Abdou CM, Hudson D, et al. "White Box" Epidemiology and the Social Neuroscience of Health Behaviors: The Environmental Affordances Model. *Soc Ment Health.* Jul 1 2013;3(2).

41. Li R, Chambless L. Test for additive interaction in proportional hazards models. *Ann. Epidemiol.* Mar 2007;17(3):227-236.

42. Boone-Heinonen J, Evenson KR, Song Y, Gordon-Larsen P. Research Built and socioeconomic environments: patterning and associations with physical activity in US adolescents. 2010.

43. Hougaard P. Shared frailty models. *Analysis of Multivariate Survival Data*: Springer New York; 2000:215-262.

44. Wienke A. *Frailty Models in Survival Analysis*: Taylor & Francis; 2010.

45. Ranchod YK, Diez Roux AV, Evenson KR, Sanchez BN, Moore K. Longitudinal associations between neighborhood recreational facilities and change in recreational physical activity in the multi-ethnic study of atherosclerosis, 2000-2007. *Am. J. Epidemiol.* Feb 1 2014;179(3):335-343.

46. Giles-Corti B, Bull F, Knuiman M, et al. The influence of urban design on neighbourhood walking following residential relocation: longitudinal results from the RESIDE study. *Soc. Sci. Med.* Jan 2013;77:20-30.

47. Calise TV, Heeren T, DeJong W, Dumith SC, Kohl HW, 3rd. Do neighborhoods make people active, or do people make active neighborhoods? Evidence from a planned community in Austin, Texas. *Prev Chronic Dis.* 2013;10:E102.

48. Boone-Heinonen J, Gordon-Larsen P, Kiefe CI, Shikany JM, Lewis CE, Popkin BM. Fast food restaurants and food stores: longitudinal associations with diet in young to middle-aged adults: the CARDIA study. *Arch. Intern. Med.* Jul 11 2011;171(13):1162-1170.

49. Cummins S, Flint E, Matthews SA. New neighborhood grocery store increased awareness of food access but did not alter dietary habits or obesity. *Health Aff. (Millwood).* Feb 2014;33(2):283-291.

50. Lucan SC, Hillier A, Schechter CB, Glanz K. Objective and self-reported factors associated with food-environment perceptions and fruit-and-vegetable consumption: a multilevel analysis. *Prev Chronic Dis.* 2014;11:E47.

51. Caspi CE, Kawachi I, Subramanian SV, Adamkiewicz G, Sorensen G. The relationship between diet and perceived and objective access to supermarkets among low-income housing residents. *Soc. Sci. Med.* Oct 2012;75(7):1254-1262.

52. Anekwe TD, Rahkovsky I. The association between food prices and the blood glucose level of US adults with type 2 diabetes. *Am. J. Public Health.* Apr 2014;104(4):678-685.

53. Beydoun MA, Powell LM, Wang Y. The association of fast food, fruit and vegetable prices with dietary intakes among US adults: is there modification by family income? *Soc. Sci. Med.* Jun 2008;66(11):2218-2229.

54. Brownson RC, Hoehner CM, Day K, Forsyth A, Sallis JF. Measuring the built environment for physical activity: state of the science. *Am. J. Prev. Med.* Apr 2009;36(4 Suppl):S99-123 e112.

55. Drewnowski A, Aggarwal A, Hurvitz PM, Monsivais P, Moudon AV. Obesity and supermarket access: proximity or price? *Am. J. Public Health.* Aug 2012;102(8):e74-80.

56. Breyer B, Voss-Andreae A. Food mirages: geographic and economic barriers to healthful food access in Portland, Oregon. *Health Place.* Nov 2013;24:131-139.

57. Burdette HL, Wadden TA, Whitaker RC. Neighborhood safety, collective efficacy, and obesity in women with young children. *Obesity (Silver Spring, Md.).* Mar 2006;14(3):518-525.

58. 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.* Aug 2013;91:15-24.

59. Bennett GG, McNeill LH, Wolin KY, Duncan DT, Puleo E, Emmons KM. Safe to walk? Neighborhood safety and physical activity among public housing residents. *PLoS medicine.* Oct 2007;4(10):1599-1606; discussion 1607.

60. Fish JS, Ettner S, Ang A, Brown AF. Association of perceived neighborhood safety with [corrected] body mass index. *Am. J. Public Health.* Nov 2010;100(11):2296-2303.

61. Reis JP, Hankinson AL, Loria CM, et al. Duration of abdominal obesity beginning in young adulthood and incident diabetes through middle age: the CARDIA study. *Diabetes Care.* May 2013;36(5):1241-1247.

62. Cole SR, Hernan MA. Fallibility in estimating direct effects. *Int. J. Epidemiol.* Feb 2002;31(1):163-165.

63. Blakely T. Commentary: estimating direct and indirect effects-fallible in theory, but in the real world? *Int. J. Epidemiol.* Feb 2002;31(1):166-167.

64. le Cessie S, Debeij J, Rosendaal FR, Cannegieter SC, Vandenbroucke JP. Quantification of bias in direct effects estimates due to different types of measurement error in the mediator. *Epidemiology.* Jul 2012;23(4):551-560.

65. Diez Roux AV. Estimating neighborhood health effects: the challenges of causal inference in a complex world. *Soc. Sci. Med.* May 2004;58(10):1953-1960.

66. Oakes JM. The (mis)estimation of neighborhood effects: causal inference for a practicable social epidemiology. *Soc. Sci. Med.* May 2004;58(10):1929-1952.

67. Sampson RJ, Sharkey P. Neighborhood selection and the social reproduction of concentrated racial inequality. *Demography.* Feb 2008;45(1):1-29.

68. Cummins S. Commentary: investigating neighbourhood effects on health--avoiding the 'local trap'. *Int. J. Epidemiol.* Apr 2007;36(2):355-357.

69. Macintyre S. Deprivation amplification revisited; or, is it always true that poorer places have poorer access to resources for healthy diets and physical activity? *Int J Behav Nutr Phys Act.* 2007;4:32.

70. Franco M, Diez Roux AV, Glass TA, Caballero B, Brancati FL. Neighborhood characteristics and availability of healthy foods in Baltimore. *Am. J. Prev. Med.* Dec 2008;35(6):561-567.

71. Duncan DT, Kawachi I, White K, Williams DR. The geography of recreational open space: influence of neighborhood racial composition and neighborhood poverty. *J. Urban Health.* Aug 2013;90(4):618-631.

72. Estabrooks PA, Lee RE, Gyurcsik NC. Resources for physical activity participation: does availability and accessibility differ by neighborhood socioeconomic status? *nn. Behav. Med.* Spring 2003;25(2):100-104.

Table 2.1 Neighborhood measures for healthy food, physical activity, and social environments, Multi-Ethnic Study of Atherosclerosis, 2000-2012

Neighborhood Summary and Component Measures	**Scale**[a]
Healthy food environment summary score	Sum of standardized component measures
GIS-based density of supermarkets/fruit and vegetable markets	Number of food stores likely to sell healthier foods (supermarkets, fruit and vegetable markets) per square mile
Survey-based healthy food availability	Likert scale, 1-5 (example: *"A large selection of fresh fruits and vegetables is available in my neighborhood"*)
Physical activity environment summary score	Sum of standardized component measures
GIS-based density of commercial recreational establishments	Number of commercial recreational establishments (gyms, pools, etc.) per square mile
Survey-based walking environment	Likert scale, 1-5 (example: *"My neighborhood offers many opportunities to be physically active"*)
Social environment summary score	Sum of standardized component measures
Survey-based social cohesion	Likert scale, 1-5 (example: *"People in my neighborhood can be trusted"*)
Survey-based safety	Likert scale, 1-5 (example: *"I feel safe walking in my neighborhood, day or night"*)

Abbreviations: GIS, geographic information system

[a] All measures are constructed such that higher values indicate more favorable environments

Table 2.2 Baseline sociodemographic, behavioral, and type 2 diabetes risk factor characteristics for the total study population, incident diabetes cases, and non-cases, Multi-Ethnic Study of Atherosclerosis, 2000-2012

		Type 2 Diabetes During Follow-up	
	Total Sample	**Yes**	**No**
No. of participants	5124	616	4508
Sociodemographics			
Age, mean (SD)	60.7 (9.9)	60.9 (9.6)	60.7 (9.9)
Female, No. (%)	2747 (53.6)	325 (52.8)	2422 (53.7)
Race/ethnicity, No. (%)			
White	2168 (42.3)	190 (30.8)	1978 (43.9)
Black	1311 (25.6)	190 (30.8)	1121 (24.9)
Hispanic	1041 (20.3)	161 (26.1)	880 (19.5)
Chinese American	604 (11.8)	75 (12.2)	529 (11.7)
Household per capita income, mean (SD), per $10,000	51.8 (34.4)	48.0 (32.5)	52.4 (34.6)
Education, mean (SD), y	13.4 (3.8)	13.0 (4.0)	13.5 (3.8)
Behavioral Characteristics and Risk Factors			
Smoking status, No. (%)			
Former	1892 (36.9)	234 (38.0)	1658 (36.8)
Current	650 (12.7)	70 (11.4)	580 (12.9)
Alcohol use, No. (%)[a]			
Moderate	1582 (30.9)	152 (24.7)	1430 (31.7)
Heavy	419 (8.2)	29 (4.7)	390 (8.7)
Alternative Healthy Eating Index 2010, mean (SD)[b]	52.1 (11.7)	50.8 (11.4)	52.2 (11.8)
Intentional physical activity, No. (%)[c]			
Low	1821 (35.5)	248 (40.3)	1573 (34.9)
Middle	1599 (31.2)	186 (30.2)	1413 (31.3)
High	1704 (33.3)	182 (29.6)	1522 (33.8)
Body mass index, No. (%)			
Normal (18-<25)	1568 (30.6)	77 (12.5)	1491 (33.1)
Overweight (25-<30)	2044 (39.9)	221 (35.9)	1823 (40.4)
Obese (≥30)	1512 (29.5)	318 (51.6)	1194 (26.5)
Family history type 2 diabetes, No. (%)	1791 (35.0)	298 (48.4)	1493 (33.1)
Neighborhood Characteristics			
Socioeconomic index, mean (SD)[d]	0.5 (1.3)	0.2 (1.2)	0.6 (1.3)
Healthy food environment, median (IQR)			
GIS-based supermarkets/fruit and vegetable markets[e]	1.0 (2.2)	1.0 (1.9)	1.0 (2.2)
Survey-based measure[f]	3.5 (0.7)	3.4 (0.6)	3.5 (0.7)
Summary measure[g]	-0.2 (2.1)	-0.4 (2.0)	-0.3 (2.2)
Physical activity environment, median (IQR)			
GIS-based commercial recreational establishments[e]	1.9 (2.9)	1.9 (2.7)	2.1 (2.9)
Survey-based measure[f]	3.9 (0.4)	3.8 (0.3)	3.9 (0.4)
Summary measure[g]	-0.1 (1.2)	-0.5 (1.0)	-0.4 (1.2)
Social environment, median (IQR)			
Survey-based social cohesion[f]	3.5 (0.3)	3.5 (0.4)	3.5 (0.3)
Survey-based safety[f]	3.7 (0.7)	3.6 (0.7)	3.7 (0.7)
Summary measure[g]	-0.0 (2.1)	-0.1 (2.5)	-0.0 (2.0)

Abbreviations: No., number; SD, standard deviation; IQR, interquartile range; GIS, geographic information system

[a] Alcohol use defined according to National Institute on Alcohol Abuse and Alcoholism definitions for men and women. Moderate drinking is defined as no more than 4 drinks on any single day and no more than 14 drinks per week for men, and no more than 3 drinks on any single day and no more than 7 drinks per week for women. Heavy drinking is defined as consumption in excess of moderate.
[b] The Alternative Healthy Eating Index 2010 is an index designed to capture a "healthy diet", and was compiled based upon a food frequency questionnaire. The index ranges from 2.5 to 87.5, and higher scores indicate a better quality diet (high intake of fruits, vegetables, soy, protein, white meat, cereal fiber, polyunsaturated fat, and multivitamins, and lower intake of alcohol, saturated fat and red meat). Some individuals are missing data for Alternative Healthy Eating Index 2010 (n=595).
[c] Refers to moderate and vigorous intentional physical activity, including walking for exercise, dance, team sports (e.g. basketball, softball), dual sports (e.g. tennis), individual activities (e.g. golf, yoga), and conditioning activities (e.g. running, swimming, cycling). Physical activity is measured in metabolic equivalent of task minutes per week (MET-min/week), and is categorized into tertiles for descriptive purposes.
[d] The neighborhood socioeconomic index includes census tract information on percent with a Bachelor's degree, percent in a managerial occupation, median home value, percent with a high school education, percent with interest, dividend, or rental income, median household income, and percent with household income > $50,000. A higher value indicates higher socioeconomic status.
[e] Number of supermarkets/fruit and vegetable markets or commercial recreational establishments per square mile within a 1-mile buffer of the participant's residential address.
[f] Survey score based upon Likert scale (1-5) rankings of healthy food and physical activity resource availability, and neighborhood social cohesion and safety within 1-mile of participant's residential address; higher scores indicate more favorable environments.
[g] Sum of standardized component measures (GIS- and survey-based measures for food and physical activity environment, or social cohesion and safety surveys for social environment); higher scores indicate more favorable environments.
[h] Number of commercial recreational and physical activity establishments, including gyms, dance studios, places to play team and racquet sports, and water-related activities, per square mile within a 1-mile buffer of the participant's residential address.

Table 2.3 Baseline sociodemographic, behavioral, and type 2 diabetes risk factor characteristics by tertiles of baseline neighborhood healthy food, physical activity, and social environment summary measures, Multi-Ethnic Study of Atherosclerosis, 2000-2012

	Total Sample	Summary Healthy Food Environment			Summary Physical Activity Environment			Summary Social Environment		
		Low	Middle	High	Low	Middle	High	Low	Middle	High
No. of participants	5124	1704	1739	1681	1785	1705	1634	1793	1680	1651
Sociodemographics										
Age, mean (SD)	60.7 (9.9)	60.7 (10.0)	61.9 (10.2)	61.8 (10.3)	61.0 (10.0)	61.3 (10.3)	62.3 (10.2)	60.9 (10.3)	61.9 (10.4)	61.7 (9.8)
Female, %	53.6	53.6	51.7	55.6	52.2	56.1	52.6	55.3	53.8	51.6
Race/ethnicity, %										
White	42.3	51.5	36.6	39.0	30.1	43.4	54.5	21.3	43.7	63.8
Black	25.6	32.2	22.2	22.4	34.5	19.1	22.6	38.4	17.7	19.7
Hispanic	20.3	13.7	18.2	29.3	22.3	37.8	15.3	32.5	19.1	8.3
Chinese American	11.8	2.6	23.1	9.4	13.2	14.4	7.5	7.8	19.5	8.2
Household per capita income, mean (SD), per $10,000	51.8 (34.4)	51.6 (32.2)	49.3 (34.0)	54.7 (36.7)	43.5 (30.2)	48.2 (32.4)	64.4 (37.0)	39.6 (29.1)	49.5 (32.9)	67.4 (35.2)
Education, mean (SD), y	13.4 (3.8)	13.6 (3.4)	13.2 (3.9)	13.5 (4.2)	12.6 (4.0)	13.2 (3.8)	14.6 (3.4)	12.2 (4.2)	13.5 (3.8)	14.7 (2.9)
Risk Factors										
Smoking status, %										
Former	36.9	39.5	34.5	36.8	35.6	34.1	41.3	34.5	36.2	40.3
Current	12.7	14.5	11.6	12.0	14.7	12.4	10.8	16.7	11.4	9.6
Alcohol use, %[a]										
Moderate	30.9	29.3	30.8	32.5	24.3	29.6	39.4	24.5	32.0	36.6
Heavy	8.2	7.5	6.9	10.2	5.4	7.3	12.1	6.5	9.1	9.2
Alternative Healthy Eating Index 2010, mean (SD)[b]	52.1 (11.7)	50.0 (11.8)	52.5 (11.2)	53.8 (11.9)	51.0 (11.5)	51.5 (11.7)	53.7 (12.0)	50.6 (11.6)	52.5 (11.6)	53.1 (12.0)
Intentional physical activity, %[c]										
Low	35.5	37.7	37.0	31.8	43.4	35.1	27.4	40.6	36.4	29.3
Middle	31.2	30.9	32.4	30.3	30.1	31.8	31.8	28.5	33.5	31.9
High	33.3	31.5	30.6	37.8	26.5	33.1	40.8	30.1	30.2	39.9
Body mass index, %										
Overweight (25-<30)	39.9	40.3	40.6	38.8	40.1	39.5	40.0	39.2	39.0	41.6
Obese (≥30)	29.5	35.9	26.3	26.3	31.9	30.1	25.2	34.8	28.1	25.3

	Total Sample	Summary Healthy Food Environment			Summary Physical Activity Environment			Summary Social Environment		
		Low	Middle	High	Low	Middle	High	Low	Middle	High
Family history type 2 diabetes, %	35.0	39.0	35.4	30.4	38.4	34.1	32.1	35.7	35.4	33.7
Neighborhood Characteristics										
Socioeconomic index, mean (SD)[d]	0.5 (1.3)	-0.0 (0.8)	0.4 (0.9)	1.3 (1.6)	-0.2 (0.8)	0.3 (0.9)	1.6 (1.4)	-0.4 (0.9)	0.4 (1.4)	1.2 (1.4)

Abbreviations: SD, standard deviation; y, years;

[a] Alcohol use defined according to National Institute on Alcohol Abuse and Alcoholism definitions for men and women (see Table 2).

[b] Some individuals are missing data for Alternative Healthy Eating Index 2010 (n=595; see Table 2).

[c] Refers to moderate and vigorous intentional physical activity. Measured in MET-min/week and categorized into tertiles (see Table 2).

[d] Higher values indicate higher socioeconomic status (see Table 2).

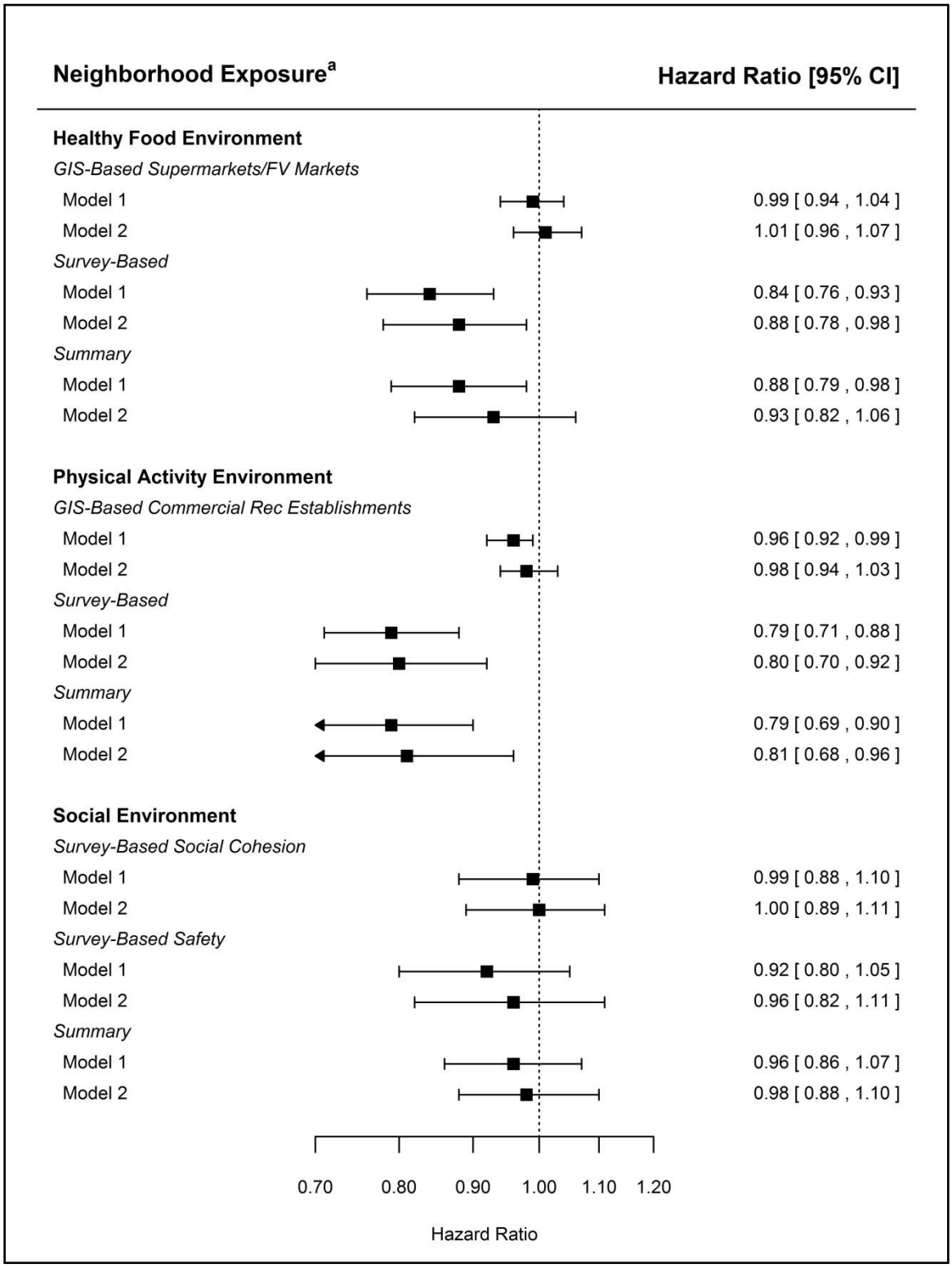
Table 2.4 Crude incidence rates of type 2 diabetes by tertiles of neighborhood food, physical activity, and social environment summary measures at baseline, Multi-Ethnic Study of Atherosclerosis, 2000-2012[a]

Neighborhood Measure	Incidence per 1000 person-years (95% CI)		
	Low Tertile, Worst	Middle Tertile	High Tertile, Best
Healthy Food Environment			
GIS-based supermarkets/FV markets	17.9 (15.7, 20.4)	15.8 (13.7, 18.1)	15.8 (13.7, 18.1)
Survey-based	17.5 (15.3, 20.0)	19.8 (17.5, 22.5)	12.1 (10.3, 14.1)
Summary	16.9 (14.8, 19.3)	18.2 (16.0, 20.8)	14.3 (12.3, 16.6)
Physical Activity Environment			
GIS-based commercial recreational establishments	20.3 (17.8, 23.3)	14.6 (12.8, 16.7)	15.4 (13.4, 17.8)
Survey-based	20.8 (18.4, 23.5)	17.8 (15.6, 20.3)	10.6 (9.0, 12.7)
Summary	20.5 (18.2, 23.2)	17.1 (15.0, 19.6)	11.8 (10.0, 13.8)
Social Environment			
Survey-based social cohesion	18.5 (16.3, 21.1)	14.6 (12.7, 16.9)	16.3 (14.2, 18.8)
Survey-based safety	18.7 (16.5, 21.3)	17.3 (15.2, 19.8)	13.4 (11.5, 15.6)
Summary	19.7 (17.4, 22.3)	15.7 (13.6, 18.1)	14.0 (12.1, 16.2)

Abbreviations: GIS, geographic information system

[a] Incidence rates were calculated using Poisson regression according to tertiles of the neighborhood exposures at baseline. Overall incidence rate in the full sample was 16.5 per 1000 person-years (95% CI; 15.2, 17.8).

Figure 2.1 Adjusted hazard ratios for type 2 diabetes incidence corresponding to an interquartile range increase in exposure to neighborhood resources, Multi-Ethnic Study of Atherosclerosis, 2000-2012



Abbreviation: GIS, geographic information system; Rec, recreational

[a] Model 1 adjusts for baseline age, gender, family history of diabetes, household per capita income, education, smoking status and alcohol consumption. Model 2 adjusts for all covariates in model 1, and adds neighborhood socioeconomic status. All exposures correspond to cumulative average exposures over time. 1 interquartile range (IQR) corresponds to the following changes for each exposure: GIS-based supermarkets/FV markets (IQR=2.2); Survey-based healthy food (IQR=0.6); Combined healthy food (IQR=2.1); GIS-based commercial recreational establishments (IQR=3.2); Survey-based physical activity (IQR=0.4); Combined physical activity (IQR=1.2); Social Cohesion (IQR=0.3); Safety (IQR=0.7); Combined social environment (IQR=2.0)

Appendix

Text 2.A1: Further description of the neighborhood GIS, survey, and summary measures for healthy food, physical activity, and social environments

GIS-based measures of access to food stores were created using data obtained from the National Establishment Time Series (NETS) database from Walls and Associates for the years 2000-2012. This data includes time-series data on establishments derived from Dun and Bradstreet (D&B) archival establishment data. Addresses were geocoded using TeleAtlas EZ-Locate web-based geocoding software (TeleAtlas, 2011). We used Standard Industrial Classification (SIC) codes to identify supermarkets and grocery stores (#5411), and fruit and vegetable markets (#5431), which we classified as healthy food stores.[1] Additional supermarket data was obtained from Nielsen/TDLinx to enhance the supermarket list.[2] We identified supermarkets as grocery stores with at least $2 million in annual sales or at least 25 employees. Additionally, we included supermarkets that had a standard chain name based on a list derived from the Nielsen/TDLinx data as described in detail elsewhere.[3] For physical activity resources, 114 SIC codes were selected to represent establishments with indoor conditioning, dance, bowling, golf, team and racquet sports, and water activities derived from lists used in previous studies.[4,5] Simple densities per square mile were created for 1-mile buffers around each address using the point density command in ArcGIS 9.3.

For the survey scales, information on neighborhood level characteristics was ascertained via questionnaire asking participants to rate the area within approximately 1 mile around their home. On the basis of a conceptual model[6] and prior work,[7] four neighborhood dimensions were assessed: walking environment (4 items, “It is pleasant to walk in my neighborhood”, “In my neighborhood it is easy to walk to places”, “I often

see other people walking in my neighborhood", and "I often see other people exercise in my neighborhood"), availability of healthy foods (2 items, "A large selection of fresh fruit and vegetables is available in my neighborhood" and "A large selection of low fat foods is available in my neighborhood"), safety (2 items, "I feel safe walking in my neighborhood day or night" and "Violence is a problem in my neighborhood"), and social cohesion (4 items, "People around here are willing to help their neighbors", "People in my neighborhood generally get along with each other", "People in my neighborhood can be trusted", and "People in my neighborhood share the same values"). Responses for each item ranged from 1 (strongly agree) to 5 (strongly disagree). Questions were reverse coded when needed to indicate a higher score being a more positive or favorable environment. Scales were based on previous work and have acceptable internal consistency (Cronbach alpha 0.64-0.82).[8] Scales based on a 1-mile buffer around the MESA participant's home address were created by taking the crude mean of the responses for all respondents living within a 1 mile buffer, excluding themselves. Respondents had to have answered all questions within the domain to be included.

To create the summary measures, we standardized the GIS- and survey-based measures by centering each measure at the sample mean and dividing by the standard deviation. We then summed the standardized measures corresponding to each domain (e.g. GIS-based supermarket/fruit and vegetable market availability and survey-based fruit and vegetable availability) to create a summary measure.

Table 2.A1 Distribution of number of respondents to the survey questionnaires used for creating each individual participant's survey-based exposure measures

	Proportion of index participants with a given number of respondents used to create survey-based exposures (%)											
	1-mile buffer						Census tract buffer					
Number of respondents used to create survey measure[a]	**1**	**2**	**3**	**4**	**≥5**	**Median**	**1**	**2**	**3**	**4**	**≥5**	**Median**
Survey-based healthy food environment	1.6	0.9	1.2	1.2	95.0	78	4.0	5.9	4.1	3.4	82.6	20
Survey-based physical activity environment	1.6	0.9	1.2	1.2	95.0	78	4.0	5.9	4.1	3.4	82.6	20
Survey-based social cohesion	1.4	1.0	1.0	1.4	95.2	76	3.6	5.8	4.1	3.4	83.1	20
Survey-based safety	1.6	0.9	1.2	1.2	95.0	78	4.0	5.9	4.1	3.4	82.6	20

[a] Note that each individual's own response was excluded from their survey-based exposure measure in order to minimize self-perception bias. Thus, the number of individuals used to create the survey response does not include the individual's own responses to survey questions.

Text 2.A2: Individual diet, physical activity, and body mass index (BMI) measurement

Diet was measured using a food frequency questionnaire administered at baseline and at exam 5. To derive an index of "healthy diet", we used the Alternative Healthy Eating Index – 2010 (AHEI-2010), which has been used in a variety of epidemiologic work due to its strong relationship to major chronic diseases.[9,10] The index ranges from 0 to 110, with higher scores indicating better diet quality (high intake of fruits, vegetables, soy, protein, white meat, cereal fiber, polyunsaturated fat and vitamins, and lower intake of alcohol, saturated fat, and red meat). Typical physical activity was measured at exams 1, 2, 3, and 5 using a standardized, semi-quantitative questionnaire adapted from the Cross-Cultural Activity Participation Study.[11] Physical activity was quantified in metabolic equivalent task minutes per week, and included all moderate and vigorous intentional physical activity, including walking for exercise, dance, team sports (e.g. basketball, softball), dual sports (e.g. tennis), individual activities (e.g. golf, yoga), and conditioning activities (e.g. running, swimming, cycling). BMI was calculated at each exam using measured height (m) and weight (kg). As potential mediators, BMI, diet, and physical activity were added to regression models as time-varying covariates, matching each mediator value to the closest preceding exposure measure. The sensitivity of our results to the use of the AHEI-2010 dietary index was tested by running additional models controlling for specific dietary components linked to type 2 diabetes in our cohort and others: percent of calories consumed from trans fats, whole grain consumption (servings per day), and consumption of nuts and seeds (servings per day). These dietary components were added to models individually and then collectively into a single model.

Text 2.A3: Further description of neighborhood socioeconomic status index

Neighborhood level scales for characteristics of socioeconomic status (SES) were obtained from the U.S. Census 2000 Summary File 1 and Summary File 3, American Community Survey (ACS) 2005-2009, and ACS 2007-2011 estimates at the census tract level. We conducted principal factor analysis with orthogonal rotation of 21 census variables which reflect aspects of race/ethnicity (percent Hispanic, percent non-Hispanic Asian, and percent non-Hispanic black), crowding (percent of households with crowing greater than 1 person per room), foreign born (percent or persons who are foreign born), education (percent of adults age 25 or older with at least a high school education and percent of adults age 25 or older with at least a Bachelor's degree), occupation (percent of persons age 16 and older with executive, managerial, or professional occupation), income and wealth (median value of housing units, percent of housing units without a telephone, percent of housing units without a vehicle, median household income, percent of households with income of at least $50,000, percent of household with interest, dividend, or net rental income, and percent of household receiving public assistance), poverty (percent below poverty level), employment (percent of those age 16 or older who are unemployed and percent of those age 16 and older who are not in the labor force), and housing (percent of occupied housing units, percent of housing units that are owner occupied, and percent of persons living in same house as previous census). Variables that represent a better SES environment were reverse coded. Five factors were kept which reflects 74% of the variance explained. Weighted scales were created by multiplying the factor weights by the standardized variables, and increasing scores represents socioeconomic disadvantage. The first factor, which we used in all analyses,

represents education, occupation, housing value, and income, and was highly weighted on % bachelor degree, % managerial occupation, median home value, % HS education, % interest/dividend/rental income, median household income, and % household income >$50,000. The scales are linked to MESA participants by census tract using Census 2000 data for years 2000-2004, ACS 2005-2009 data for years 2005-2007, and ACS 2007-2011 data for years 2008-2012.

Text 2.A4: Description of models using baseline and change since baseline neighborhood measures as the exposures of interest

The parameterization of longitudinal neighborhood exposures as time-varying cumulative averages in the main models of the paper reflects both theory and biological plausibility regarding how neighborhood exposures are likely to influence the risk for type 2 diabetes, a slow, progressive onset chronic disease. Nonetheless, there is interest in evaluating if change in the neighborhood environment is associated with risk for diabetes. We ran additional Cox proportional hazards models parameterizing the neighborhood exposures as two separate regression coefficients: a baseline value, which estimates the association between the baseline level of exposure and the hazard for developing diabetes, and a change since baseline value, which estimates the association between the change in the level of exposure from baseline to the most recent follow-up exam and the hazard for developing diabetes. All models adjusted for the same covariates as the models in the main paper, and the results of these analyses are presented in Table 2.A7. For simplicity, all hazard ratios and 95% confidence intervals are estimated for a 1-unit change in the exposures.

Table 2.A2 Baseline values and mean 10-year changes for neighborhood health food, physical activity, and social environment measures

Neighborhood Summary and Component Measures	Baseline values, median (IQR)	Mean 10-year changes (95% CI)
Healthy food environment summary score	-0.31 (2.14)	0.83 (0.83, 0.84)
GIS-based density of favorable food stores	0.96 (2.23)	-0.20 (-0.21, -0.19)
Survey-based healthy food availability	3.49 (0.65)	0.48 (0.48, 0.48)
Physical activity environment summary score	-0.48 (1.17)	0.54 (0.53, 0.54)
GIS-based density of physical activity resources	1.91 (2.87)	2.01 (1.98, 2.03)
Survey-based walking environment	3.86 (0.35)	0.09 (0.09, 0.09)
Social environment summary score	-0.03 (2.09)	0.26 (0.25, 0.27)
Survey-based social cohesion	3.54 (0.33)	0.07 (0.07, 0.08)
Survey-based safety	3.68 (0.68)	0.01 (0.01, 0.01)

Table 2.A3 Hazard ratios associated with an IQR increase in cumulative average neighborhood exposures, comparing models with and without BMI, diet and physical activity to evaluate possible mediation

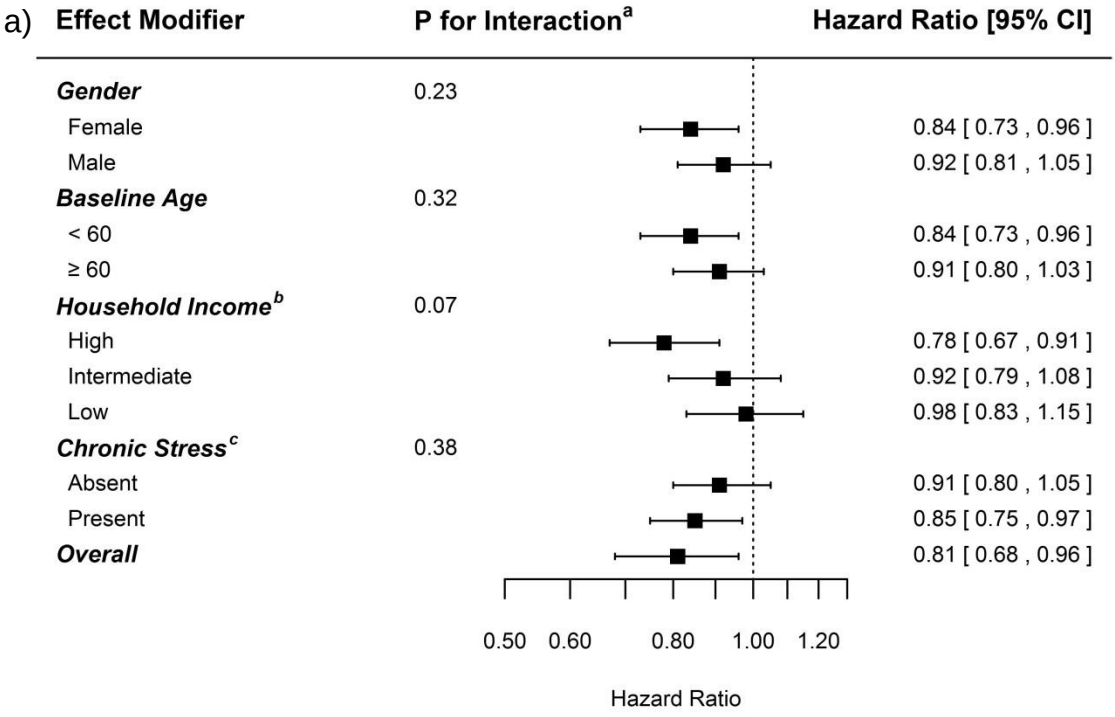
Neighborhood Exposures	**Model 1: All individual-level covariates[a]**	**Model 2: Model 1 + BMI[b]**	**Model 3: Model 2 + diet and physical activity[c]**
	HR (95% CI)	**HR (95% CI)**	**HR (95% CI)**
Healthy Food Environment			
GIS-based supermarkets/FV markets	0.99 (0.94, 1.04)	1.00 (0.95, 1.05)	1.00 (0.95, 1.05)
Survey-based	0.84 (0.76, 0.93)	0.87 (0.78, 0.96)	0.85 (0.77, 0.95)
Summary	0.88 (0.79, 0.98)	0.91 (0.82, 1.02)	0.89 (0.79, 1.01)
Physical Activity Environment			
GIS-based commercial rec establishments	0.96 (0.92, 0.99)	0.97 (0.93, 1.01)	0.97 (0.92, 1.00)
Survey-based	0.79 (0.71, 0.88)	0.81 (0.73, 0.91)	0.80 (0.70, 0.88)
Summary	0.79 (0.69, 0.90)	0.82 (0.71, 0.93)	0.80 (0.68, 0.91)
Social Environment			
Survey-based social cohesion	0.97 (0.77, 1.23)	0.96 (0.75, 1.22)	0.98 (0.75, 1.29)
Survey-based safety	0.96 (0.90, 1.03)	0.96 (0.89, 1.02)	0.96 (0.89, 1.04)
Summary	0.96 (0.86, 1.07)	0.96 (0.85, 1.09)	0.96 (0.85, 1.09)

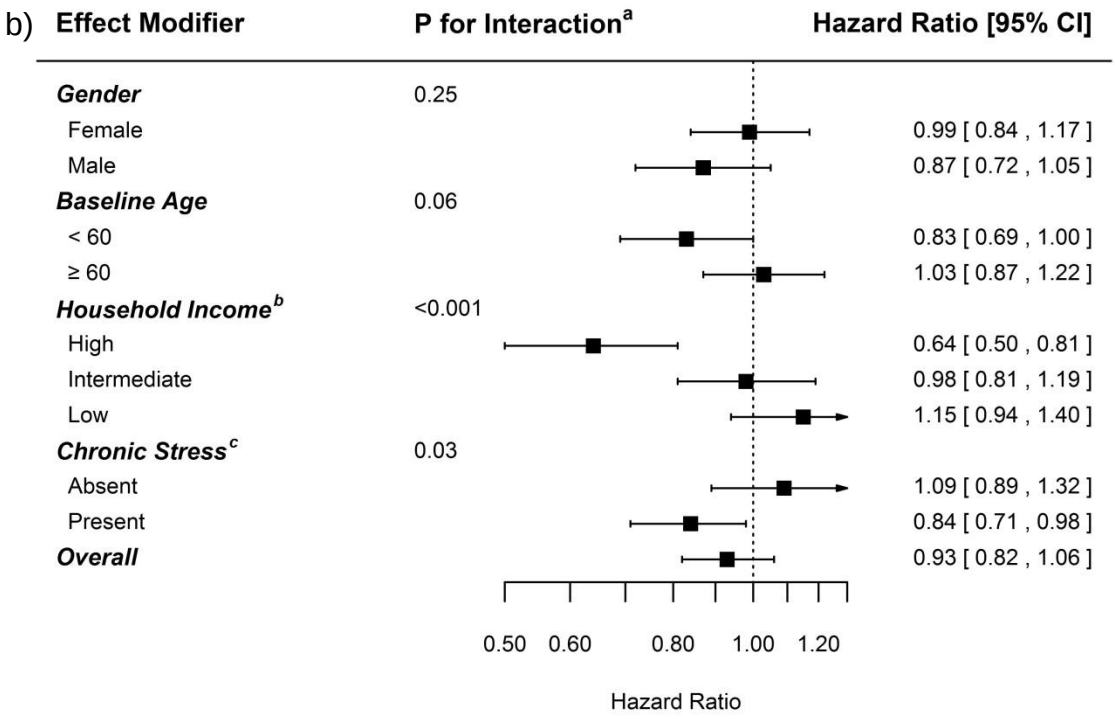
[a] Model 1 is the same as model 1 in the main paper, and controls for baseline age, gender, family history of diabetes, education, household income per capita, race/ethnicity, smoking status, and alcohol consumption.

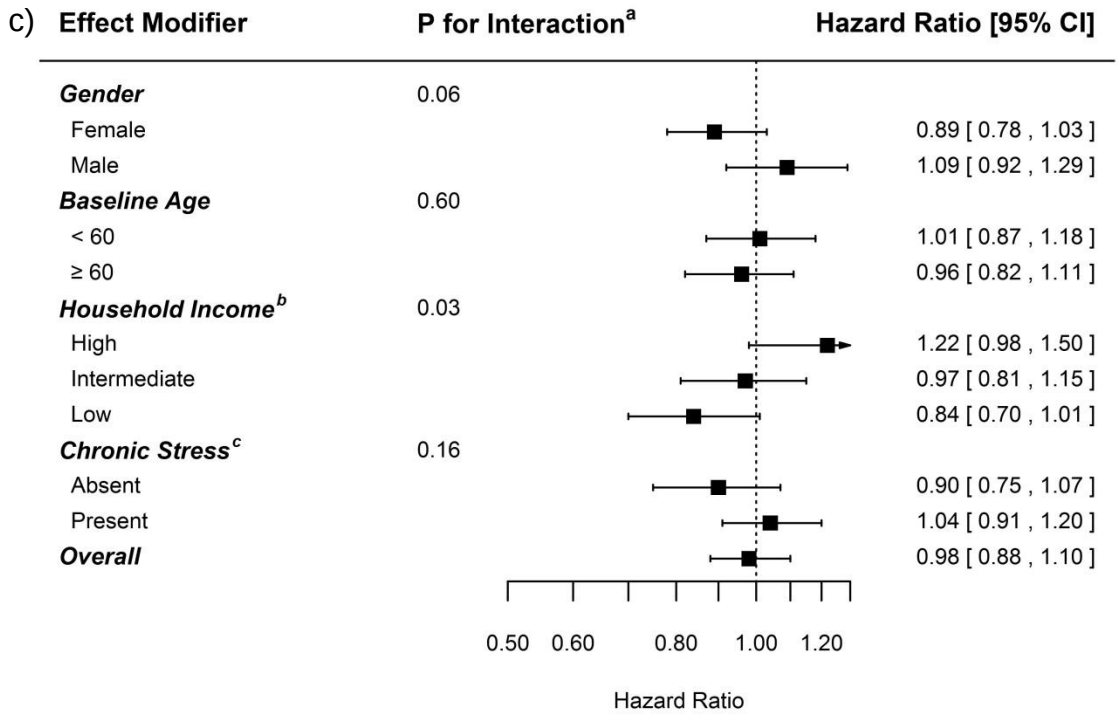
[b] Model 2 controls for all covariates in model 1, and adds time-varying BMI as a potential mediator.

[c] Model 3 controls for all covariates in model 2, and adds diet (measured as the AHEI 2010 dietary index) and physical activity (total intentional physical activity measured in MET-mins/wk) as potential mediators. For additional details regarding the measurement of diet and physical activity, see Text 2.A2.Results when including specific dietary features (% of calories from tans fat, whole grain consumption [servings/day], and nuts/seed consumption [servings/day]) were nearly identical to the results presented in the table.

Figure 2.A1 Effect modification of adjusted hazard ratios for type 2 diabetes incidence for an IQR increase in cumulative neighborhood exposure by gender, baseline age, household income, and chronic stress status for summary a) healthy food, b) physical activity, and social environments







[a] P-values for interaction come from a model adjusting for baseline age, gender, family history of diabetes, household per capita income, education, smoking status, alcohol consumption, and neighborhood SES index, and including an interaction term between the neighborhood exposure and effect modifier of interest. P-values are from Wald Chi-square tests for departures from multiplicative joint effects.

[b] Household income per capita is divided into tertiles.

[c] Chronic stress corresponds to self-reported problems due to money, job status, health concerns, or relationships that have lasted for greater than 6 months. Respondents answering yes to any chronic problems in the domains specified were classified as "present".

Table 2.A4 Hazard ratios associated with an IQR increase in cumulative average neighborhood exposures, using interval censored survival models[a]

Neighborhood Exposures[b]	Model 1: All individual-level covariates[c] HR (95% CI)	Model 2: Model 1 + Neighborhood SES HR (95% CI)
Healthy Food Environment		
GIS-based supermarkets/FV markets	0.99 (0.94, 1.04)	1.04 (0.98, 1.10)
Survey-based	0.65 (0.58, 0.72)	0.67 (0.59, 0.74)
Summary	0.72 (0.63, 0.82)	0.78 (0.67, 0.89)
Physical Activity Environment		
GIS-based commercial rec establishments	0.92 (0.88, 0.97)	0.96 (0.91, 1.01)
Survey-based	0.73 (0.66, 0.81)	0.77 (0.67, 0.87)
Summary	0.80 (0.74, 0.87)	0.83 (0.74, 0.91)
Social Environment		
Survey-based social cohesion	0.94 (0.84, 1.04)	0.94 (0.85, 1.04)
Survey-based safety	0.93 (0.80, 1.06)	0.97 (0.84, 1.10)
Summary	0.94 (0.84, 1.03)	0.96 (0.86, 1.05)

[a] All analyses use accelerated failure time models with a Weibull distribution to account for interval censoring of diabetes events. Standard errors and confidence intervals were calculated using the delta method.

[b] All exposure measures correspond to the most recent cumulative average exposure at the time of interval censoring, or at the end of follow-up for those remaining free of diabetes.

[c] Model 1 controls for baseline age, gender, family history of diabetes, education, and race/ethnicity, and most recently reported household income per capita, smoking status, and alcohol consumption.

Table 2.A5 Sensitivity analyses for adjusted hazard ratios for type 2 diabetes incidence corresponding to an IQR increase in exposure to neighborhood resources[a]

Neighborhood Exposure	Alternative geographic scale[b]	Control for population density[c]	Control for study site	Shared frailty models[d]	1-year lagged exposure[e]
	HR (95% CI)	HR (95% CI)	HR (95% CI)	HR (95% CI)	HR (95% CI)
Healthy food environment					
GIS-based supermarkets/FV markets	1.00 (0.97, 1.04)	0.99 (0.87, 1.14)	0.96 (0.88, 1.06)	1.00 (0.95, 1.05)	0.99 (0.94, 1.04)
Survey-based	0.83 (0.74, 0.94)	0.83 (0.75, 0.92)	0.85 (0.75, 0.97)	0.88 (0.79, 0.97)	
Summary	0.89 (0.79, 1.00)	0.74 (0.63, 0.88)	0.80 (0.67, 0.96)	0.92 (0.82, 1.02)	
Physical activity environment					
GIS-based commercial rec establishments	0.96 (0.92, 1.00)	0.93 (0.89, 0.98)	0.95 (0.91, 1.00)	0.94 (0.90, 0.98)	0.95 (0.92, 0.99)
Survey-based	0.81 (0.72, 0.92)	0.78 (0.70, 0.87)	0.80 (0.71, 0.90)	0.83 (0.75, 0.92)	
Summary	0.87 (0.79, 0.95)	0.74 (0.64, 0.86)	0.78 (0.67, 0.91)	0.85 (0.80, 0.92)	
Social environment					
Survey-based social cohesion	1.03 (0.92, 1.17)	0.96 (0.84, 1.10)	0.92 (0.80, 1.05)	[f]	
Survey-based safety	1.02 (0.96, 1.09)	0.90 (0.78, 1.04)	0.85 (0.73, 0.99)	0.93 (0.81, 1.08)	
Summary	1.05 (0.92, 1.19)	0.93 (0.83, 1.05)	0.89 (0.78, 1.01)	0.94 (0.85, 1.05)	

[a] All models control for baseline age, gender, family history of diabetes, education, household income per capita, race/ethnicity, smoking status, and alcohol consumption.
[b] For GIS-based measures, simple 3-mile buffers were used. For survey-based measures, including social cohesion and safety, and summary measures, census tracts were used. Alternative geographic scale measures were created in the same manner as those described in the methods section.
[c] Population density, measured as persons per square mile within a 1-mile buffer of the participant's address, was calculated based on block-level census population. Each block was weighted by the percent of the block area that falls within the participant buffer. The total population within that block was then multiplied by this weight and the weighted populations were summed together for the total population within the buffer. The total population was divided by total buffer area in square miles. For dates prior to January 2006, population counts originated from the 2000 Census (Census, 2000). For dates on and after January 2006, population counts originated from the 2010 Census.
[d] Shared frailty models are the random effects analogue of the Cox models presented in the main analyses. Rather than using robust standard errors to account for geographic clustering of the outcome, the shared frailty models use a random intercept for each census tract to account for geographic clustering of incident cases within census tracts. The advantage to including the random intercept for census tract is that it may help to control for residual confounding at the neighborhood level due to unmeasured or mismeasured factors (e.g. confounding not accounted for by covariates in our model, such as socioeconomic index).[12] The disadvantage is that such models assume homogeneity of unobserved factors within census tracts, which may be incorrect, especially in larger census tracts. All shared frailty models assumed a lognormal frailty distribution.
[e] 1-year lagged exposures were only available for GIS-based measures, since these exposures were collected annually. Survey-based measures were not collected annually, and hence comparable exposure measures could not be created.
[f] Shared frailty models for social cohesion failed to converge (a recognized problem with such models[13])

Table 2.A6 Hazard ratios associated with IQR increase in cumulative average neighborhood exposures, with additional adjustment for diabetes risk factors at baseline

Neighborhood Exposures	Model 1: All individual-level covariates[a]	Model 2: Model 1 + baseline BMI	Model 2: Model 1 + baseline hypertension[b]	Model 4: Model 1 + baseline high cholesterol[c]	Model 5: Model 1 + baseline BMI, hypertension, and high cholesterol
	HR (95% CI)	HR (95% CI)	HR (95% CI)	HR (95% CI)	HR (95% CI)
Healthy Food Environment					
GIS-based supermarkets/ FV markets	0.99 (0.94, 1.04)	1.00 (0.95, 1.05)	0.99 (0.94, 1.04)	0.99 (0.95, 1.05)	1.00 (0.95, 1.05)
Survey-based	0.84 (0.76, 0.93)	0.86 (0.78, 0.96)	0.85 (0.77, 0.94)	0.84 (0.76, 0.92)	0.86 (0.78, 0.95)
Summary	0.88 (0.79, 0.98)	0.91 (0.81, 1.02)	0.89 (0.79, 0.99)	0.88 (0.79, 0.99)	0.91 (0.81, 1.02)
Physical Activity Environment					
GIS-based commercial rec establishments	0.96 (0.92, 0.99)	0.96 (0.92, 1.01)	0.96 (0.93, 1.00)	0.96 (0.92, 0.99)	0.96 (0.92, 1.01)
Survey-based	0.79 (0.71, 0.88)	0.81 (0.72, 0.90)	0.81 (0.72, 0.90)	0.79 (0.72, 0.88)	0.82 (0.73, 0.91)
Summary	0.79 (0.69, 0.90)	0.88 (0.81, 0.96)	0.88 (0.81, 0.95)	0.87 (0.80, 0.94)	0.89 (0.82, 0.96)
Social Environment					
Survey-based social cohesion	0.97 (0.77, 1.23)	0.99 (0.88, 1.11)	0.98 (0.88, 1.09)	0.99 (0.88, 1.10)	0.98 (0.88, 1.10)
Survey-based safety	0.96 (0.90, 1.03)	0.92 (0.79, 1.06)	0.91 (0.79, 1.05)	0.93 (0.81, 1.07)	0.92 (0.80, 1.06)
Summary	0.96 (0.86, 1.07)	0.96 (0.86, 1.07)	0.95 (0.86, 1.06)	0.96 (0.86, 1.07)	0.96 (0.86, 1.07)

[a] Model 1 controls for baseline age, gender, family history of diabetes, education, household income per capita, race/ethnicity, smoking status, and alcohol consumption.
[b] Baseline hypertension was defined as systolic blood pressure ≥140, or diastolic blood pressure ≥90, or taking antihypertensive medications.
[c] Baseline high cholesterol was defined as LDL cholesterol ≥ 160 or taking cholesterol-lowering medications.

Table 2.A7 Adjusted hazard ratios for type 2 diabetes incidence corresponding to 1-unit increases in baseline and change from baseline exposure measures

	Model 1: All individual-level covariates[a]		Model 2: Model 1 + Neighborhood SES	
Neighborhood Exposure	**Baseline exposure HR (95% CI)**	**Change from baseline exposure HR (95% CI)**	**Baseline exposure HR (95% CI)**	**Change from baseline exposure HR (95% CI)**
Health food environment				
GIS-based supermarkets/ FV markets	0.99 (0.97, 1.02)	1.03 (0.95, 1.10)	1.01 (0.98, 1.03)	1.04 (0.96, 1.12)
Survey-based	0.78 (0.67, 0.90)	0.94 (0.74, 1.19)	0.83 (0.69, 0.99)	0.96 (0.76, 1.22)
Summary	0.94 (0.89, 1.00)	1.00 (0.88, 1.14)	0.97 (0.91, 1.03)	1.01 (0.89, 1.15)
Physical activity environment				
GIS-based commercial rec establishments	0.98 (0.97, 1.00)	1.01 (0.99, 1.03)	0.96 (0.94, 0.99)	1.02 (0.97, 1.06)
Survey-based	0.55 (0.41, 0.73)	0.64 (0.41, 0.99)	0.54 (0.41, 0.73)	0.64 (0.41, 0.99)
Summary	0.89 (0.84, 0.95)	0.92 (0.81, 1.04)	0.90 (0.82, 0.98)	0.92 (0.81, 1.05)
Social environment				
Survey-based social cohesion	0.96 (0.67, 1.37)	0.76 (0.46, 1.26)	0.99 (0.70, 1.41)	0.79 (0.48, 1.29)
Survey-based safety	0.89 (0.72, 1.10)	0.79 (0.55, 1.14)	0.95 (0.76, 1.20)	0.80 (0.56, 1.14)
Summary	0.98 (0.93, 1.04)	0.94 (0.86, 1.04)	0.99 (0.94, 1.05)	0.95 (0.87, 1.04)

[a] Model 1 controls for baseline age, gender, family history of diabetes, education, household income per capita, race/ethnicity, smoking status, and alcohol consumption. Exposures are all parameterized for a 1-unit increase.

Appendix References:

1. U.S. Department of Labor. SIC Manual. Available at: http://www.osha.gov/pls/imis/sic_manual.html Accessed May 2012.

2. Nielsen Company. *Retail Site Database - The Ultimate Source*: Trade Dimensions, a subsidiary of Nielsen Company; 2008.

3. Auchincloss AH, Moore KA, Moore LV, Diez Roux AV. Improving retrospective characterization of the food environment for a large region in the United States during a historic time period. *Health Place.* Nov 2012;18(6):1341-1347.

4. Powell LM, Chaloupka FJ, Slater SJ, Johnston LD, O'Malley PM. The availability of local-area commercial physical activity-related facilities and physical activity among adolescents. *Am. J. Prev. Med.* Oct 2007;33(4 Suppl):S292-300.

5. Gordon-Larsen P, Nelson MC, Page P, Popkin BM. Inequality in the built environment underlies key health disparities in physical activity and obesity. *Pediatrics.* Feb 2006;117(2):417-424.

6. Diez Roux AV. Residential environments and cardiovascular risk. *J. Urban Health.* Dec 2003;80(4):569-589.

7. Echeverria SE, Diez-Roux AV, Link BG. Reliability of self-reported neighborhood characteristics. *J. Urban Health.* Dec 2004;81(4):682-701.

8. Mujahid MS, Diez Roux AV, Morenoff JD, Raghunathan T. Assessing the measurement properties of neighborhood scales: from psychometrics to ecometrics. *Am. J. Epidemiol.* Apr 15 2007;165(8):858-867.

9. Chiuve SE, Fung TT, Rimm EB, et al. Alternative dietary indices both strongly predict risk of chronic disease. *J. Nutr.* Jun 2012;142(6):1009-1018.

10. Samieri C, Sun Q, Townsend MK, et al. The association between dietary patterns at midlife and health in aging: an observational study. *Ann. Intern. Med.* Nov 5 2013;159(9):584-591.

11. Bild DE, Bluemke DA, Burke GL, et al. Multi-ethnic study of atherosclerosis: objectives and design. *Am. J. Epidemiol.* Nov 1 2002;156(9):871-881.

12. Wienke A. *Frailty Models in Survival Analysis*: Taylor & Francis; 2010.

13. Diez Roux AV. Estimating neighborhood health effects: the challenges of causal inference in a complex world. *Soc. Sci. Med.* May 2004;58(10):1953-1960.

CHAPTER 3 :

THE INTERACTION OF NEIGHBORHOOD ENVIRONMENTS AND GENETIC RISK FOR TYPE 2 DIABETES: THE MULTI-ETHNIC STUDY OF ATHEROSCLEROSIS (MESA)

Introduction:

Type 2 diabetes is an etiologically complex disease that affects an estimated 18.8 million adults in the US.[1] With temporal increases in prevalence and incidence over the last 20 years, type 2 diabetes is now the 7th and 8th leading cause of Years of Life Lost and Years Lived with Disability, respectively.[2] Dramatic changes in obesity levels caused by shifts in diet and physical activity patterns are thought to underlie the rise of type 2 diabetes.[3-5] However, type 2 diabetes also has a substantial genetic component, and genome wide association studies (GWAS) have now identified over 70 loci that confer increased risk.[6] Given the importance of both lifestyle and genetics in the development of diabetes, there is growing interest in understanding how these factors may interact to explain population patterns of diabetes.[7]

Prior research exploring gene-environment interactions in type 2 diabetes has been focused largely on individual diet and physical activity. In early studies of biological candidate genes, the effects of several diabetes-associated variants were shown to be attenuated in individuals with higher physical activity levels and specific dietary patterns.[7-10] More recently, genetic risk scores that combine diabetes-associated genetic

variants into a single measure have become available. Gene-environment interaction studies using these risk scores have shown that an individuals' overall genetic risk can be modified by dietary patterns.[7] For instance, in the Health Professionals' Follow-up Study, researchers found that a Western dietary pattern led to increased risk of diabetes in individuals with higher, but not lower, genetic risk scores.[11]

Notwithstanding the importance of individual behaviors in the development of type 2 diabetes, a substantial body of research has demonstrated that such behaviors are partially shaped by the larger social and economic contexts in which people live.[12-14] Factors like neighborhood disadvantage and the spatial patterning of health promoting resources and norms fundamentally support or constrain peoples' abilities to engage in healthy behaviors.[15-17] Despite this recognition and calls in the literature for broader conceptualizations of "environment",[18,19] empirical examinations of the interaction between genetic risk and these larger environmental features remain rare, and no such studies have focused on diabetes. To the extent that area-level factors both shape health behaviors and represent policy-relevant realms for intervention, studies investigating the interaction between genetic risk for type 2 diabetes and broader neighborhood environmental features are needed.[18] Using longitudinal data from the Multi-Ethnic Study of Atherosclerosis (MESA), this study sought to examine if the neighborhood environment, characterized by the availability of healthy food and physical activity resources and neighborhood socioeconomic status (SES), modifies the effect of genetic predisposition for type 2 diabetes, summarized using a diabetes risk score. Our hypothesis was that genetic risk for type 2 diabetes would be most pernicious for persons

residing in neighborhood environments characterized by fewer healthy food and physical activity resources, and lower SES.

Methods:

Study population and analytic sample:

MESA is a longitudinal cohort composed of 6814 non-institutionalized adults (45-84 years at baseline) who self-identify as white, African American, Hispanic, or Chinese. Beginning in 2000, individuals free of clinical cardiovascular disease were recruited from 6 locations (New York, New York; Baltimore, Maryland; Forsyth County, North Carolina; Chicago, Illinois; St. Paul, Minnesota; and Los Angeles, California). Baseline examinations took place from 2000 to 2002, and 4 follow-up exams have occurred an average of 1.6, 3.1, 4.8, and 9.5 years after baseline. Retention rates at exams 2 through 5 were 92%, 89%, 87%, and 76%, respectively.

For this analysis, we use data from participants consenting to both geocoding of their home address as part of the ancillary MESA Neighborhood Study (6191) and genotyping (6429). Combining these datasets yielded 5838 individuals. For our primary analyses, we included individuals with both prevalent and incident diabetes. We excluded individuals who were missing exposure, outcome, or covariate data (n=134), leaving 5704 individuals available for analyses of prevalent and incident diabetes. In supplementary analyses using incident cases only, we excluded 649 individuals with prevalent diabetes at baseline.

Measurement of type 2 diabetes:

The primary outcome was type 2 diabetes identified at each exam according to the American Diabetes Association 2003 criteria[20]: fasting plasma glucose level ≥126 mg/dL

(7 mmol/L), or use of oral anti-hyperglycemic medications or insulin. Glucose levels were obtained from blood samples taken after a 12-hour fast as previously described.[21] Information on the use of oral medications and insulin was obtained by visual inspection of medications, or by self-report on the study questionnaire.

Neighborhood exposure variables:

Based on previous research demonstrating their associations with risk for type 2 diabetes and various health behaviors, we identified three neighborhood exposures of interest: the availability of stores selling healthy food, the availability of recreational establishments, and neighborhood SES.[22-24] Neighborhood-level availability of healthy food and physical activity resources were measured using methods from prior studies.[17,25-28] In brief, we constructed Geographic Information System (GIS)-based measures of access to food stores more likely to sell healthy foods(supermarkets and fruit and vegetable markets) and commercial recreational establishments (facilities for indoor conditioning, dance, bowling, golf, team and racquet sports, and water activities) using annual tax parcel information from the National Establishment Time Series (NETS) database for years 2000-2012 (See Text 3.A1 for details).[29] Simple densities per square mile were created for 1-mile buffers around each participant's residence using ArcGIS, version 9.3 (Esri, Redlands, California). Time-varying densities for each year were linked to participants based on home addresses to account for changes over time and/or participant relocation. One-mile densities were chosen as proxies for neighborhoods based on an area in which most individuals could reasonably walk and federal government definitions of access to services.[30]

An extensive literature exists regarding the definition and measurement of neighborhood-level SES.[31-33] Drawing from previous studies of the relationship between neighborhood disadvantage and health outcomes,[34,35] we selected the following indicators of neighborhood socioeconomic position *a priori* to combine into a summary index: percent of adults age 25 and older with at least a high school education, percent of adults age 25 and older with at least a Bachelor's degree, median household income, percent of residents living below the poverty level, and percent of households receiving public assistance income. Following recommended methods, we created the index by first transforming variables to remove skewness, z-scoring the variables, and then summing the z-scores.[36] All indices were created at the census tract-level and scaled so that an increasing score indicates greater SES. The neighborhood SES index was linked to MESA participants by census tract using Census 2000 data for years 2000-2004, American Community Survey (ACS) 2005-2009 data for years 2005-2007, and ACS 2007-2011 data for years 2008-2012.

Genotyping:

Participants were genotyped on the Affymetrix Human SNP array 6.0 (Affymetrix Inc., Santa Clara, CA). Sample quality control (QC) was based on call rates and contrast QC statistics. Additional details regarding genotyping and quality control have been described elsewhere.[37] Genotypes were imputed using IMPUTE v2.2.2[38] and the 1000 Genomes Phase I integrated variant set (all ancestries)[39] for each ethnicity separately. All SNPs used in the analyses were taken from the 1000 genomes imputation data.

Genetic risk score:

Genome-wide association studies have identified multiple loci associated with the risk for developing type 2 diabetes. While most loci associated with increased diabetes risk were originally found in European-ancestry individuals, recent trans-ethnic meta-analyses and replication studies using non-European-ancestry samples have revealed that many of these loci, or nearby loci in strong linkage disequilibrium, are reproducible across racial/ethnic groups.[6,40] Drawing upon these meta-analyses and previous work demonstrating similar performance of genetic risk scores across racial/ethnic groups in several biracial cohorts,[41-43] we selected 62 SNPs for inclusion in a genetic risk score. Of these 62 SNPs, 55 had an imputation quality of 0.8 or higher (see Table 3.A1 for details). We calculated an unweighted genetic risk score for diabetes as the sum of the number of risk alleles (0, 1, or 2) at each locus for each individual. We chose not to calculate a weighted risk score due to the paucity of GWAS studies with reliable effect sizes in Hispanic and African American individuals. Because of concern about the heterogeneity of the genetic risk score-diabetes association across racial/ethnic group, we constructed a second genetic risk score that was restricted to those SNPs with consistent direction of effects in all racial/ethnic groups as reported in a recent trans-ethnic meta-analysis and a meta-analysis of African American individuals (hereafter referred to as the "restricted genetic risk score").[6,44] After excluding SNPs with imputation quality less than 0.8, 16 SNPs were included in the restricted genetic risk score. Of these 16 SNPs, 13 were part of the original list of 55, and 3 additional SNPs were included based upon the trans-ethnic meta-analysis results.[6] Neither of the genetic risk scores showed significant heterogeneity in their association with diabetes across racial/ethnic groups (p values for raceXgenetic risk score interaction, 0.35 and 0.78 for the full and restricted genetic risk scores,

respectively; Table 3.A2), though the full genetic risk score had a slightly weaker association with diabetes in African Americans compared with other racial/ethnic groups.

Measurement of covariates:

Information on covariates was obtained via administered questionnaire at baseline and follow-up exams. Covariates measured at baseline included sex, race/ethnicity, education, and family history of diabetes. Time-varying information available at baseline and follow-up included age, annual household income, alcohol use (no, moderate, or heavy use according to established guidelines[45]) and smoking status (current, former, or never). Because allele frequencies, and hence disease risk marked by those alleles, can vary across populations of different ancestry, we included 5 eigenvector variables to control for population stratification that is not captured by self-reported race/ethnicity. The eigenvectors were created using principal components analysis (PCA) of the pooled MESA cohort following recommended methods to control for population stratification.[46,47] Ethnic-specific PCAs were also performed and the first 5 eigenvectors used in supplementary, race-specific analyses.

Statistical analysis:

We began by assessing the distribution of sociodemographic and diabetes risk characteristics across categories of both the genetic risk score and neighborhood environments. For descriptive purposes, categories of neighborhood healthy food stores and recreational establishments were defined based upon theoretical differences between resource availability, while neighborhood SES and genetic risk score were categorized as tertiles and quartiles, respectively.

We used parametric, interval censored survival analyses with age as the time scale to model the association between neighborhood environments, genetic risk score, and their interaction with type 2 diabetes. To maximize statistical power and adequately capture the effect of genetic risk on diabetes, which likely operates by causing individuals to develop disease at a younger age, we included both prevalent and incident diabetes cases in our analyses. Prevalent cases were treated as left censored, with age at baseline serving as the upper interval boundary, while incident cases were censored within the interval defined by age at the last diabetes-free exam and age at the exam where diabetes was first reported. Individuals remaining diabetes-free were considered right censored at their age of last follow-up corresponding to either study drop out or administrative censoring at exam 5. We used age as the time scale, as is appropriate when the start of study follow-up is at an arbitrary time point given the exposures of interest.[48-50] We selected a Weibull distribution for the hazard based on graphical evidence (plot of log(-log(Survival))) and model fit (AIC values).[51] More flexible specifications of age did not improve model fit.

Because cumulative neighborhood exposures are hypothesized to be most relevant for disease risk,[15] we parameterized our exposures as cumulative averages, defined as the average across all years between baseline and each follow-up exam. For prevalent cases, this corresponded to neighborhood exposure values at baseline, while for incident cases and those remaining diabetes free, it corresponded to the cumulative average at the most recent exam. To control for possible confounding of both neighborhood and genetic risk score associations, all models adjusted for sex, self-reported race/ethnicity, annual household income, education, alcohol use, cigarette smoking, and the first 5 eigenvectors

from the pooled sample. In models with healthy food stores and recreational establishments as the exposures of interest, we also controlled for neighborhood SES. Models with neighborhood SES as the main exposure did not control for healthy food stores and recreational establishments which are hypothesized to be mediators of the association between neighborhood SES and diabetes. All estimates of association were reported as hazard ratios (HRs) with corresponding 95% confidence intervals (CIs).

We assessed gene-environment interaction in two ways. First, we used continuous measures of both neighborhood exposures and the genetic risk score to estimate HRs for diabetes associated with a 1-unit increase in neighborhood exposure and a 10-allele increase in genetic risk score. We added interaction terms to this model to assess effect modification on the multiplicative scale, and computed p-values using Wald tests. Continuous measures were mean-centered to facilitate interpretation. To illustrate modification of the genetic risk by neighborhood environment, we estimated the HR associated with a 10-allele increase in genetic risk score at the 10^{th} and 90^{th} percentiles of the neighborhood exposure distribution.

Second, because living in a neighborhood environment with no healthy food stores or recreational establishments may be qualitatively different than living in a neighborhood with at least 1 of these resources, we dichotomized the healthy food store and recreational establishment measures into cumulative average < 1 versus ≥ 1. Since no similar theoretical thresholds exist for neighborhood SES or genetic risk score, we dichotomized these exposures at their medians. To assess interaction, we determined whether the joint associations of both neighborhood exposure and genetic risk score were greater than, equal to, or less than the expected joint effects[52,53] on the additive scale

($HR_{10}+HR_{01}-1=HR_{11}$), and multiplicative scale ($HR_{10}*HR_{01}=HR_{11}$). [48,49,54] In line with recent recommendations, we also computed the relative excess risk due to interaction (RERI) as a measure of additive interaction.[53,55] Additive interaction is of particular interest due to its direct public health relevance and its ability to detect the types of synergistic effects that underlie the concept of gene-environment interaction.[55] In all models, standard errors were computed using the delta method[56], and no violations of the proportional hazards assumption were found.

We performed several sensitivity analyses. Because population density and regional norms may affect health behaviors independent of neighborhood resources,[25,57] we ran additional models controlling for population density and study site. For individuals with prevalent diabetes at baseline, it is possible that they changed environments over time such that their exposure measure is misclassified. To assess this possibility and its potential effects on our results, we re-ran the main analyses excluding those with prevalent diabetes. Due to concerns about heterogeneity of the genetic risk score-diabetes association in different racial/ethnic groups, we performed three additional analyses: first, we repeated the main analyses using the restricted genetic risk score with only those SNPs showing consistent directions of effect in all racial/ethnic groups; second, we ran race/ethnicity-specific models controlling for the first 5 ethnic-specific eigenvectors to assess if the direction of the gene-environment interaction was similar across racial/ethnic groups; and third, we added interactions between genetic risk score and self-reported race/ethnicity to the main models. Finally, because gene by neighborhood environment interaction may be a proxy for gene by individual SES

interaction, we ran additional models controlling for the interaction of genetic risk score with both individual income and education.

Results:

The baseline characteristics of the cohort overall and by quartile of genetic risk score are shown in Table 1. Of 5704 individuals included at baseline, 649 had prevalent type 2 diabetes, and another 622 developed type 2 diabetes during follow-up (median follow-up of 9.0 years). The overall sample was 40.6% white, 24.6% African American, 22.5% Hispanic, and 12.3% Chinese. White participants were more likely to be in the lowest quartile of genetic risk score, while Hispanic and Chinese participants were more likely to be in the top quartile. Relative to individuals in the lower three quartiles, individuals in the top quartile of genetic risk had lower annual household income, were more likely to have prevalent diabetes or develop incident diabetes over follow-up, and were more likely to have a family history of diabetes. There was no marked variation in body mass index (BMI), smoking and alcohol use, or neighborhood exposures across quartiles of genetic risk.

Table 2 shows the characteristics of participants by categories of the neighborhood exposures at baseline. In general, residents of neighborhoods with < 1 healthy food store or recreational establishment were more likely to be white and African American than Hispanic or Chinese. Healthy food store and recreational establishment availability was not highly patterned by individual income, education, or genetic risk score, but residents of neighborhoods in the highest category of resource density had a lower prevalence of obesity and family history of diabetes relative to those in the lowest category. For neighborhood SES, relative to those in the lowest tertile, residents of higher

SES neighborhoods were more likely to be white, had higher individual income and education, and lower levels of obesity and cigarette smoking. Prevalent and incident diabetes was less common in the highest categories of neighborhood recreational establishments and SES than in the lowest categories, but no consistent pattern was observed across healthy food store categories.

In models adjusted for gender, race, income, education, alcohol use, cigarette smoking, neighborhood SES, the density of healthy food stores, and the first 5 eigenvectors for population stratification, a 10-allele increase in genetic risk score was associated with a 25% higher risk of diabetes (HR, 1.25; 95% CI, 1.16,1.35; Table 3, Model 1). This elevated risk was consistent across models with different neighborhood exposures. Neighborhood exposures were also associated with diabetes risk, such that 1-unit increases in healthy food store density, recreational establishment density, and neighborhood SES were associated with 3%, 3%, and 6% lower risk for diabetes, respectively (HR, 0.97; 95% CI, 0.95, 0.99; HR, 0.97; 95% CI 0.96, 0.98; and HR, 0.94; 95% CI 0.92, 0.95, respectively).

In models adding an interaction term between neighborhood exposures and genetic risk score, an interaction was observed for healthy food stores such that the association between increasing genetic risk and diabetes was weakened at higher levels of healthy food store density (p=0.05; Table 3, Model 2). The association of genetic risk score with diabetes decreased slightly at higher levels of recreational density and at lower neighborhood SES levels, but these differences were not statistically significant (interaction p=0.28 and 0.33 respectively). Models estimating the association for a 10-allele increase in genetic risk score at the 10^{th} and 90^{th} percentiles of neighborhood

exposure are presented in Figure 1. For healthy food store availability, a 10-allele increase in genetic risk was associated with a 32% higher risk of diabetes at the 10th percentile of availability (HR, 1.32; 95% CI, 1.20, 1.45), but only an 11% higher risk of diabetes at the 90th percentile of availability (HR, 1.11; 95% CI, 0.96, 1.27).

Table 4 shows the observed independent and joint associations of dichotomous neighborhood and genetic risk score exposures, and compares them to the expected joint associations if the exposures were perfectly additive or multiplicative. For individuals with a genetic risk score above the median and living in neighborhoods with < 1 healthy food store on average, the joint HR for diabetes relative to those with genetic risk below the median and living in a neighborhood with 1 or more healthy food stores was 1.92 (95% CI, 1.60, 2.25). This was higher than expected if genetic risk and neighborhood environment acted independently in either an additive or multiplicative manner (p-values for additive and multiplicative interaction, <0.01 and 0.23, respectively). Similar and stronger results were observed for recreational establishments: the observed HR for those jointly exposed to low resource availability and above-median genetic risk was 2.42 (95% CI 1.99, 2.84), again greater than that expected under additive or multiplicative joint effects (p-values for additive and multiplicative interaction, <0.0001 and 0.12, respectively). The joint association of below-median neighborhood SES and genetic risks score revealed an HR of 1.75 (95% CI, 1.43, 2.06), which was quite similar to the expected additive joint effect (p-values for additive and multiplicative interaction, 0.61 and 0.24, respectively).

Sensitivity analyses controlling for population density and study site did not change the interaction results (Table 3.A3). Models restricted to incident cases only

showed associations that were smaller in magnitude for the genetic risk score, consistent with the hypothesis that genetic risk likely functions by causing individuals to develop disease at an earlier age (Table 3.A4). However, the interaction results were consistent in direction for healthy food stores and neighborhood SES compared to models including prevalent cases, though the estimates were predictably less precise given the reduced number of cases. Analyses using the restricted genetic risk score including only SNPs with directionally consistent effects in all racial/ethnic groups also showed similar interaction results for healthy food stores and neighborhood SES (Table 3.A5). In race/ethnicity-specific models controlling for ethnic-specific eigenvectors and using continuous measures of genetic risk score and neighborhood environment, the interaction results were consistent in direction across all racial/ethnic groups for healthy food store and recreational establishment availability (i.e. increasing resource availability associated with decrease in the genetic risk), but not for neighborhood SES (Table 3.A6). Adding genetic risk score by race/ethnicity interactions to the main models did not alter our results (data not shown). Finally, models controlling for genetic risk score interactions with individual income and education produced nearly identical results to those shown in Table 3 (data not shown).

Discussion:

In this longitudinal cohort, we found suggestive evidence that genetic predisposition for type 2 diabetes is modified by the availability of healthy food stores and recreational establishments. Specifically, our analyses suggest that increased neighborhood access to healthy food and recreational establishments may dampen the effects of genetic risk for diabetes, as the association between higher genetic risk and

type 2 diabetes was weaker in environments with greater healthy food resources and recreational establishments. This was especially evident in models with dichotomous exposures, where the joint associations of higher genetic risk and having <1 healthy food store or recreational facility were significantly greater than expected on the additive scale, suggesting a type of synergy between genetic risk and neighborhood environment. Models using continuous exposures demonstrated significant modification of genetic risk by healthy food availability, but results were weaker for recreational establishments. Increasing neighborhood SES slightly increased the association of genetic risk with diabetes, though the interaction between neighborhood SES and genetic risk was not-significant in any of the models. All results were robust to control for multiple individual and neighborhood-level confounders, and the direction of interaction for healthy food stores and recreational establishments was consistent across racial/ethnic groups.

Type 2 diabetes likely results from a complex interplay of genetic susceptibility, behavior, and environmental exposures.[5,58] Previous studies of gene-environment interaction in diabetes have focused almost exclusively on specific individual behaviors.[58] To our knowledge, this is the first study to expand the notion of "environment" to include physical and socioeconomic characteristics of neighborhoods and to evaluate their interaction with individual genetic risk for diabetes. Our finding that neighborhoods with more healthy food stores, and to a lesser extent recreational establishments, dampen the effects of the risk alleles is consistent with studies that have shown that diabetes risk alleles have a stronger association in individuals with low levels of physical activity and in those with dietary patterns characterized by high intake of processed and red meats, high fat dairy, and refined grains.[7,11,59] They are furthermore

consistent with a previous study which showed that the risk of metabolic syndrome associated with a single risk allele was lower for individuals residing in neighborhoods with greater numbers of recreational facilities.[60] These results also add to a growing literature on gene-by-neighborhood environment interaction that has shown neighborhood modification of genetic effects on phenotypes ranging from older adult cognition to adolescent antisocial behavior.[35,61,62]

Neighborhood physical and social environments may modify the genetic risk of type 2 diabetes through several mechanisms. Previous work in the MESA has demonstrated that the availability of supermarkets/fruit and vegetable markets and recreational establishments is related to higher diet quality and intentional physical activity levels, providing plausible behavioral pathways to explain the interaction.[17,63] Our environmental measures may also serve as proxies for historical or current neighborhood conditions or social norms that shape individual health behaviors and social relationships related to diabetes.[18] Recent work demonstrating that cohort of birth may modify genetic risk for obesity suggests that exposure to broadly obesogenic environments over the lifecourse may be most important for gene-environment interaction.[64] We are unable to evaluate if our neighborhood exposures capture such dynamics, though there is research to suggest that neighborhood conditions can be surprisingly stable within and across generations, particularly among those living in the most disadvantaged environments.[65-67]

Models dichotomizing neighborhood and genetic risk exposures showed that living in a neighborhood with less than 1 healthy food store or recreational establishment resulted in substantial increases in diabetes risk that were greater than what would be

expected if the genetic risk or environment variables acted independently.[55,68] Under specific conditions (no unmeasured confounding, monotonic associations for both exposures), this interaction is indicative of synergism,[51,66,67] and suggests that the effect of genetic risk for diabetes may be stronger, or may only become manifest, in neighborhood environments with few health-promoting resources. While we cannot empirically evaluate the assumptions required for such synergism, additional studies showing similar results may strengthen our confidence in the associations observed.

Our study has several strengths. In line with recent calls to expand the notion of environment to include "multilevel, multidimensional, longitudinal" measurements of context,[18,69] we utilized neighborhood-level exposures that help shape and constrain health behaviors that have been the subjects of most prior gene-environment research. Doing so not only places individual risk behaviors like diet and physical activity in context, but helps focus attention on modifiable environmental features to which entire populations are exposed. We also utilized longitudinal data to create cumulative average neighborhood exposure measures and to update covariate values to control for confounding.

Our results should also be viewed in light of several limitations. First, though novel and an improvement over studies focusing only on behaviors without attention to context, our environmental measures are limited, simplistic measures of neighborhood healthy food, physical activity, and socioeconomic environments. Simple counts of healthy food stores and recreational facilities based on tax parcel data ignore important determinants of resource use including quality and affordability, [70-72] and other relevant features of the environment like aesthetic quality and walkability.[73,74] Using such

measures however, allows other studies to attempt to replicate our results. Second, our genetic risk scores, though based on recent meta-analyses, do not necessarily include the strongest SNPs or causal variants in each racial/ethnic group, which may have decreased the strength of our associations and caused slight differences across racial/ethnic groups. Third, the strong correlation between race and neighborhood SES due to racial and socioeconomic segregation made gene-environment associations involving neighborhood SES difficult to interpret, as comparisons of the genetic risk score associations across the distribution of neighborhood SES may inherently involve comparisons across racial/ethnic groups.[75] While our genetic risk scores did not show significant heterogeneity by race/ethnicity, and we made several attempts to minimize potential biases induced by the co-segregation of race/ethnicity and neighborhood resources, additional replication studies with samples large enough to evaluate within-race interactions are needed to increase our confidence in the associations observed.[18] Fourth, as in all observational studies of residential contexts and health outcomes, it is possible that individuals at higher risk for diabetes by virtue of genetics or health behaviors selected or were sorted (via social stratification) into neighborhoods with fewer resources.[76] We attempted to account for this by controlling for known, measured predictors of neighborhood selection, but unobserved confounding remains a possibility. Such unmeasured confounders would, however, not influence our interaction results unless they themselves also interact with the genetic risk score.[77] Finally, we utilized a single sample of mid to late-life adults. To the extent that neighborhood environments or genetic predisposition have a greater influence on diabetes risk earlier in life, we may be missing a critical window to evaluate their interaction. Furthermore, many gene-

environment interactions have a poor replication record, and thus replication of our findings in independent cohorts is needed.[58]

In summary, using a genetic risk score and three theoretically grounded measures of residential neighborhood environments, we found suggestive evidence that greater availability of healthy food stores and recreational establishments modifies genetic risk for type 2 diabetes such that genetic predisposition had a smaller effect in neighborhoods with more health-promoting resources. In light of recent calls for population-based chronic disease prevention,[78] our results suggest that modifying neighborhood environments may represent a useful approach to diabetes prevention that helps offset genetic predisposition, though replication of our findings is needed.

References:

1. Ali MK, Bullard KM, Saaddine JB, Cowie CC, Imperatore G, Gregg EW. Achievement of goals in U.S. diabetes care, 1999-2010. *N. Engl. J. Med.* Apr 25 2013;368(17):1613-1624.

2. Murray CJ, Abraham J, Ali MK, et al. The State of US Health, 1990-2010: Burden of Diseases, Injuries, and Risk Factors. *JAMA.* Jul 10 2013.

3. Hu FB, Manson JE, Stampfer MJ, et al. Diet, lifestyle, and the risk of type 2 diabetes mellitus in women. *N. Engl. J. Med.* Sep 13 2001;345(11):790-797.

4. Mozaffarian D, Kamineni A, Carnethon M, Djousse L, Mukamal KJ, Siscovick D. Lifestyle risk factors and new-onset diabetes mellitus in older adults: the cardiovascular health study. *Arch. Intern. Med.* Apr 27 2009;169(8):798-807.

5. Hu FB. Globalization of diabetes: the role of diet, lifestyle, and genes. *Diabetes Care.* Jun 2011;34(6):1249-1257.

6. Genome-wide trans-ancestry meta-analysis provides insight into the genetic architecture of type 2 diabetes susceptibility. *Nat. Genet.* Mar 2014;46(3):234-244.

7. Cornelis MC, Hu FB. Gene-environment interactions in the development of type 2 diabetes: recent progress and continuing challenges. *Annu. Rev. Nutr.* Aug 21 2012;32:245-259.

8. Cornelis MC, Qi L, Kraft P, Hu FB. TCF7L2, dietary carbohydrate, and risk of type 2 diabetes in US women. *Am. J. Clin. Nutr.* Apr 2009;89(4):1256-1262.

9. Fisher E, Boeing H, Fritsche A, Doering F, Joost HG, Schulze MB. Whole-grain consumption and transcription factor-7-like 2 (TCF7L2) rs7903146: gene-diet interaction in modulating type 2 diabetes risk. *Br. J. Nutr.* Feb 2009;101(4):478-481.

10. Florez JC, Jablonski KA, Bayley N, et al. TCF7L2 polymorphisms and progression to diabetes in the Diabetes Prevention Program. *N. Engl. J. Med.* Jul 20 2006;355(3):241-250.

11. Qi L, Cornelis MC, Zhang C, van Dam RM, Hu FB. Genetic predisposition, Western dietary pattern, and the risk of type 2 diabetes in men. *Am. J. Clin. Nutr.* May 2009;89(5):1453-1458.

12. Williams GH. The determinants of health: structure, context and agency. *Sociol. Health Illn.* 2003;25:131-154.

13. Link BG. Epidemiological sociology and the social shaping of population health. *J. Health Soc. Behav.* Dec 2008;49(4):367-384.

14. Krieger N. Theories for social epidemiology in the 21st century: an ecosocial perspective. *Int. J. Epidemiol.* Aug 2001;30(4):668-677.

15. Diez Roux AV, Mair C. Neighborhoods and health. *Ann. N. Y. Acad. Sci.* Feb 2010;1186:125-145.

16. Franco M, Diez-Roux AV, Nettleton JA, et al. Availability of healthy foods and dietary patterns: the Multi-Ethnic Study of Atherosclerosis. *Am. J. Clin. Nutr.* Mar 2009;89(3):897-904.

17. Moore LV, Diez Roux AV, Nettleton JA, Jacobs DR, Jr. Associations of the local food environment with diet quality--a comparison of assessments based on surveys and geographic information systems: the multi-ethnic study of atherosclerosis. *Am. J. Epidemiol.* Apr 15 2008;167(8):917-924.

18. Boardman JD, Daw J, Freese J. Defining the environment in gene-environment research: lessons from social epidemiology. *Am. J. Public Health.* Oct 2013;103 Suppl 1:S64-72.

19. Glymour MM, Osypuk TL, Rehkopf DH. Invited commentary: Off-roading with social epidemiology--exploration, causation, translation. *Am. J. Epidemiol.* Sep 15 2013;178(6):858-863.

20. Genuth S, Alberti KG, Bennett P, et al. Follow-up report on the diagnosis of diabetes mellitus. *Diabetes Care.* Nov 2003;26(11):3160-3167.

21. Bertoni AG, Burke GL, Owusu JA, et al. Inflammation and the incidence of type 2 diabetes: the Multi-Ethnic Study of Atherosclerosis (MESA). *Diabetes Care.* Apr 2010;33(4):804-810.

22. Auchincloss AH, Diez Roux AV, Brown DG, Erdmann CA, Bertoni AG. Neighborhood resources for physical activity and healthy foods and their association with insulin resistance. *Epidemiology.* Jan 2008;19(1):146-157.

23. Auchincloss AH, Diez Roux AV, Mujahid MS, Shen M, Bertoni AG, Carnethon MR. Neighborhood resources for physical activity and healthy foods and incidence of type 2 diabetes mellitus: the Multi-Ethnic study of Atherosclerosis. *Arch. Intern. Med.* Oct 12 2009;169(18):1698-1704.

24. Schootman M, Andresen EM, Wolinsky FD, et al. The Effect of Adverse Housing and Neighborhood Conditions on the Development of Diabetes Mellitus among Middle-aged African Americans. *Am. J. Epidemiol.* August 15, 2007 2007;166(4):379-387.

25. Boone-Heinonen J, Diez-Roux AV, Goff DC, et al. The neighborhood energy balance equation: does neighborhood food retail environment + physical activity environment = obesity? The CARDIA study. *PloS one.* 2013;8(12):e85141.

26. Auchincloss AH, Moore KA, Moore LV, Diez Roux AV. Improving retrospective characterization of the food environment for a large region in the United States during a historic time period. *Health Place.* Nov 2012;18(6):1341-1347.

27. Moore LV, Diez Roux AV. Associations of neighborhood characteristics with the location and type of food stores. *Am. J. Public Health.* Feb 2006;96(2):325-331.

28. Mujahid MS, Diez Roux AV, Morenoff JD, Raghunathan T. Assessing the measurement properties of neighborhood scales: from psychometrics to ecometrics. *Am. J. Epidemiol.* Apr 15 2007;165(8):858-867.

29. Walls & Associates. National Establishment Time-Series (NETS) Database. 2010; Database Description: www.youreconomy.org/nets/NETSDatabaseDescription.pdf.

30. Ver PLoeg M, Bereneman V, Farrigan T, et al. *Access to affordable and nutritions food - measuring and understanding food deserts and their consequences: Report to congress.* Washington, DC: United States Department of Agriculture;2009.

31. Braveman PA, Cubbin C, Egerter S, et al. Socioeconomic status in health research: One size does not fit all. *JAMA.* 2005;294(22):2879-2888.

32. Krieger N, Chen JT, Waterman PD, Rehkopf DH, Subramanian SV. Painting a truer picture of US socioeconomic and racial/ethnic health inequalities: the Public Health Disparities Geocoding Project. *Am. J. Public Health.* Feb 2005;95(2):312-323.

33. Krieger N, Chen JT, Waterman PD, Rehkopf DH, Subramanian SV. Race/ethnicity, gender, and monitoring socioeconomic gradients in health: a comparison of area-based socioeconomic measures--the public health disparities geocoding project. *Am. J. Public Health.* Oct 2003;93(10):1655-1671.

34. Sampson RJ, Sharkey P, Raudenbush SW. Durable effects of concentrated disadvantage on verbal ability among African-American children. *Proceedings of the National Academy of Sciences.* January 22, 2008 2008;105(3):845-852.

35. Beaver KM, Gibson CL, DeLisi M, Vaughn MG, Wright JP. The Interaction Between Neighborhood Disadvantage and Genetic Factors in the Prediction of Antisocial Outcomes. *Youth Violence and Juvenile Justice.* January 1, 2012 2012;10(1):25-40.

36. Gilthorpe MS. The importance of normalisation in the construction of deprivation indices. *J. Epidemiol. Community Health.* Dec 1995;49 Suppl 2:S45-50.

37. Rasmussen-Torvik LJ, Guo X, Bowden DW, et al. Fasting glucose GWAS candidate region analysis across ethnic groups in the Multiethnic Study of Atherosclerosis (MESA). *Genet. Epidemiol.* May 2012;36(4):384-391.

38. Howie BN, Donnelly P, Marchini J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS genetics.* Jun 2009;5(6):e1000529.

39. An integrated map of genetic variation from 1,092 human genomes. *Nature.* 2012;491(7422):56-65.

40. Anand SS, Meyre D, Pare G, et al. Genetic information and the prediction of incident type 2 diabetes in a high-risk multiethnic population: the EpiDREAM genetic study. *Diabetes Care.* Sep 2013;36(9):2836-2842.

41. Vassy JL, Dasmahapatra P, Meigs JB, et al. Genotype prediction of adult type 2 diabetes from adolescence in a multiracial population. *Pediatrics.* Nov 2012;130(5):e1235-1242.

42. Vassy JL, Durant NH, Kabagambe EK, et al. A genotype risk score predicts type 2 diabetes from young adulthood: the CARDIA study. *Diabetologia.* Oct 2012;55(10):2604-2612.

43. Vassy JL, Hivert MF, Porneala B, et al. Polygenic type 2 diabetes prediction at the limit of common variant detection. *Diabetes.* Feb 11 2014.

44. Ng MC, Shriner D, Chen BH, et al. Meta-analysis of genome-wide association studies in African Americans provides insights into the genetic architecture of type 2 diabetes. *PLoS genetics.* Aug 2014;10(8):e1004517.

45. Allen JP, Wilson V, eds. *Assessing alcohol probelms: A guide for clinicians and researchers.* 2nd ed. Bethesda, MD: National Institute on Alcohol Abuse and Alcoholism; 2003.

46. Patterson N, Price AL, Reich D. Population structure and eigenanalysis. *PLoS genetics.* Dec 2006;2(12):e190.

47. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. *Nat. Genet.* Aug 2006;38(8):904-909.

48. Korn EL, Graubard BI, Midthune D. Time-to-event analysis of longitudinal follow-up of a survey: choice of the time-scale. *Am. J. Epidemiol.* Jan 1 1997;145(1):72-80.

49. Thiebaut AC, Benichou J. Choice of time-scale in Cox's model analysis of epidemiologic cohort data: a simulation study. *Stat. Med.* Dec 30 2004;23(24):3803-3820.

50. Cologne J, Hsu WL, Abbott RD, et al. Proportional hazards regression in epidemiologic follow-up studies: an intuitive consideration of primary time scale. *Epidemiology.* Jul 2012;23(4):565-573.

51. Allison PD. *Survival Analysis Using SAS: A Practical Guide*: SAS Institute; 2010.

52. Rothman KJ, Greenland S, Lash TL. *Modern Epidemiology*: Wolters Kluwer Health/Lippincott Williams & Wilkins; 2008.

53. Knol MJ, VanderWeele TJ. Recommendations for presenting analyses of effect modification and interaction. *Int. J. Epidemiol.* Apr 2012;41(2):514-520.

54. Li R, Chambless L. Test for additive interaction in proportional hazards models. *Ann. Epidemiol.* Mar 2007;17(3):227-236.

55. VanderWeele Tyler J, Knol Mirjam J. A Tutorial on Interaction. *Epidemiologic Methods.* Vol 32014:33.

56. Hosmer DW, Lemeshow S. Confidence interval estimation of interaction. *Epidemiology.* Sep 1992;3(5):452-456.

57. Boone-Heinonen J, Evenson KR, Song Y, Gordon-Larsen P. Research Built and socioeconomic environments: patterning and associations with physical activity in US adolescents. 2010.

58. Franks PW, Pearson E, Florez JC. Gene-environment and gene-treatment interactions in type 2 diabetes: progress, pitfalls, and prospects. *Diabetes Care.* May 2013;36(5):1413-1421.

59. Langenberg C, Sharp SJ, Franks PW, et al. Gene-lifestyle interaction and type 2 diabetes: the EPIC interact case-cohort study. *PLoS medicine.* May 2014;11(5):e1001647.

60. Wang SH, Chen WJ, Chuang LM, Hsiao PC, Liu PH, Hsiao CK. Inference of cross-level interaction between genes and contextual factors in a matched case-control metabolic syndrome study: a Bayesian approach. *PloS one.* 2013;8(2):e56693.

61. Lee BK, Glass TA, James BD, Bandeen-Roche K, Schwartz BS. Neighborhood psychosocial environment, apolipoprotein E genotype, and cognitive function in older adults. *Arch. Gen. Psychiatry.* Mar 2011;68(3):314-321.

62. Boardman JD, Barnes LL, Wilson RS, Evans DA, Mendes de Leon CF. Social disorder, APOE-E4 genotype, and change in cognitive function among older adults living in Chicago. *Soc. Sci. Med.* May 2012;74(10):1584-1590.

63. Ranchod YK, Diez Roux AV, Evenson KR, Sanchez BN, Moore K. Longitudinal associations between neighborhood recreational facilities and change in recreational physical activity in the multi-ethnic study of atherosclerosis, 2000-2007. *Am. J. Epidemiol.* Feb 1 2014;179(3):335-343.

64. Rosenquist JN, Lehrer SF, O'Malley AJ, Zaslavsky AM, Smoller JW, Christakis NA. Cohort of birth modifies the association between FTO genotype and BMI. *Proc. Natl. Acad. Sci. U. S. A.* Jan 13 2015;112(2):354-359.

65. Sampson RJ, Sharkey P. Neighborhood selection and the social reproduction of concentrated racial inequality. *Demography.* Feb 2008;45(1):1-29.

66. Sharkey P, Elwert F. The legacy of disadvantage: multigenerational neighborhood effects on cognitive ability. *AJS.* May 2011;116(6):1934-1981.

67. Patrick Sharkey. The Intergenerational Transmission of Context. *American Journal of Sociology.* 2008;113(4):931-969.

68. VanderWeele TJ. Causal interactions in the proportional hazards model. *Epidemiology.* Sep 2011;22(5):713-717.

69. Bookman EB, McAllister K, Gillanders E, et al. Gene-environment interplay in common complex diseases: forging an integrative model-recommendations from an NIH workshop. *Genet. Epidemiol.* May 2011;35(4):217-225.

70. Caspi CE, Kawachi I, Subramanian SV, Adamkiewicz G, Sorensen G. The relationship between diet and perceived and objective access to supermarkets among low-income housing residents. *Soc. Sci. Med.* Oct 2012;75(7):1254-1262.

71. Anekwe TD, Rahkovsky I. The association between food prices and the blood glucose level of US adults with type 2 diabetes. *Am. J. Public Health.* Apr 2014;104(4):678-685.

72. Beydoun MA, Powell LM, Wang Y. The association of fast food, fruit and vegetable prices with dietary intakes among US adults: is there modification by family income? *Soc. Sci. Med.* Jun 2008;66(11):2218-2229.

73. Brownson RC, Hoehner CM, Day K, Forsyth A, Sallis JF. Measuring the built environment for physical activity: state of the science. *Am. J. Prev. Med.* Apr 2009;36(4 Suppl):S99-123 e112.

74. Sallis JF, Floyd MF, Rodríguez DA, Saelens BE. Role of Built Environments in Physical Activity, Obesity, and Cardiovascular Disease. *Circulation.* February 7, 2012 2012;125(5):729-737.

75. Massey DS, Rothwell J, Domina T. The Changing Bases of Segregation in the United States. *The ANNALS of the American Academy of Political and Social Science.* November 1, 2009 2009;626(1):74-90.

76. Oakes JM. The (mis)estimation of neighborhood effects: causal inference for a practicable social epidemiology. *Soc. Sci. Med.* May 2004;58(10):1929-1952.

77. Vanderweele TJ, Ko YA, Mukherjee B. Environmental confounding in gene-environment interaction studies. *Am. J. Epidemiol.* Jul 1 2013;178(1):144-152.

78. Kindig D, Mullahy J. Comparative Effectiveness-of What? Evaluating Strategies to Improve Population Health. *JAMA.* Aug 25 2010;304(8):901-902.

Table 3.1 Baseline characteristics of MESA study participants overall and by quartile of genetic risk score, Multi-Ethnic Study of Atherosclerosis

Sample Characteristics	Overall	Genetic Risk Score Quartiles[a]			
		Q1	Q2	Q3	Q4
No. of participants	5704	1367	1431	1429	1477
Age, mean (SD)	61.93 (10.13)	62.04 (10.27)	62.19 (10.28)	61.51 (10.01)	62.00 (9.97)
Female, No. (%)	2962 (51.93)	699 (51.13)	749 (52.34)	757 (52.97)	757 (51.25)
Race/ethnicity, No. (%)					
White	2318 (40.64)	670 (49.01)	592 (41.37)	533 (37.30)	523 (35.41)
African American	1401 (24.56)	367 (26.85)	375 (26.77)	344 (24.55)	315 (22.48)
Hispanic	1284 (22.51)	234 (17.12)	294 (20.55)	342 (23.93)	414 (28.03)
Chinese	701 (12.29)	96 (7.02)	170 (11.88)	210 (14.70)	225 (15.23)
Prevalent type 2 diabetes, No. (%)	649 (11.38)	103 (7.53)	163 (11.39)	164 (11.48)	219 (14.83)
Incident type 2 diabetes, No. (%)[b]	622 (12.30)	127 (10.05)	155 (12.22)	170 (13.44)	170 (13.51)
Household income, mean (SD), per $10,000	5.00 (3.42)	5.05 (3.33)	5.00 (3.43)	5.06 (3.42)	4.87 (3.51)
Education, mean (SD), y	13.18 (4.01)	13.44 (3.60)	13.20 (3.86)	13.15 (4.16)	12.97 (4.34)
Behavioral Characteristics and Risk Factors					
Body mass index, No. (%)[c]					
Normal (18-<25)	1686 (29.58)	371 (27.20)	415 (29.04)	443 (31.00)	457 (30.94)
Overweight (25-<30)	2148 (37.69)	512 (37.54)	549 (38.42)	519 (36.32)	568 (38.46)
Obese (≥30)	1865 (32.73)	481 (35.26)	465 (32.54)	467 (32.68)	452 (30.60)
Smoking status, No. (%)					
Former	2099 (36.80)	526 (48.28)	526 (36.76)	513 (35.90)	534 (36.15)
Current	720 (12.62)	181 (13.24)	173 (12.09)	183 (12.81)	183 (12.39)
Alcohol use, No. (%)[d]					
Moderate	1824 (31.98)	467 (34.16)	454 (31.73)	436 (30.51)	467 (31.62)
Heavy	304 (5.33)	86 (6.29)	71 (4.96)	66 (4.62)	81 (5.48)
Family history type 2 diabetes, No. (%)[e]	2097 (37.69)	436 (32.71)	525 (37.66)	550 (39.68)	586 (40.39)
Neighborhood Characteristics					
Healthy food stores, median (IQR)[f]	0.96 (1.91)	0.64 (1.59)	0.96 (1.91)	0.96 (2.23)	0.96 (2.23)
Recreational establishments, median (IQR)[f]	1.91 (2.87)	1.91 (2.55)	1.91 (2.87)	1.91 (2.87)	2.23 (2.87)
Neighborhood SES, mean (SD)[g]	-0.79 (4.45)	-0.57 (4.25)	-0.77 (4.37)	-0.90 (4.55)	-0.89 (4.62)

Abbreviations: SD, standard deviation; IQR, interquartile range; SES, socioeconomic status

[a] Genetic risk score composed of 55 SNPs

[b] Number of cases of incident diabetes that developed over time among those free of diabetes at baseline

[c] n=5699

[d] Alcohol use defined according to National Institute on Alcohol Abuse and Alcoholism definitions for men and women. Moderate drinking is defined as no more than 4 drinks on any single day and no more than 14 drinks per week for men, and no more than 3 drinks on any single day and no more than 7 drinks per week for women. Heavy drinking is defined as consumption in excess of moderate.

[e] n=5564

[f] Number of supermarkets/fruit and vegetable markets or commercial recreational establishments per square mile within a 1-mile buffer of the participant's residential address.

[g] The neighborhood socioeconomic status index includes census tract information on percent with a Bachelor's degree, percent with a high school degree, median household income, percent living in poverty, and percent receiving public assistance income. All variables recoded such that a higher value indicates higher socioeconomic status.

Table 3.2 Baseline characteristics of participants by categories of neighborhood exposures, Multi-Ethnic Study of Atherosclerosis

Sample Characteristics	Healthy Food Stores			Recreational Establishments			Neighborhood SES		
	<1	1-3	>3	<1	1-3	>3	Tertile 1	Tertile 2	Tertile 3
No. of participants	3380	1236	1088	1606	2258	1840	1885	1939	1880
Age, mean (SD)	61.55 (9.98)	62.69 (10.46)	62.28 (10.17)	61.52 (9.79)	61.81 (1.028)	62.45 (10.23)	61.82 (10.37)	61.64 (10.16)	62.36 (9.85)
Female, No. (%)	1719 (50.86)	652 (52.75)	591 (54.32)	781 (48.63)	1204 (53.32)	977 (53.10)	1029 (54.59)	1007 (51.93)	926 (49.26)
Race/ethnicity, No. (%)									
White	1589 (47.01)	439 (35.52)	290 (26.65)	693 (43.15)	920 (40.74)	705 (38.32)	275 (14.59)	863 (44.51)	1180 (62.77)
African American	885 (26.18)	213 (17.23)	303 (27.85)	540 (33.62)	471 (20.86)	390 (21.20)	241 (12.79)	212 (10.93)	248 (13.19)
Hispanic	591 (17.49)	231 (18.69)	462 (42.46)	248 (15.44)	517 (22.90)	519 (28.21)	620 (32.89)	513 (26.46)	268 (14.26)
Chinese	315 (9.32)	353 (28.56)	33 (3.03)	125 (7.78)	350 (15.50)	226 (12.28)	749 (39.73)	351 (18.10)	184 (9.79)
Prevalent type 2 diabetes, No. (%)	375 (11.09)	139 (11.25)	135 (12.41)	185 (11.52)	264 (11.69)	200 (10.87)	275 (14.59)	233 (12.02)	141 (7.50)
Incident type 2 diabetes, No. (%)	378 (12.58)	122 (11.12)	122 (12.80)	208 (14.64)	227 (11.38)	187 (11.40)	228 (14.16)	215 (12.60)	179 (10.29)
Household per capita income, mean (SD), per $10,000	5.22 (3.34)	4.97 (3.74)	4.34 (3.22)	5.32 (3.23)	4.70 (3.27)	5.08 (3.65)	3.39 (2.65)	4.68 (3.05)	6.91 (3.54)
Education, mean (SD), y	13.51 (3.63)	12.97 (4.49)	12.40 (4.42)	13.36 (3.64)	12.93 (4.12)	13.34 (4.16)	11.33 (4.50)	13.17 (3.57)	15.06 (2.89)
Genetic Risk Score, mean (SD)	61.49 (4.47)	62.03 (4.48)	62.06 (4.46)	61.56 (4.52)	61.71 (4.51)	61.86 (4.40)	61.95 (4.35)	61.51 (4.54)	61.69 (4.52)
Body mass index, No. (%)[a]									
Normal (18-<25)	925 (27.39)	463 (37.52)	298 (27.39)	393 (24.50)	694 (30.76)	599 (32.57)	507 (26.93)	514 (26.52)	665 (35.41)
Overweight (25-<30)	1244 (36.84)	480 (38.90)	424 (38.97)	606 (37.78)	830 (36.79)	712 (38.72)	686 (36.43)	715 (36.89)	747 (39.78)
Obese (≥30)	1208 (35.77)	291 (23.58)	366 (33.64)	605 (37.72)	732 (32.45)	528 (28.71)	690 (36.64)	709 (36.58)	466 (24.81)
Smoking status, No. (%)									
Former	1294 (38.28)	412 (33.33)	393 (36.12)	628 (39.10)	785 (34.77)	686 (37.28)	615 (32.63)	713 (36.77)	771 (41.01)
Current	435 (12.87)	126 (10.19)	159 (14.61)	193 (12.02)	304 (13.46)	223 (12.12)	295 (15.65)	260 (13.41)	165 (8.78)
Alcohol use, No. (%)									
Moderate	1128 (33.37)	375 (30.34)	321 (29.50)	493 (30.70)	718 (31.80)	613 (33.32)	473 (25.09)	580 (29.91)	771 (41.01)
Heavy	174 (5.15)	74 (5.99)	56 (5.15)	75 (4.67)	111 (4.92)	118 (6.41)	62 (3.29)	104 (5.36)	138 (7.34)
Family history type 2 diabetes, No. (%)[b]	1349 (40.99)	382 (31.60)	366 (34.40)	657 (42.25)	862 (38.99)	578 (32.15)	694 (37.76)	788 (41.89)	615 (33.33)

Abbreviations: SES, socioeconomic status; SD, standard deviation; No., number

[a] n=5699

[b] n=5564

Table 3.3 Associations of neighborhood exposure, genetic risk score, and neighborhood exposure by genetic risk score interaction with the risk for type 2 diabetes, Multi-Ethnic Study of Atherosclerosis, 2000-2012

	Model 1; HR (95% CI)[a]		Model 2; HR (95% CI)[b]			
Neighborhood Domain	**Neighborhood Exposure[c]**	**Genetic Risk Score[d]**	**Neighborhood Exposure[c]**	**Genetic Risk Score[d]**	**Interaction**	**P-value for Interaction**
Healthy Food Stores	0.97 (0.95, 0.99)	1.25 (1.16, 1.35)	0.97 (0.95, 0.99)	1.26 (1.16, 1.35)	0.98 (0.96, 1.00)	0.05
Recreational Establishments	0.97 (0.96, 0.98)	1.25 (1.16, 1.35)	0.97 (0.96, 0.99)	1.24 (1.15, 1.34)	0.99 (0.98, 1.01)	0.28
Neighborhood SES	0.94 (0.92, 0.95)	1.24 (1.15, 1.33)	0.94 (0.92, 0.95)	1.25 (1.16, 1.35)	1.01 (0.99, 1.03)	0.33

Abbreviations: HR, hazard ratio; CI, confidence interval; SES, socioeconomic status

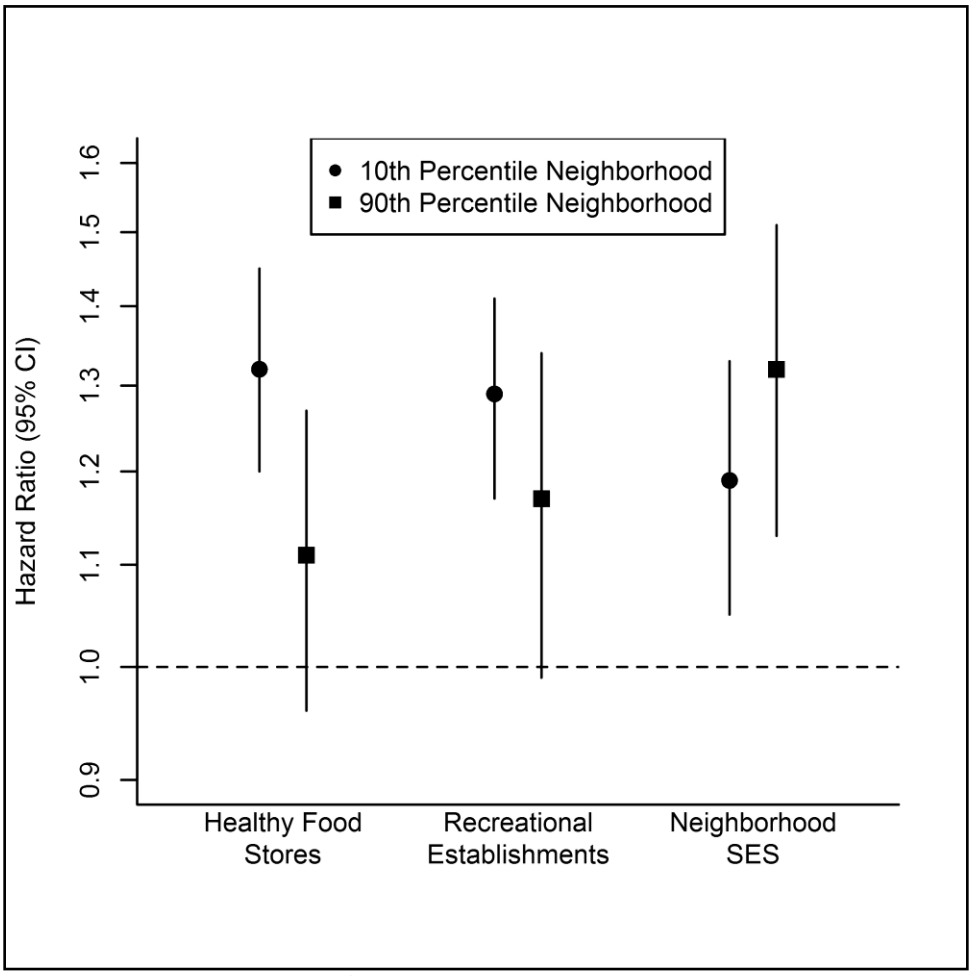
[a] Model 1 controls for gender, race, income, education, alcohol use, cigarette smoking, neighborhood SES (except for models with neighborhood SES as the exposure), and the first 5 eigenvectors from the pooled sample to control for population stratification. All standard errors and confidence intervals were computed using the delta method.

[b] Model 2 controls for all covariates in model 1, and adds an interaction between the genetic risk score and the neighborhood exposure.

[c] Estimates are for a 1-unit increase in neighborhood exposure

[d] Estimates are for a 10-allele increase in genetic risk score

Figure 3.1 Risk of type 2 diabetes associated with a 10-allele increase in the genetic risk score, estimated at the 10th and 90th percentiles of the neighborhood environment, Multi-Ethnic Study of Atherosclerosis, 2000-2012[a]



Abbreviations: SES, socioeconomic status; CI, confidence interval

[a] All models control for gender, race, income, education, alcohol use, cigarette smoking, neighborhood SES (except for models with neighborhood SES as the exposure), and the first 5 eigenvectors from the pooled sample to control for population stratification. All standard errors and confidence intervals were computed using the delta method.

Table 3.4 The independent and joint associations of neighborhood environment and genetic risk with the risk for developing type 2 diabetes, Multi-Ethnic Study of Atherosclerosis, 2000-2012[a]

	HR (95% CI)			
	Genetic Risk Score		Expected Interaction If…	
	Below Median	Above Median	Additive Joint Effects[b]	Multiplicative Joint Effects[c]
Healthy Food Stores[d]				
≥ 1 stores	1.00 (ref)	1.16 (0.96, 1.36)		
< 1 store	1.44 (1.18, 1.70)	1.92 (1.60, 2.25)	1.60 (1.20, 2.01)	1.67 (1.16, 2.19)
Recreational establishments[e]				
≥ 1 establishment	1.00 (ref)	1.19 (1.03, 1.34)		
< 1 establishment	1.66 (1.33, 1.99)	2.42 (1.99, 2.84)	1.85 (1.44, 2.25)	1.97 (1.43, 2.51)
Neighborhood SES[f]				
Above median	1.00 (ref)	1.37 (1.12, 1.61)		
Below median	1.47 (1.20, 1.74)	1.75 (1.43, 2.06)	1.83 (1.38, 2.29)	2.01 (1.36, 2.67)

Abbreviations: SES, socioeconomic status; HR, hazard ratio; CI, confidence interval; RERI, relative excess risk due to interaction

[a] Models control for gender, race, income, education, alcohol use, cigarette smoking, neighborhood SES (except for models with neighborhood SES as the exposure), and the first 5 eigenvectors from the pooled sample to control for population stratification. All confidence intervals were computed using the delta method.

[b] Expected joint effect for additive interaction: HR01 + HR10 -1 = HR11

[c] Expected joint effect for multiplicative interaction: HR01 * HR10 = HR11

[d] RERI for healthy food stores = 1.92 – 1.16 – 1.44 + 1 = 0.32 (0.09, 0.55); p-values for additive and multiplicative interaction, <0.01 and 0.23, respectively.

[e] RERI for recreational establishments = 2.42 – 1.66 – 1.19 + 1 = 0.57 (0.32, 0.82); p-values for additive and multiplicative interaction, <0.0001 and 0.12, respectively.

[f] RERI for neighborhood SES = 1.75 – 1.47 – 1.37 +1 = -0.09 (-0.40, 0.22); p-values for additive and multiplicative interaction, 0.61 and 0.24, respectively.

Appendix

Text 3.A1: Further description of the neighborhood GIS measures for healthy food store and recreational establishment availability

GIS-based measures of access to food stores were created using data obtained from the National Establishment Time Series (NETS) database from Walls and Associates for the years 2000-2012. This data includes time-series data on establishments derived from Dun and Bradstreet (D&B) archival establishment data. Addresses were geocoded using TeleAtlas EZ-Locate web-based geocoding software (TeleAtlas, 2011). We used Standard Industrial Classification (SIC) codes to identify supermarkets and grocery stores (#5411), and fruit and vegetable markets (#5431), which we classified as healthy food stores.[1] Additional supermarket data was obtained from Nielsen/TDLinx to enhance the supermarket list.[2] We identified supermarkets as grocery stores with at least $2 million in annual sales or at least 25 employees. Additionally, we included supermarkets that had a standard chain name based on a list derived from the Nielsen/TDLinx data as described in detail elsewhere.[3] For physical activity resources, 114 SIC codes were selected to represent establishments with indoor conditioning, dance, bowling, golf, team and racquet sports, and water activities derived from lists used in previous studies.[4,5] Simple densities per square mile were created for 1-mile buffers around each address using the point density command in ArcGIS 9.3.

Table 3.A1 Characteristics of established diabetes risk alleles used in the analyses[a]

SNP	Chr	Build 36 Position (bp)	Risk Allele	Other Allele	RAF White	RAF African American	RAF Hispanic	RAF Chinese	Nearest Gene
rs10203174	2	43543534	C	T	0.89	0.66	0.89	0.99	THADA
rs10401969*	19	19268718	C	T	0.07	0.17	0.08	0.09	CLIP2
rs10758593	9	4282083	A	G	0.44	0.50	0.48	0.44	GLIS3
rs10811661	9	22124094	T	C	0.83	0.93	0.87	0.60	CDKN2A/B
rs10830963	11	92348358	G	C	0.27	0.08	0.20	0.45	MTNR1B
rs10842994	12	27856417	C	T	0.80	0.94	0.84	0.79	KLHDC5
rs10923931	1	120319482	T	G	0.10	0.30	0.13	0.03	NOTCH2
rs1111875*	10	94452862	C	T	0.60	0.76	0.65	0.27	HHEX/IDE
rs11634397	15	78219277	G	A	0.64	0.43	0.57	0.09	ZFAND6
rs11717195	3	124565088	T	C	0.77	0.88	0.74	1.00	ADCY5
rs12242953	10	70535348	G	A	0.94	0.93	0.93	0.87	VPS26A
rs12497268	3	64065403	G	C	0.83	0.90	0.88	0.70	PSMD6
rs12571751*	10	80612637	A	G	0.55	0.53	0.51	0.57	ZMIZ1
rs12899811	15	89345080	G	A	0.31	0.62	0.61	0.97	PRC1
rs12970134*	18	56035730	A	A	0.26	0.15	0.18	0.17	MC4R
rs13233731	7	130088229	G	A	0.53	0.72	0.62	0.72	KLF14
rs13389219	2	165237122	C	T	0.61	0.30	0.68	0.90	GRB14
rs1359790*	13	79615157	G	A	0.72	0.88	0.68	0.72	SPRY2
rs1496653	3	23429794	A	G	0.82	0.64	0.84	0.80	UBE2E2
rs1552224	11	72110746	A	C	0.86	0.96	0.93	0.92	ARAP1(CENTD2)
rs163184*	11	2803645	G	T	0.49	0.20	0.42	0.46	KCNQ1
rs16927668	9	8359533	T	C	0.22	0.71	0.53	0.49	PTPRD
rs17168486	7	14864807	T	C	0.18	0.12	0.35	0.47	DGKB
rs17301514	3	188096103	A	G	0.12	0.06	0.05	0.04	ST64GAL1
rs17791513	9	81095410	A	G	0.95	0.96	0.88	0.95	TLE4
rs17867832	7	126784073	T	G	0.92	0.87	0.95	0.93	GCC1
rs1801282	3	12368125	C	G	0.88	0.98	0.91	0.96	PPARG
rs2075423	1	212221342	G	T	0.63	0.63	0.68	0.83	PROX1
rs2261181	12	64498585	T	C	0.12	0.21	0.10	0.11	HMGA2
rs2334499	11	1653425	T	C	0.40	0.16	0.44	0.82	DUSP8
rs243088*	2	60422249	T	A	0.47	0.52	0.58	0.68	BCL11A
rs2796441*	9	83498768	G	A	0.61	0.82	0.58	0.39	TLE1
rs2943640	2	226801829	C	A	0.66	0.90	0.81	0.94	IRS1
rs3802177*	8	118254206	G	A	0.70	0.90	0.76	0.54	SLC30A8
rs4299828	6	38285645	A	G	0.81	0.75	0.79	0.92	ZFAND3
rs4402960	3	186994381	T	G	0.34	0.52	0.28	0.25	IGF2BP2
rs4458523	4	6340887	G	T	0.60	0.57	0.68	0.92	WFS1
rs4502156	15	60170447	T	C	0.59	0.28	0.41	0.52	C2CD4A
rs459193*	5	55842508	G	A	0.75	0.59	0.72	0.50	ANKRD55
rs4812829	20	42422681	A	G	0.18	0.10	0.39	0.43	HNF4A

SNP	Chr	Build 36 Position (bp)	Risk Allele	Other Allele	RAF White	RAF African American	RAF Hispanic	RAF Chinese	Nearest Gene
rs516946	8	41638405	C	T	0.75	0.77	0.79	0.86	ANK1
rs5215*	11	17365206	C	T	0.36	0.11	0.33	0.36	KCNJ11
rs6795735	3	64680405	C	T	0.53	0.20	0.29	0.27	ADAMTS9
rs6819243	4	1283245	T	C	0.97	0.66	0.73	0.55	MAEA
rs7177055	15	75619817	A	G	0.71	0.36	0.60	0.33	HMG20A
rs7202877*	16	73804746	T	G	0.90	0.82	0.90	0.81	BCAR1
rs7569522	2	161054693	A	G	0.46	0.44	0.53	0.29	RBMS1
rs7756992	6	20787688	G	A	0.29	0.54	0.35	0.49	CDKAL1
rs780094	2	27594741	C	T	0.57	0.82	0.67	0.53	GCKR
rs7845219	8	96006678	T	C	0.50	0.65	0.44	0.26	TP53INP1
rs7903146*	10	114748339	T	C	0.70	0.70	0.74	0.98	TCF7L2
rs7955901	12	69719560	C	T	0.47	0.26	0.51	0.65	TSPAN8
rs8182584	19	38601550	T	G	0.40	0.41	0.41	0.65	PEPD
rs849135	7	28162938	G	A	0.49	0.73	0.64	0.99	JAZF1
rs9936385	16	52376670	C	T	0.41	0.49	0.32	0.14	FTO
rs2028299**	15	88175261	C	A	0.28	0.29	0.21	0.20	AP3S2
rs7041847**	9	4277466	A	G	0.53	0.87	0.60	0.48	GLIS3
rs7593730**	2	160879700	C	T	0.77	0.62	0.81	0.82	RBMS1

Abbreviations: SNP, single nucleotide polymorphism; bp, base pair; RAF, risk allele frequency
[a] All SNPs were taken from GWAS meta-analyses, please see citations [6-8] for more details.
* SNPs included in the restricted genetic risk score (corresponding to SNPs with consistent direction of effects in all racial/ethnic groups) and in the full genetic risk score
** SNPs included in the restricted genetic risk score only, and not the full genetic risk score

Table 3.A2 Genetic risk score performance across racial/ethnic groups[a]

Genetic risk score	Association with diabetes; HR (95% CI)	P-value for genetic risk score*race/ethnicity interaction in model pooling across race/ethnicity
Full genetic risk score (55 SNPs)		
White	1.05 (1.02, 1.07)	
African American	1.02 (1.00, 1.04)	
Hispanic	1.04 (1.02, 1.07)	
Chinese	1.05 (1.00, 1.09)	
Overall/Pooled	1.04 (1.02, 1.05)	0.35
Restricted genetic risk score (16 SNPs)		
White	1.04 (1.00, 1.09)	
African American	1.05 (1.01, 1.10)	
Hispanic	1.02 (0.98, 1.06)	
Chinese	1.03 (0.97, 1.09)	
Overall/Pooled	1.04 (1.01, 1.06)	0.78

Abbreviations: HR, hazard ratio; CI, confidence interval; SNP, single nucleotide polymorphism

[a] Estimates of association are from a parametric, interval censored survival model with a Weibull distribution, as described in the methods section of the paper. Race-specific models control for sex, and first 5 ethnic-specific eigenvectors. Estimates of association are for a 1-unit change in the genetic risk scores. Standard errors and confidence intervals were computed using the delta method. P-values correspond to a Type 3 Wald χ^2 test with 3 degrees of freedom. Analyses pooling across race/ethnic groups control for sex, and first 5 eigenvectors from the pooled sample.

Table 3.A3 Additional controls for possible neighborhood-level confounders[a]

	Control for population density; HR (95% CI)				Control for MESA site; HR (95% CI)			
Neighborhood Domain	**Neighborhood Exposure[b]**	**Genetic Risk Score[c]**	**Interaction**	**P value for interaction**	**Neighborhood Exposure[b]**	**Genetic Risk Score[c]**	**Interaction**	**P value for interaction**
Healthy Food Stores	0.99 (0.94, 1.04)	1.26 (1.16, 1.35)	0.98 (0.96, 1.00)	0.05	1.00 (0.97, 1.04)	1.26 (1.16, 1.35)	0.98 (0.96, 1.00)	0.05
Recreational Establishments	0.97 (0.96, 0.99)	1.24 (1.15, 1.34)	0.99 (0.98, 1.01)	0.28	0.98 (0.97, 1.00)	1.25 (1.15, 1.34)	0.99 (0.98, 1.01)	0.26
Neighborhood SES	0.93 (0.91, 0.95)	1.26 (1.16, 1.36)	1.01 (0.99, 1.03)	0.29	0.93 (0.91, 0.96)	1.26 (1.16, 1.36)	1.01 (0.99, 1.03)	0.28

Abbreviations: HR, hazard ratio; CI, confidence interval; SES, socioeconomic status

[a] Models control for gender, race, income, education, alcohol use, cigarette smoking, neighborhood SES index (except for models with neighborhood SES index as the exposure), and the first 5 eigenvectors from the pooled sample to control for population stratification. All confidence intervals were computed using the delta method. Population density, measured as persons per square mile within a 1-mile buffer of the participant's address, was calculated based on block-level census population. Each block was weighted by the percent of the block area that falls within the participant buffer. The total population within that block was then multiplied by this weight and the weighted populations were summed together for the total population within the buffer. The total population was divided by total buffer area in square miles. For dates prior to January 2006, population counts originated from the 2000 Census (Census, 2000). For dates on and after January 2006, population counts originated from the 2010 Census.

[b] Estimates are for a 1-unit increase in neighborhood exposure

[c] Estimates are for a 10-allele increase in genetic risk score

Table 3.A4 Interval censored regression models using incident cases only

	Model 1; HR (95% CI)[a]		Model 2; HR (95% CI)[b]			
Neighborhood Domain	**Neighborhood Exposure[c]**	**Genetic Risk Score[d]**	**Neighborhood Exposure[c]**	**Genetic Risk Score[d]**	**Interaction**	**P value for interaction**
Healthy Food Stores	0.97 (0.94, 1.00)	1.11 (0.99, 1.24)	0.97 (0.95, 1.00)	1.12 (1.00, 1.24)	0.98 (0.95, 1.02)	0.30
Recreational Establishments	0.97 (0.96, 0.99)	1.12 (1.00, 1.24)	0.97 (0.96, 0.99)	1.12 (0.99, 1.24)	1.00 (0.98, 1.02)	0.70
Neighborhood SES	0.94 (0.92, 0.96)	1.10 (0.98, 1.22)	0.94 (0.91, 0.96)	1.11 (0.99, 1.23)	1.01 (0.98, 1.04)	0.39

Abbreviations: HR, hazard ratio; CI, confidence interval; SES, socioeconomic status

[a] Model 1 controls for gender, race, income, education, alcohol use, cigarette smoking, neighborhood SES index (except for models with neighborhood SES index as the exposure), and the first 5 eigenvectors from the pooled sample to control for population stratification. All standard errors and confidence intervals were computed using the delta method.

[b] Model 2 controls for all covariates in model 1, and adds an interaction between the genetic risk score and the neighborhood exposure.

[c] Estimates are for a 1-unit increase in neighborhood exposure

[d] Estimates are for a 10-allele increase in genetic risk score

Table 3.A5 Associations using restricted genetic risk score, including only SNPs with consistent direction of effect in all racial/ethnic groups

	Model 1; HR (95% CI)[a]		Model 2; HR (95% CI)[b]			
Neighborhood Domain	**Neighborhood Exposure[c]**	**Restricted Genetic Risk Score[d]**	**Neighborhood Exposure[c]**	**Restricted Genetic Risk Score[d]**	**Interaction**	**P value for interaction**
Healthy Food Stores	0.97 (0.95, 0.99)	1.13 (1.05, 1.21)	0.97 (0.95, 0.99)	1.13 (1.05, 1.21)	0.98 (0.96, 1.00)	0.13
Recreational Establishments	0.94 (0.91, 0.97)	1.12 (1.04, 1.20)	0.94 (0.91, 0.97)	1.12 (1.04, 1.20)	1.00 (0.96, 1.04)	0.89
Neighborhood SES	0.94 (0.92, 0.95)	1.12 (1.04, 1.20)	0.94 (0.92, 0.95)	1.13 (1.04, 1.21)	1.01 (0.99, 1.02)	0.38

Abbreviations: SNP, single nucleotide polymorphisms; HR, hazard ratio; CI, confidence interval; SES, socioeconomic status

[a] Model 1 controls for gender, race, income, education, alcohol use, cigarette smoking, neighborhood SES index (except for models with neighborhood SES index as the exposure), and the first 5 eigenvectors from the pooled sample to control for population stratification. All standard errors and confidence intervals were computed using the delta method.

[b] Model 2 controls for all covariates in model 1, and adds an interaction between the genetic risk score and the neighborhood exposure.

[c] Estimates are for a 1-unit increase in neighborhood exposure

[d] Estimates are for a 4-allele increase in genetic risk score

Table 3.A6 Race-specific associations of neighborhood exposures, genetic risk score, and neighborhood exposures by genetic risk score interactions with type 2 diabetes[a]

	HR (95% CI)		
Neighborhood Domain	**Neighborhood Exposure[b]**	**Genetic Risk Score[c]**	**Interaction**
Healthy Food Stores			
White	0.99 (0.94, 1.04)	1.33 (1.12, 1.54)	0.98 (0.92, 1.03)
African American	0.98 (0.96, 1.01)	1.11 (0.95, 1.26)	0.99 (0.95, 1.03)
Hispanic	0.94 (0.90, 0.97)	1.31 (1.12, 1.50)	0.97 (0.94, 1.01)
Chinese	0.84 (0.66, 1.02)	1.35 (0.88, 1.81)	0.97 (0.68, 1.26)
Recreational Establishments			
White	0.98 (0.96, 1.00)	1.36 (1.15, 1.56)	0.99 (0.97, 1.02)
African American	0.97 (0.95, 1.00)	1.09 (0.93, 1.25)	0.99 (0.96, 1.02)
Hispanic	0.97 (0.94, 0.99)	1.25 (1.08, 1.42)	0.98 (0.96, 1.01)
Chinese	0.91 (0.82, 1.01)	1.31 (0.85, 1.78)	0.98 (0.84, 1.12)
Neighborhood SES			
White	0.95 (0.91, 0.98)	1.39 (1.16, 1.62)	0.98 (0.94, 1.03)
African American	0.96 (0.93, 0.99)	1.15 (0.97, 1.32)	1.02 (0.99, 1.06)
Hispanic	0.91 (0.88, 0.94)	1.36 (1.08, 1.63)	1.02 (0.98, 1.06)
Chinese	0.91 (0.86, 0.95)	1.40 (1.06, 1.74)	1.00 (0.94, 1.06)

Abbreviations: HR, hazard ratio; CI, confidence interval

[a] Models control for gender, income, education, alcohol use, cigarette smoking, neighborhood SES index (except for models with neighborhood SES index as the exposure), and the first 5 ethnic-specific eigenvectors. All standard errors and confidence intervals were computed using the delta method.

[b] Estimates are for a 1-unit increase in neighborhood exposure

[c] Estimates are for a 4-allele increase in genetic risk score

Appendix References:

1. Genome-wide trans-ancestry meta-analysis provides insight into the genetic architecture of type 2 diabetes susceptibility. *Nat. Genet.* Mar 2014;46(3):234-244.

2. Auchincloss AH, Moore KA, Moore LV, Diez Roux AV. Improving retrospective characterization of the food environment for a large region in the United States during a historic time period. *Health Place.* Nov 2012;18(6):1341-1347.

3. Gordon-Larsen P, Nelson MC, Page P, Popkin BM. Inequality in the built environment underlies key health disparities in physical activity and obesity. *Pediatrics.* Feb 2006;117(2):417-424.

4. Ng MC, Shriner D, Chen BH, et al. Meta-analysis of genome-wide association studies in African Americans provides insights into the genetic architecture of type 2 diabetes. *PLoS genetics.* Aug 2014;10(8):e1004517.

5. Nielsen Company. *Retail Site Database - The Ultimate Source*: Trade Dimensions, a subsidiary of Nielsen Company; 2008.

6. Powell LM, Chaloupka FJ, Slater SJ, Johnston LD, O'Malley PM. The availability of local-area commercial physical activity-related facilities and physical activity among adolescents. *Am. J. Prev. Med.* Oct 2007;33(4 Suppl):S292-300.

7. U.S. Department of Labor. SIC Manual. Available at: http://www.osha.gov/pls/imis/sic_manual.html Accessed May 2012.

CHAPTER 4 :

INCLUSION OF INDIVIDUAL AND AREA-LEVEL SOCIOECONOMIC STATUS IN RISK PREDICTION MODELLING: AN APPLICATION TO TYPE 2 DIABETES IN THE MULTI-ETHNIC STUDY OF ATHEROSCLEROSIS (MESA)

Introduction:

Diabetes is an important cause of death and disability worldwide.[1] In the United States, an estimated 18.8 million adults have diagnosed diabetes mellitus,[2] and if current trends persist, as many as 1 in 3 Americans could have diabetes by 2050.[3] The continued increase in prevalence and incidence has led to calls for cost-effective prevention strategies, including identifying individuals at high risk for developing disease.[4,5]

In light of evidence that nearly two-thirds of diabetes cases in high risk individuals can be prevented through behavioral and pharmacologic interventions, risk prediction models are increasingly recommended for use in clinical practice and public health planning.[4-6] These models, which use available clinical information, are seen as complements to traditional approaches for identifying high-risk individuals, such as hemoglobin A_{1C} tests, and could help identify high-risk individuals before lab tests can identify prediabetes.[6]

With the growth of risk prediction as a tool to guide preventive and therapeutic interventions in the US and abroad,[5] much research has focused upon improving prediction models. For type 2 diabetes, this has often taken the form of including novel

biological information, especially genetic risk.[7-11] Virtually no research, however, has explored the use of social and environmental information – including individual and area-level socioeconomic status (SES) – in diabetes risk prediction models. This is surprising given the pronounced disparities in type 2 diabetes by SES,[12] and the increasing recognition of the importance of social and area-level factors in the development of diabetes.[13,14] Studies of cardiovascular risk models have shown that ignoring SES can result in underestimation of risk for low-SES individuals, yet no similar studies exist for type 2 diabetes.[15,16] Given the growth of electronic health records (EHRs) and the ability to link "non-medical" environmental data to medical records,[17] empirical assessments of the utility of including multi-level social information in risk assessment are warranted.

Using a large, multi-ethnic, prospective cohort, we used several approaches to evaluate the utility of adding individual and area-level SES information to diabetes risk prediction models. Specifically, we investigated (1) whether adding individual and area-level SES to models based on traditional risk factors aids in the discrimination of people who developed diabetes from those who remained diabetes-free; (2) if diabetes risk prediction models based upon traditional risk factors underestimate risk in low-SES individuals or those living in low-SES environments, and whether the addition of SES information improves prediction accuracy (i.e. model calibration) across the SES distribution; and (3)whether adding social information to risk prediction models results in risk reclassification.

Methods:

Study population and analytic sample:

The Multi-Ethnic Study of Atherosclerosis (MESA) is a prospective cohort study of non-institutionalized adults 45-84 years old who self-identified as white, black, Hispanic, or Chinese. Participants were free of clinical cardiovascular disease at baseline and were recruited from 6 US locations (New York, New York; Baltimore, Maryland; Forsyth County, North Carolina; Chicago, Illinois; St. Paul, Minnesota; and Los Angeles, California). The first examination took place between 2000 and 2002, and 4 follow-up exams were completed an average of 1.6, 3.1, 4.8, and 9.5 years after baseline. Written informed consent was obtained from participants, and the study was approved by the institutional review boards at each site.

A total of 6814 adults were enrolled at baseline. For this analysis, we excluded individuals with type 2 diabetes at baseline (n=736). We also excluded individuals who did not agree to have their residential address geocoded and linked to area-level data (n=623), who were missing data on diabetes risk factors (n=404), or for whom no follow-up information was available regarding diabetes status (n=29). Our total analytic sample thus consisted of 5021 adults.

Follow-up and measurement of type 2 diabetes

Incident type 2 diabetes was determined at each follow-up exam according to the American Diabetes Association's 2003 criteria[18]: fasting plasma glucose level ≥126 mg/dL (7 mmol/L) or use of oral anti-hyperglycemic medications or insulin. Information on the use of oral medication and insulin was obtained at each exam via self-report and visual inspection of medication bottles.

Traditional diabetes risk factors

Traditional diabetes risk factors of interest were identified based upon established diabetes risk prediction models.[5,7,19] Non-laboratory risk factors measured at baseline included age, sex, family history of type 2 diabetes, systolic blood pressure (mmHg), receiving treatment for hypertension (yes/no), waist circumference (cm), waist:hip ratio, body mass index (BMI, kg/m^2), height (m), and smoking status (yes/no). Self-identified race/ethnicity, utilized in many diabetes risk scores, was also available at baseline. For laboratory-based risk factors, we utilized 12-hour fasting blood samples to measure plasma glucose levels (mg/dL), triglycerides (mmol/L), and HDL cholesterol (mmol/L).

Socioeconomic status

An extensive literature exists regarding the measurement of SES at different points during the lifecourse.[20,21] Since we were interested in exploring variables that could be feasibly incorporated into medical records or public health databases, we elected to focus on simple SES measures, and to combine measures only in ways that could be reasonably replicated with other data. At the individual level, we measured SES using highest level of attained education in years, annual household income, and annual household income per individual supported by the income in the household (hereafter, called household income per capita), all of which were assessed via questionnaire at baseline. Because individual-level SES variables are not always available and area-level features may influence chronic disease development independent of individual-level factors, [22] we also evaluated several area-level SES variables. Area-level variables were defined at the census tract-level and included median household income, percentage of adults ages 25 and older with a bachelor's degree or higher, and percentage of people living below the federally defined poverty threshold. In the MESA sample and in other

studies, these variables have been shown to be related to other area-level exposures that may be relevant to the development of type 2 diabetes, such as the availability of healthy food and physical activity resources,[23-25] perceived safety,[26,27] and levels of air pollution.[28,29] All area-level predictors were taken from the U.S. Census 2000 and linked to geocoded participant residential addresses.

Since single measures of SES may fail to capture the intersections between different domains of social standing and risk, we created SES indices at both the individual- and area-levels, as well as an overall SES index combining individual and area measures. Following recommended methods, indices were created by first transforming variables to remove skewness, z-scoring the variables, and then summing the z-scores to create a composite index.[30] We selected variables *a priori* to include in SES indices based upon prior work in our cohort and others demonstrating clusters of variables that are associated with increased risk for developing diabetes and cardiovascular disease.[19,31,32] At the individual-level, we included education and annual household income. At the area-level, we included percent of adults with a high school education, percent of adults with a bachelor's degree, median household income (log transformed), median home value (log transformed), percent of adults in a managerial occupation, and percent of households with income from interest, dividends, or rental properties, all measured at the census tract-level. For the combined SES index, we summed the standardized individual and area SES indices. All indices were created so that an increasing score indicates greater SES.

Statistical analysis

We began our analysis by evaluating the sociodemographic characteristics and distribution of diabetes risk factors in our sample overall and by incident diabetes status. Sample means and standard deviations were used to summarize normally distributed continuous variables, while sample medians and interquartile ranges were used for skewed variables.

We developed two separate prediction models to estimate 10-year incident diabetes risk in MESA, one employing only non-laboratory variables that would be available during a routine medical visit (hereafter, the "Clinical Model"), and one using laboratory variables from a fasting blood sample (hereafter, the "Laboratory Model"). The primary purpose of fitting our own internal prediction models was to ensure that any improvements in model performance with the addition of SES predictors was not due to simply improving the fit of poorly modeled variables in an external prediction model. We began by fitting Cox proportional hazards models with all risk factors from the ARIC diabetes risk model,[33] which has been shown to perform well in MESA and other validation studies.[19,34] This included age, family history of diabetes, race/ethnicity, systolic blood pressure, height, waist circumference, fasting glucose, fasting triglycerides, HDL cholesterol. For parsimony, we excluded variables that were not marginally associated with diabetes incidence ($p \geq 0.1$). We then added additional risk factors shown to be predictive in at least 3 other diabetes risk scores (BMI, waist:hip ratio, smoking status)[5,35] and retained those that were marginally significant ($p<0.1$) in likelihood ratio tests. Akaike Information Criterion (AIC) values were used to decide between predictors that were highly correlated (e.g. BMI and waist circumference). Individuals were considered at risk until diagnosis of diabetes, last follow-up visit, or administrative

censoring at exam 5, and incident diabetes cases were assigned to the midpoint between their previous diabetes-free and current exam dates.

For continuous predictors, we checked for linearity of the exposure-outcome relationship by fitting penalized b splines with 2 degrees of freedom and retained splines for continuous predictors with visual and statistical evidence of non-linearity. [36] We compared the sensitivity of our approach to other methods including adding square terms and fitting multivariable fractional polynomials.[37] For all models, the proportional hazards assumption was investigated graphically using scaled Shoenfeld residuals and log-time since baseline, and found to be satisfied for all variables. The final Clinical Model included age (spline), race/ethnicity, family history of type 2 diabetes, systolic blood pressure, waist circumference, and hypertension treatment, and our Laboratory Model added fasting glucose and triglycerides (log transformed, spline).

We added SES predictors individually in separate models to both Clinical and Laboratory models. We assessed the value of each addition by evaluating model discrimination, calibration, and risk reclassification in models with and without the SES predictors. Model discrimination refers to the ability of a model to differentiate who will and will not have an event: in this case, incident diabetes.[38] We evaluated discrimination by assessing the statistical significance of the predictors using likelihood ratio tests of nested models, computing Harrell's C statistics (the equivalent to area under the curve for survival models), and plotting receiver operating characteristic (ROC) curves.[39]

For model calibration, we were particularly interested in whether the Clinical and Laboratory models underestimated risk for individuals of lower individual-SES or residing in low-SES neighborhoods. Following recommended methods, we calculated

observed risks using Kaplan Meier estimates and predicted risks using our Cox proportional hazards models.[37,40] To evaluate differential calibration across the SES distribution, we then calculated mean differences (and 95% confidence intervals) between observed and predicted risks across tertiles (or natural groupings) of the SES predictors. We elected to use tertiles because they matched the natural categories for variables like education, but we also calculated mean differences across quintiles of the SES predictors to evaluate the sensitivity of our results to the number of categories. For brevity, we assessed calibration for the two individual and area-level SES variables that were most predictive of incident diabetes across the Clinical and Laboratory models.

To place model discrimination in the context of potential preventive interventions, we also assessed risk reclassification comparing models with and without SES predictors. Given that no established risk threshold exists for instituting preventive pharmacotherapy or particular preventive interventions, we chose *a priori* 10-year risk categories of 0 to <10%, 10-20%, and >20%. These categories are similar to those used in risk prediction models for cardiovascular disease. Using these categories, we calculated the number of individuals classified in each category comparing models with and without the two most predictive individual- and area-level SES variables.

We performed several sensitivity analyses. Because individual- and area-level variables may be independently associated with type 2 diabetes,[41,42] we evaluated models including both individual- and area-level SES variables together. Similarly, individual and area-level SES predictors may interact with each other or with traditional risk factors to influence diabetes risk. To evaluate this possibility, we tested interactions between each SES predictor with other SES predictors and with age, sex, and race to see if such

interactions improved model discrimination or calibration. To assess the sensitivity of our calibration results to the use of our internal prediction model, we also applied two established diabetes risk prediction models to our sample and assessed their calibration using the same methods described above. Based upon their performance in prior validation studies,[19,34] we chose to utilize the original ARIC diabetes risk score[33] and the Framingham Offspring Study diabetes risk score.[43] We calculated the predicted 10-year risk of diabetes for both of these risk scores using the published model coefficients, and included the risk score as a covariate in a Cox proportional hazards model to account for censoring in our sample. To each of these models, we added the same SES variables as above, both independently and interacted with the risk score, and evaluated changes in calibration with the addition of the SES variables. All analyses were conducted in R version 3.1.2 (R Foundation for Statistical Computing, Vienna, Austria), and STATA 12 (Stata Corp, College Station, Texas).

Results:

Of 5022 individuals without diabetes at baseline, 615 developed diabetes over a median of 9.2 years of follow-up. Compared to individuals who did not develop diabetes, individuals who developed diabetes were more likely to be African-American and Hispanic, to have a parent or sibling with diabetes, and to be obese (Table 1). Individuals developing diabetes also had higher mean baseline values of traditional diabetes risk factors, including fasting plasma glucose, triglycerides, waist circumference, and systolic blood pressure, relative to those who did not develop diabetes. 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 levels of education, lower median household incomes, and a greater proportion of persons living below the poverty line. SES indices at the individual- and area-level were considerably lower (indicating lower SES) for individuals who developed diabetes.

Table 2 displays the association of each SES predictor with diabetes, as well as model fit and model discrimination for multivariable Clinical and Laboratory prediction models. All SES variables were highly predictive of incident diabetes in univariable models (all $p<0.01$, data not shown), but the associations were attenuated considerably in the multivariable models. The strongest SES predictors in the Clinical model were the SES index at the individual level (HR for a standard deviation [SD] increase = 0.91; 95% CI [0.82, 1.00]) and percent of adults with a bachelor's degree at the area-level (HR for a SD increase = 0.91, 95% CI [0.83, 1.01]). In the Laboratory model, the strongest individual-level SES predictor was categorical household income (HR comparing highest to lowest category = 0.74; 95% CI [0.57, 0.95]), though continuous household income, and household income per capita were also relatively 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 the individual SES measures. Hazard ratios for all model predictors are listed in Tables 4.A1 and 2, and Figure 4.A1.

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 ROC plots were largely overlapping (Figure 4.A2).

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

Most SES predictors reclassified 2-3% of individuals in a given risk category (Tables 4.A3-6). 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 possible synergies between individual and area-level SES measures (Table 4.A7), sensitivity analyses adding the variables to the same models did not produce better discrimination or calibration. Analyses adding interactions of SES predictors with each other and with age, sex, and race, also failed to

yield improvements in discrimination and calibration, and no interactions were consistently predictive across Clinical and Laboratory models (data not shown).

Sensitivity analyses evaluating the calibration of two established risk scores across tertiles of select SES predictors varied according to the risk score used (Table 4.A8). Both scores generally overestimated risk for high-SES groups, but only the Framingham score significantly underestimated risk for low-SES groups (by as much as 2.60% in one case). The addition of SES predictors to the risk scores significantly 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:

The inclusion of socioeconomic information in risk prediction is of increasing interest, with recent applications in cardiovascular event prediction[11,12,43] and hospital readmission assessments.[44-46] Given that individual and area-level socioeconomic characteristics are strong predictors of a many chronic conditions and operate through a variety of pathways which are difficult to measure, socioeconomic characteristics have been hypothesized to aid in risk prediction. In our study, we find limited support for the use of individual or area-level socioeconomic characteristics in type 2 diabetes risk prediction to improve model discrimination. While several socioeconomic characteristics were indeed predictive of incident type 2 diabetes independent of demographic, anthropometric, and laboratory predictors, no SES variable altered the overall ability of the models to discriminate between those who would and would not go on to develop

diabetes. Adding SES predictors also failed to reclassify most individuals across selected risk categories.

With respect to model calibration, models without socioeconomic predictors 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 improved calibration across the SES distribution and eliminated significant differences between SES groups, though results varied in the established risk models. The magnitude of under- and overestimation varied according to the SES predictor and model, but generally never exceeded a couple of percentage points. Whether the differences between observed and predicted risks in the different SES groups are meaningful for clinical or public health applications is debatable, and depends upon the specific thresholds for potential interventions which are currently not well defined for type 2 diabetes.

While we are unaware of other studies directly investigating the benefit of adding socioeconomic variables to type 2 diabetes prediction models, our results are largely consistent with similar studies from the cardiovascular event prediction literature. In several studies evaluating the utility of adding socioeconomic variables to the Framingham Risk Score, researchers found that SES predictors offer little improvement in model discrimination.[16,47] However, the same studies also document systematic underestimation of risk in low-SES individuals that is eliminated when SES information is added to the prediction model. The failure to include SES in this case could lead to under-treatment of low-SES individuals with therapies known to be effective for preventing cardiovascular events, such as statins.[15]

Why risk prediction models might perform differentially for individuals in different socioeconomic strata is not known and may merit further investigation. It is notable that most risk prediction models in the US, including those for type 2 diabetes, are created using large longitudinal cohorts.[19] To the extent that loss to follow up in these studies is higher in low-SES groups, and differential attrition by SES is not captured by variables in the risk prediction model, it is possible that prediction models based on these cohorts are better fit to higher SES individuals. Whether such differential attrition, which has been demonstrated in the literature,[48,49] results in poor calibration of prediction models across the distribution of SES is unknown. Given the increasing use of risk prediction models to help guide clinical and public health decision-making, future research that compares models which ignore SES with models which explicitly account for differential attrition (e.g. via survival models or inverse probability weighting) would help ensure that risk models perform well across the SES distribution.

Our study has several strengths. We utilized a large, diverse cohort with excellent measured data over 10 years on both traditional diabetes risk factors and socioeconomic variables. By using SES data at the individual and area-level, we were able to compare a multitude of SES measures, which may be useful when considering the inclusion of social variables in electronic health records. Employing SES predictors at both the individual and area-level also allowed us to assess their possible interactions when predicting diabetes risk. Finally, we performed several sensitivity analyses to ensure that our modeling approach and results were robust to different specifications and assumptions.

Our study also has several limitations. First, we use a single cohort of middle-aged and older adults. The absence of model improvement in our sample does not mean that SES predictors may not perform differently in other samples, such as those in less urban areas or composed of younger individuals. Second, we chose SES predictors *a priori* based upon evidence from the literature of their effects on chronic disease development.[21,50,51] Other SES predictors not included in our models (e.g. occupation), or SES predictors defined at different levels of geographic resolution (e.g. census block groups or zip codes), may yield different results. Third, we did not attempt to build a prediction model using more data-driven approaches (e.g. ensemble methods) which can obviate the need to specify whether and/or how to include predictor variables.[52,53] Whether utilizing such methods would result in the selection of social variables over more traditional predictors is unknown, and should be evaluated in further research. Finally, it should be noted that placing individuals into risk categories based, in part, upon social characteristics is not without potential harms, including activating the implicit biases of clinicians and care providers.[54,55] Ultimately, whether social variables like SES and race should be included in risk prediction models is not only a statistical question, but an ethical one that requires weighing the potential benefits and harms.

In conclusion, 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, particularly area education, to the models largely mitigated these differences, though the absolute difference in risk was small. While no SES predictor aided in discriminating between those who did and did not develop diabetes, whether such predictors should be included in risk prediction models to aid in

calibration remains an open question. Further work exploring different sets of SES predictors in additional populations is merited and would help guide efforts to assure that risk prediction models do not exacerbate social disparities in disease outcomes due to systematic misestimation of risk.

References:

1. Danaei G, Finucane MM, Lu Y, et al. National, regional, and global trends in fasting plasma glucose and diabetes prevalence since 1980: systematic analysis of health examination surveys and epidemiological studies with 370 country-years and 2.7 million participants. *Lancet.* Jul 2 2011;378(9785):31-40.

2. Ali MK, Bullard KM, Saaddine JB, Cowie CC, Imperatore G, Gregg EW. Achievement of goals in U.S. diabetes care, 1999-2010. *N. Engl. J. Med.* Apr 25 2013;368(17):1613-1624.

3. Boyle JP, Thompson TJ, Gregg EW, Barker LE, Williamson DF. Projection of the year 2050 burden of diabetes in the US adult population: dynamic modeling of incidence, mortality, and prediabetes prevalence. *Popul Health Metr.* 2010;8:29.

4. Seaquist ER. ADdressing the burden of diabetes. *JAMA.* 2014;311(22):2267-2268.

5. Noble D, Mathur R, Dent T, Meads C, Greenhalgh T. Risk models and scores for type 2 diabetes: systematic review. *BMJ.* 2011;343:d7163.

6. National Institute for Health and Care Excellence. Preventing type 2 diabetes: Risk identification and interventions for individauls at high risk. *NICE Public Health Guidance 38.* 2012; http://www.nice.org.uk/guidance/ph38. Accessed July 5, 2014.

7. Buijsse B, Simmons RK, Griffin SJ, Schulze MB. Risk assessment tools for identifying individuals at risk of developing type 2 diabetes. *Epidemiol. Rev.* Jul 2011;33(1):46-62.

8. Vassy JL, Durant NH, Kabagambe EK, et al. A genotype risk score predicts type 2 diabetes from young adulthood: the CARDIA study. *Diabetologia.* Oct 2012;55(10):2604-2612.

9. Vassy JL, Hivert MF, Porneala B, et al. Polygenic type 2 diabetes prediction at the limit of common variant detection. *Diabetes.* Feb 11 2014.

10. Meigs JB, Shrader P, Sullivan LM, et al. Genotype score in addition to common risk factors for prediction of type 2 diabetes. *N. Engl. J. Med.* Nov 20 2008;359(21):2208-2219.

11. 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.* Jan 2013;36(1):70-76.

12. Agardh E, Allebeck P, Hallqvist J, Moradi T, Sidorchuk A. Type 2 diabetes incidence and socio-economic position: a systematic review and meta-analysis. *Int. J. Epidemiol.* Jun 2011;40(3):804-818.

13. Auchincloss AH, Diez Roux AV, Mujahid MS, Shen M, Bertoni AG, Carnethon MR. Neighborhood resources for physical activity and healthy foods and incidence of type 2 diabetes mellitus: the Multi-Ethnic study of Atherosclerosis. *Arch. Intern. Med.* Oct 12 2009;169(18):1698-1704.

14. Ludwig J, Sanbonmatsu L, Gennetian L, et al. Neighborhoods, obesity, and diabetes--a randomized social experiment. *N. Engl. J. Med.* Oct 20 2011;365(16):1509-1519.

15. Fiscella K, Tancredi D. Socioeconomic status and coronary heart disease risk prediction. *JAMA.* Dec 10 2008;300(22):2666-2668.

16. Fiscella K, Tancredi D, Franks P. Adding socioeconomic status to Framingham scoring to reduce disparities in coronary risk assessment. *Am. Heart J.* Jun 2009;157(6):988-994.

17. Weber GM, Mandl KD, Kohane IS. Finding the Missing Link for Big Biomedical Data. *JAMA.* May 22 2014.

18. Genuth S, Alberti KG, Bennett P, et al. Follow-up report on the diagnosis of diabetes mellitus. *Diabetes Care.* Nov 2003;26(11):3160-3167.

19. 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.

20. Oakes JM, Kaufman JS. *Methods in Social Epidemiology*: Wiley; 2006.

21. Galobardes B, Lynch J, Smith GD. Measuring socioeconomic position in health research. *Br. Med. Bull.* January 1, 2007 2007;81-82(1):21-37.

22. Kawachi I, Berkman LF. *Neighborhoods and Health*: Oxford University Press, USA; 2003.

23. Moore LV, Diez Roux AV. Associations of neighborhood characteristics with the location and type of food stores. *Am. J. Public Health.* Feb 2006;96(2):325-331.

24. Duncan DT, Kawachi I, White K, Williams DR. The geography of recreational open space: influence of neighborhood racial composition and neighborhood poverty. *J. Urban Health.* Aug 2013;90(4):618-631.

25. Moore LV, Diez Roux AV, Evenson KR, McGinn AP, Brines SJ. Availability of recreational resources in minority and low socioeconomic status areas. *Am. J. Prev. Med.* Jan 2008;34(1):16-22.

26. Burdette HL, Wadden TA, Whitaker RC. Neighborhood safety, collective efficacy, and obesity in women with young children. *Obesity (Silver Spring, Md.).* Mar 2006;14(3):518-525.

27. 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.* Aug 2013;91:15-24.

28. 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.* Nov-Dec 2013;121(11-12):1325-1333.

29. Havard S, Deguen S, Zmirou-Navier D, Schillinger C, Bard D. Traffic-Related Air Pollution and Socioeconomic Status: A Spatial Autocorrelation Study to Assess Environmental Equity on a Small-Area Scale. *Epidemiology.* 2009;20(2):223-230 210.1097/EDE.1090b1013e31819464e31819461.

30. Gilthorpe MS. The importance of normalisation in the construction of deprivation indices. *J. Epidemiol. Community Health.* Dec 1995;49 Suppl 2:S45-50.

31. Diez Roux AV, Merkin SS, Arnett D, et al. Neighborhood of residence and incidence of coronary heart disease. *N. Engl. J. Med.* Jul 12 2001;345(2):99-106.

32. Krishnan S, Cozier YC, Rosenberg L, Palmer JR. Socioeconomic Status and Incidence of Type 2 Diabetes: Results From the Black Women's Health Study. *Am. J. Epidemiol.* March 1, 2010 2010;171(5):564-570.

33. Schmidt MI, Duncan BB, Bang H, et al. Identifying individuals at high risk for diabetes: The Atherosclerosis Risk in Communities study. *Diabetes Care.* Aug 2005;28(8):2013-2018.

34. Mann DM, Bertoni AG, Shimbo D, et al. Comparative Validity of 3 Diabetes Mellitus Risk Prediction Scoring Models in a Multiethnic US Cohort. *Am. J. Epidemiol.* May 1 2010;171(9):980-988.

35. Collins GS, Mallett S, Omar O, Yu LM. Developing risk prediction models for type 2 diabetes: a systematic review of methodology and reporting. *BMC medicine.* 2011;9:103.

36. Peng RD, Dominici F. *Statistical Methods for Environmental Epidemiology with R: A Case Study in Air Pollution and Health*: Springer; 2008.

37. Royston P, Moons KG, Altman DG, Vergouwe Y. Prognosis and prognostic research: Developing a prognostic model. *BMJ.* 2009;338:b604.

38. Ioannidis JP, Tzoulaki I. What makes a good predictor?: the evidence applied to coronary artery calcium score. *JAMA.* Apr 28 2010;303(16):1646-1647.

39. Harrell FE, Jr., Lee KL, Califf RM, Pryor DB, Rosati RA. Regression modelling strategies for improved prognostic prediction. *Stat. Med.* Apr-Jun 1984;3(2):143-152.

40. Cook NR, Ridker PM. Advances in Measuring the Effect of Individual Predictors of Cardiovascular Risk: The Role of Reclassification Measures. *Ann. Intern. Med.* Jun 2 2009;150(11):795-W143.

41. Diez Roux AV. Estimating neighborhood health effects: the challenges of causal inference in a complex world. *Soc. Sci. Med.* May 2004;58(10):1953-1960.

42. Diez Roux AV, Mair C. Neighborhoods and health. *Ann. N. Y. Acad. Sci.* Feb 2010;1186:125-145.

43. 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.* May 28 2007;167(10):1068-1074.

44. Krumholz HM, Bernheim SM. Considering the Role of Socioeconomic Status in Hospital Outcomes MeasuresConsidering the Role of Socioeconomic Status in Hospital Outcomes Measures. *Ann. Intern. Med.* 2014;161(11):833-834.

45. Fiscella K, Burstin HR, Nerenz DR. Quality measures and sociodemographic risk factors: to adjust or not to adjust. *JAMA.* Dec 24-31 2014;312(24):2615-2616.

46. Nagasako EM, Reidhead M, Waterman B, Claiborne Dunagan W. Adding Socioeconomic Data To Hospital Readmissions Calculations May Produce More Useful Results. *Health Aff. (Millwood).* May 1, 2014 2014;33(5):786-791.

47. Brindle PM, McConnachie A, Upton MN, Hart CL, Smith GD, Watt GCM. The accuracy of the Framingham risk-score in different socioeconomic groups: a prospective study. *The British Journal of General Practice.* 2005;55(520):838-845.

48. Howe LD, Tilling K, Galobardes B, Lawlor DA. Loss to follow-up in cohort studies: bias in estimates of socioeconomic inequalities. *Epidemiology.* Jan 2013;24(1):1-9.

49. Fitzgerald J, Gottschalk P, Moffitt R. An Analysis of Sample Attrition in Panel Data: The Michigan Panel Study of Income Dynamics. *The Journal of Human Resources.* 1998;33(2):251-299.

50. Krieger N, Chen JT, Waterman PD, Rehkopf DH, Subramanian SV. Painting a truer picture of US socioeconomic and racial/ethnic health inequalities: the Public Health Disparities Geocoding Project. *Am. J. Public Health.* Feb 2005;95(2):312-323.

51. Krieger N, Chen JT, Waterman PD, Rehkopf DH, Subramanian SV. Race/ethnicity, gender, and monitoring socioeconomic gradients in health: a comparison of area-based socioeconomic measures--the public health disparities geocoding project. *Am. J. Public Health.* Oct 2003;93(10):1655-1671.

52. Kennedy EH, Wiitala WL, Hayward RA, Sussman JB. Improved cardiovascular risk prediction using nonparametric regression and electronic health record data. *Med. Care.* Mar 2013;51(3):251-258.

53. Hastie T, Tibshirani R, Friedman JH. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*: Springer; 2001.

54. Gómez LE, López N, Valdez RB, et al. *Mapping "Race": Critical Approaches to Health Disparities Research*: Rutgers University Press; 2013.

55. van Ryn M, Burke J. The effect of patient race and socio-economic status on physicians' perceptions of patients. *Soc. Sci. Med.* 2000;50(6):813-828.

Table 4.1 Baseline sociodemographic characteristics and diabetes risk factors, Multi-Ethnic Study of Atherosclerosis, 2000

Characteristics	Overall	Incident Cases	Non-Cases
Sample, No. of participants	5021	615	4406
Age, mean (SD), y	61.31 (10.16)	60.61 (9.48)	61.41 (10.25)
Female, No. (%)	2664 (53.06)	326 (53.01)	2338 (53.06)
Race/Ethnicity, No. (%)			
White	2153 (42.88)	191 (31.06)	1962 (44.53)
African-American	1217 (24.24)	180 (29.27)	1037 (23.54)
Hispanic	1046 (20.83)	168 (27.32)	878 (19.93)
Chinese-American	605 (12.05)	76 (12.36)	529 (12.01)
Parent or sibling with diabetes, No. (%)	1756 (34.97)	302 (49.11)	1454 (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)
Body mass index, mean (SD), kg/m^2	27.98 (5.26)	31.09 (5.93)	27.53 (5.01)
25-29.9, No. (%)	2011 (40.05)	222 (36.10)	1789(40.60)
≥30, No. (%)	1473 (29.34)	314 (51.06)	1159 (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, No. (%)[b]	2010 (40.03)	324 (52.68)	1686 (32.86)
Taking anti-hypertensive medication, No. (%)	1616 (32.18)	272 (44.23)	1344 (30.50)
Individual-Level Socioeconomic Status Variables			
Education, mean (SD), y	13.50 (3.83)	12.99 (3.99)	13.57 (3.81)
High school or less, No. (%)	1602 (31.91)	226 (36.75)	1376 (31.23)
Some college or associates/technical degree, No. (%)	1438 (28.64)	190 (30.89)	1248 (28.33)
Bachelor's degree or higher, No. (%)	1981 (39.45)	199 (32.36)	1782 (40.44)
Household income, mean (SD), per $10K	52.16 (34.46)	47.81 (32.24)	52.77 (34.72)
<25,000, No. (%)	1380 (27.48)	183 (29.76)	1197 (27.17)
25,000-75,000, No. (%)	2356 (46.92)	315 (51.22)	2041 (46.32)
75,000+, No. (%)	1285 (25.59)	117 (19.02)	1168 (26.51)
Household income per capita, mean (SD), per $10K[c]	2.77 (2.12)	2.42 (1.83)	2.82 (2.15)
Socioeconomic status index, mean (SD)[d]	0.00 (1.71)	-0.30 (1.62)	0.04 (1.71)
Area-Level Socioeconomic Status Variables			
Percent of adults 25+ 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 $10K	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)
Socioeconomic status index, mean (SD)[d]	0.79 (5.66)	-0.25 (5.56)	0.94 (5.66)
Combined Individual and Area-Level Socioeconomic Status			
Combined socioeconomic status index, mean (SD)[d]	0.00 (1.72)	-0.36 (1.64)	0.05 (1.73)

Abbreviations: HDL, high-density lipoprotein; SD, standard deviation; IQR, inter-quartile range

[a] Defined as waist circumference >88 cm for women and >102 cm for men.

[b] Hypertension defined as blood pressure ≥140/90 mmHg, or treatment with anti-hypertensive medication

[c] Defined as annual household income divided by the number of persons supported by that income

[d] Higher values indicate higher socioeconomic status

Table 4.2 Association of individual and area-level socioeconomic status variables with incident type 2 diabetes in prediction models containing clinical and laboratory variables, Multi-Ethnic Study of Atherosclerosis, 2000-2012[a]

Socioeconomic Status Predictors[b]	Multivariable Clinical Model[c]				Multivariable Lab Model[d]			
	HR (95% CI)[e]	*P* value	C	AIC	HR (95% CI)[e]	*P* value	C	AIC
None			0.72	9941.5			0.83	9254.2
Individual-level								
Education (continuous)	0.94 (0.86, 1.03)	0.17	0.72	9941.5	1.00 (0.91, 1.10)	0.99	0.83	9256.2
Education (categorical)								
High school or less	1.00 (Ref)	0.36[f]	0.72	9943.5	1.00 (Ref)	0.97[f]	0.83	9258.1
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	9942.2	0.91 (0.83, 1.01)	0.07	0.83	9252.3
Household income (categorical)								
<25,000	1.00 (Ref)	0.18[f]	0.72	9941.9	1.00 (Ref)	0.02[f]	0.83	9250.3
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	9939.8	0.91 (0.82, 1.02)	0.09	0.83	9252.7
Socioeconomic status index	0.91 (0.82, 1.00)	0.05	0.72	9939.5	0.95 (0.86, 1.05)	0.33	0.83	9255.1
Area-level[g]								
Percent of adults 25+ with a bachelor's degree	0.91 (0.83, 1.01)	0.07	0.72	9939.9	0.92 (0.83, 1.02)	0.13	0.83	9253.5
Median household income	0.97 (0.89, 1.07)	0.60	0.72	9943.2	0.92 (0.83, 1.02)	0.12	0.83	9253.1
Percent of persons living below the poverty line	1.02 (0.94, 1.12)	0.61	0.72	9943.2	1.07 (0.98, 1.16)	0.14	0.83	9254.0
Socioeconomic status index	0.93 (0.84, 1.02)	0.12	0.72	9940.9	0.91 (0.83, 1.01)	0.07	0.83	9252.5
Combined individual and area-level								
Overall socioeconomic status index	0.90 (0.81, 0.99)	0.06	0.72	9939.8	0.92 (0.83, 1.01)	0.09	0.83	9253.0

Abbreviations: HR, hazard ratio; CI, confidence interval.

[a] Number of incident diabetes cases = 615

[b] Socioeconomic status 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 standard deviation (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 socioeconomic status index, 22% for percent of adults 25+ 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 socioeconomic status index, and 1.72 units for overall socioeconomic status index.

[f] P-value corresponds to a likelihood ratio test for all categories combined

[g] Area-level variables measured at the census-tract level

Figure 4.1 Calibration of clinical models with and without socioeconomic information by tertiles of individual household income per capita, individual socioeconomic status (SES) index, area-level education, and area SES index, Multi-Ethnic Study of Atherosclerosis, 2000-2012

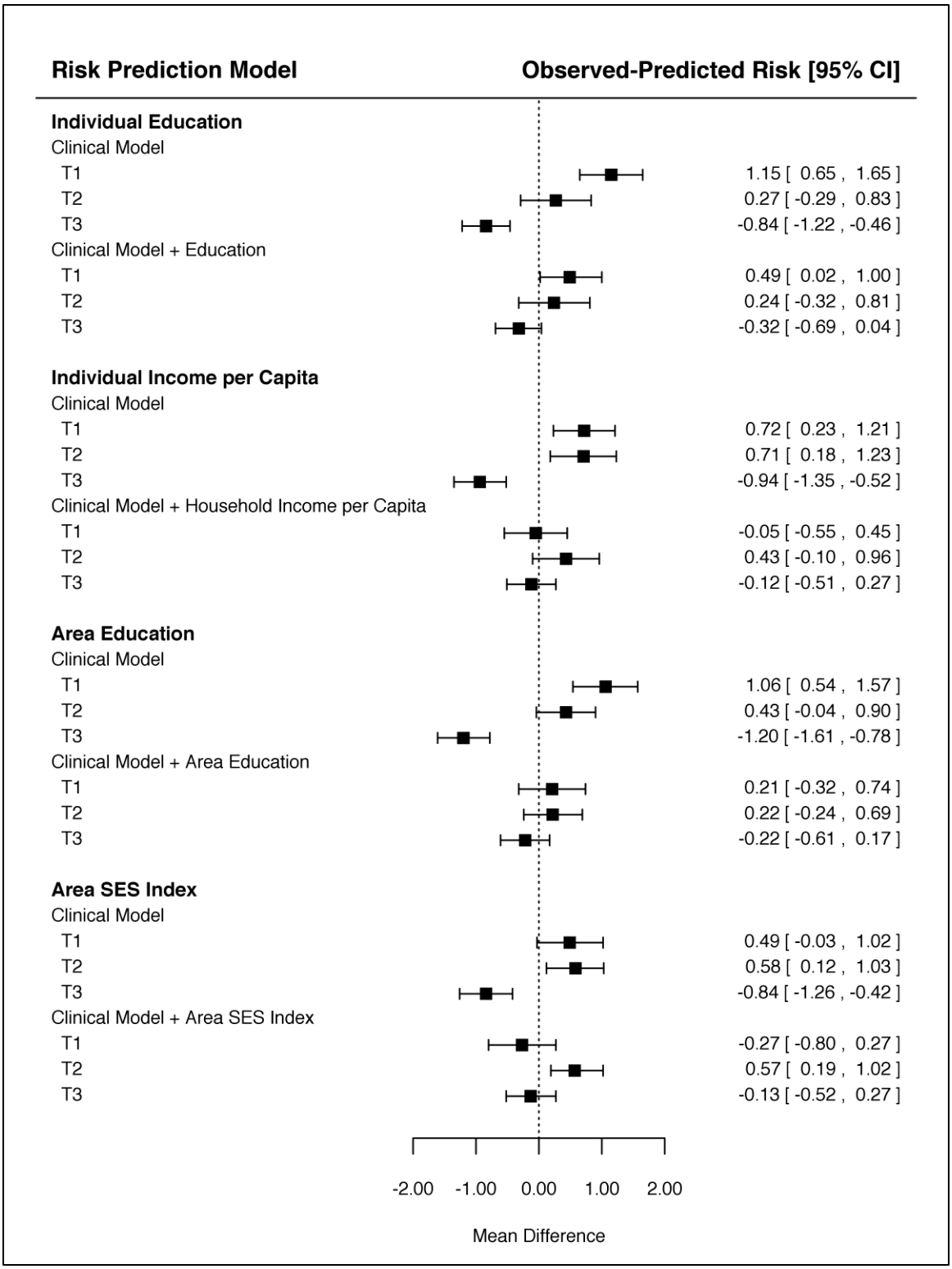
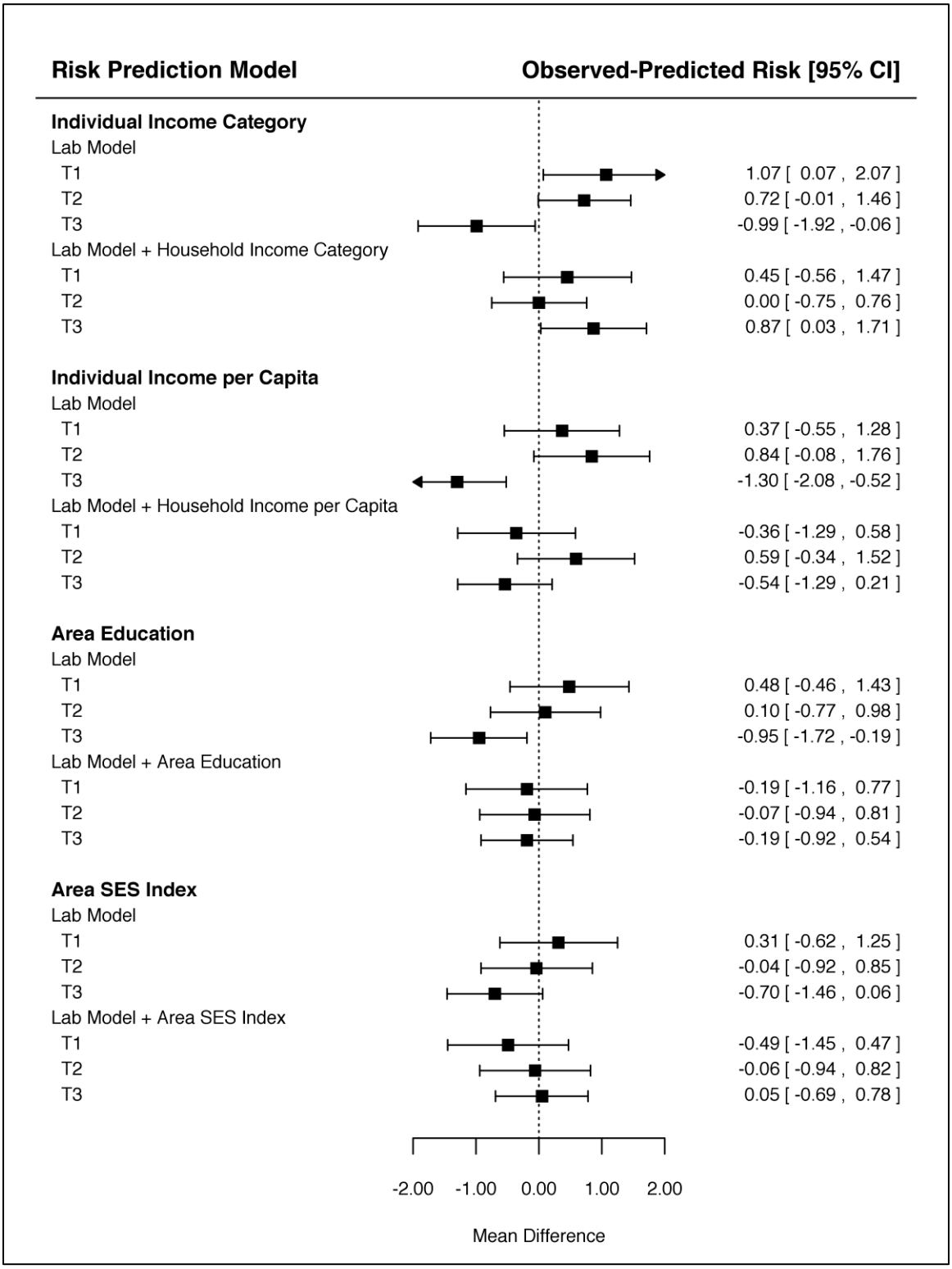


Figure 4.2 Calibration of laboratory models with and without socioeconomic information by tertiles of individual income category, individual household income per capita, area education, and area socioeconomic status (SES) index, Multi-Ethnic Study of Atherosclerosis, 2000-2012



Appendix

Table 4.A1 All model coefficients from clinical models without and with the most predictive socioeconomic variables[a]

Model Variable	Clinical Only		Clinical + Individual Household Income per Capita		Clinical + Individual SES Index		Clinical + Area-Level Education		Clinical + Area SES Index	
	HR (95 % CI)	P-value	HR (95 % CI)	P-value	HR (95 % CI)	P-value	HR (95 % CI)	P-value	HR (95 % CI)	P-value
Age										
Linear term	[b]	<0.01	[b]	<0.01	[b]	<0.01	[b]	<0.01	[b]	<0.01
Non-linear term	[b]	<0.01	[b]	<0.001	[b]	<0.001	[b]	<0.001	[b]	<0.001
Race/ethnicity										
White (ref)	1.00	<0.0001[c]	1.00	<0.0001[c]	1.00	<0.0001[c]	1.00	<0.0001[c]	1.00	<0.0001[c]
Chinese	2.33 (1.78, 3.05)	<0.0001	2.15 (1.62, 2.96)	<0.0001	2.18 (1.64, 2.89)	<0.0001	2.27 (1.74, 2.98)	<0.0001	2.30 (1.75, 3.01)	<0.0001
Black	1.34 (1.08, 1.66)	<0.01	1.30 (1.05, 1.61)	0.02	1.30 (1.05, 1.62)	0.02	1.26 (1.02, 1.57)	0.04	1.26 (1.01, 1.58)	0.04
Hispanic	1.62 (1.31, 2.01)	<0.0001	1.51 (1.20, 1.89)	<0.001	1.46 (1.14, 1.86)	<0.001	1.51 (1.21, 1.89)	<0.001	1.50 (1.19, 1.90)	<0.001
Family history of type 2 diabetes	1.65 (1.40, 1.94)	<0.0001	1.65 (1.40, 1.94)	<0.0001	1.65 (1.40, 1.94)	<0.0001	1.65 (1.40, 1.94)	<0.0001	1.66 (1.41, 1.95)	<0.0001
Systolic blood pressure	1.09 (1.05, 1.14)	<0.0001	1.09 (1.05, 1.14)	<0.0001	1.09 (1.05, 1.13)	<0.0001	1.09 (1.05, 1.14)	<0.0001	1.09 (1.05, 1.14)	<0.0001
Waist circumference	1.43 (1.36, 1.51)	<0.0001	1.43 (1.35, 1.50)	<0.0001	1.43 (1.36, 1.50)	<0.0001	1.43 (1.36, 1.50)	<0.0001	1.43 (1.36, 1.51)	<0.0001
Hypertension treatment	1.42 (1.19, 1.70)	<0.0001	1.43 (1.20, 1.70)	<0.0001	1.43 (1.20, 1.70)	<0.0001	1.42 (1.19, 1.69)	<0.0001	1.41 (1.19, 1.68)	<0.001
Individual SES variables										
Household income per capita			0.91 (0.82, 1.01)	0.07						
SES index					0.91 (0.82, 1.00)	0.05				
Area SES variables										
Percent with bachelor's degree or higher							0.91 (0.83, 1.01)	0.07		
SES index									0.93 (0.84, 1.02)	0.12

Abbreviations: HR, hazard ratio; CI, confidence interval; SES, socioeconomic status

[a] Unit changes for estimates of continuous variables: systolic blood pressure, 10 mmHg; waist circumference, 10 cm; SES predictors, standard deviations ($20,120 for individual household income per capita, 1.71 units for individual socioeconomic status index, 22% for percent of adults 25+ with a bachelor's degree, and 5.66 units for area socioeconomic status index).

[b] Age modeled using penalized b spline with 2 degrees of freedom; For effect estimates across the range of values, see Figure 4.A1.

[c] P-value from likelihood ratio test for all categories comparing nested models with and without race/ethnicity.

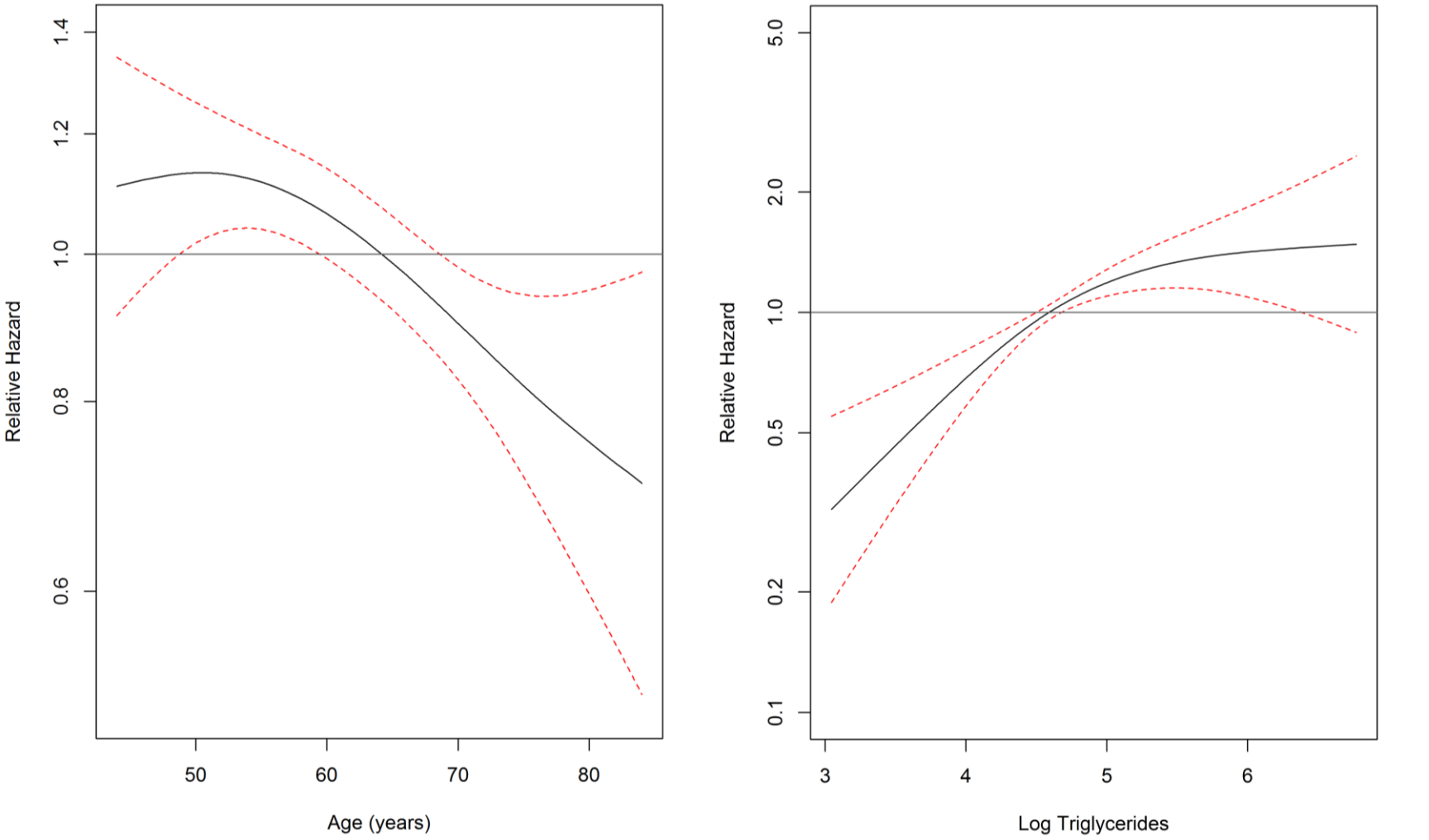
Table 4.A2 All model coefficients from laboratory models without and with the most predictive socioeconomic variables[a]

Model Variable	Lab Model		Lab + Individual Income Category		Lab + Individual Household Income per Capita		Lab + Area-Level Education		Lab + Area-Level SES Index	
	HR (95 % CI)	P-value	HR (95 % CI)	P-value	HR (95 % CI)	P-value	HR (95 % CI)	P-value	HR (95 % CI)	P-value
Age										
Linear term	[b]	<0.0001	[b]	<0.0001	[b]	<0.0001	[b]	<0.0001	[b]	<0.0001
Non-linear term	[b]	0.03	[b]	0.02	[b]	0.02	[b]	0.03	[b]	0.03
Race/ethnicity										
White (ref)	1.00	0.04[c]	1.00	0.10[c]	1.00	0.11[c]	1.00	0.13[c]	1.00	0.15[c]
Chinese	1.29 (0.96, 1.72)	0.09	1.23 (0.92, 1.66)	0.17	1.20 (0.88, 1.63)	0.25	1.26 (0.95, 1.69)	0.11	1.27 (0.95, 1.69)	0.11
Black	1.34 (1.07, 1.68)	0.01	1.30 (1.03, 1.64)	0.03	1.30 (1.03, 1.64)	0.03	1.28 (1.01, 1.61)	0.04	1.25 (0.99, 1.59)	0.06
Hispanic	1.17 (0.93, 1.48)	0.19	1.10 (0.86, 1.40)	0.46	1.09 (0.85, 1.41)	0.48	1.10 (0.87, 1.41)	0.42	1.07 (0.84, 1.38)	0.57
Family history of type 2 diabetes	1.49 (1.26, 1.76)	<0.0001	1.49 (1.26, 1.77)	<0.0001	1.49 (1.25, 1.76)	<0.0001	1.49 (1.26, 1.77)	<0.0001	1.50 (1.27, 1.77)	<0.0001
Systolic blood pressure	1.05 (1.01, 1.10)	0.03	1.05 (1.01, 1.10)	0.03	1.05 (1.01, 1.10)	0.03	1.05 (1.00, 1.10)	0.03	1.05 (1.00, 1.10)	0.03
Waist circumference	1.21 (1.13, 1.29)	<0.0001	1.20 (1.12, 1.28)	<0.0001	1.21 (1.13, 1.29)	<0.0001	1.21 (1.14, 1.29)	<0.0001	1.21 (1.14, 1.29)	<0.0001
Hypertension treatment	1.25 (1.04, 1.51)	0.02	1.26 (1.04, 1.51)	0.02	1.26 (1.05, 1.51)	0.01	1.24 (1.03, 1.49)	0.02	1.24 (1.03, 1.48)	0.02
Fasting glucose	2.58 (2.38, 2.80)	<0.0001	2.60 (2.39, 2.82)	<0.0001	2.58 (2.38, 2.80)	<0.0001	2.58 (2.38, 2.79)	<0.0001	2.59 (2.38, 2.80)	<0.0001
Log Triglycerides										
Linear term	[b]	<0.001	[b]	<0.001	[b]	<0.001	[b]	<0.001	[b]	<0.0001
Non-linear term	[b]	<0.0001	[b]	<0.0001	[b]	<0.0001	[b]	<0.0001	[b]	<0.0001
Individual SES variables										
Income category										
<25,000 (Ref)			1.00	.02[c]						
25,000-74,999			1.01 (0.83, 1.23)	0.94						
≥75,000			0.75 (0.58, 0.97)	0.03						
Household income per capita					0.91 (0.82, 1.02)	0.09				
Area SES variables										
% bachelor's degree or higher							0.92 (0.83, 1.02)	0.13		
SES Index									0.98 (0.97, 1.00)	0.07

Abbreviations: HR, hazard ratio; CI, confidence interval; SES, socioeconomic status

[a] Unit changes for estimates of continuous variables: systolic blood pressure, 10 mmHg; waist circumference, 10 cm; fasting glucose, 10mg/dL; SES predictors, standard deviations ($20,120 for individual household income per capita, 22% for percent of adults 25+ with a bachelor's degree, and 5.66 units for area socioeconomic status index).
[b] Age and triglycerides modeled using penalized b splines with 2 degrees of freedom. For effect estimates for triglycerides across the range of values, see Figure 4.A1.
[c] P-value from likelihood ratio test for all categories comparing nested models with and without the categorical variable.

Figure 4.A1 Graphical representation of non-linearity of age (left) and triglyceride (right) effects on risk for type 2 diabetes[a]



[a] Age and triglyceride effects from fully-adjusted Cox proportional hazards models estimated using penalized b splines with 2 degrees of freedom. Red hashed lines represent 95% confidence intervals.

Figure 4.A2 Receiver operator characteristic curves showing area under the curve for incident type 2 diabetes in clinical (left) and laboratory (right) prediction models after 10 years of follow-up

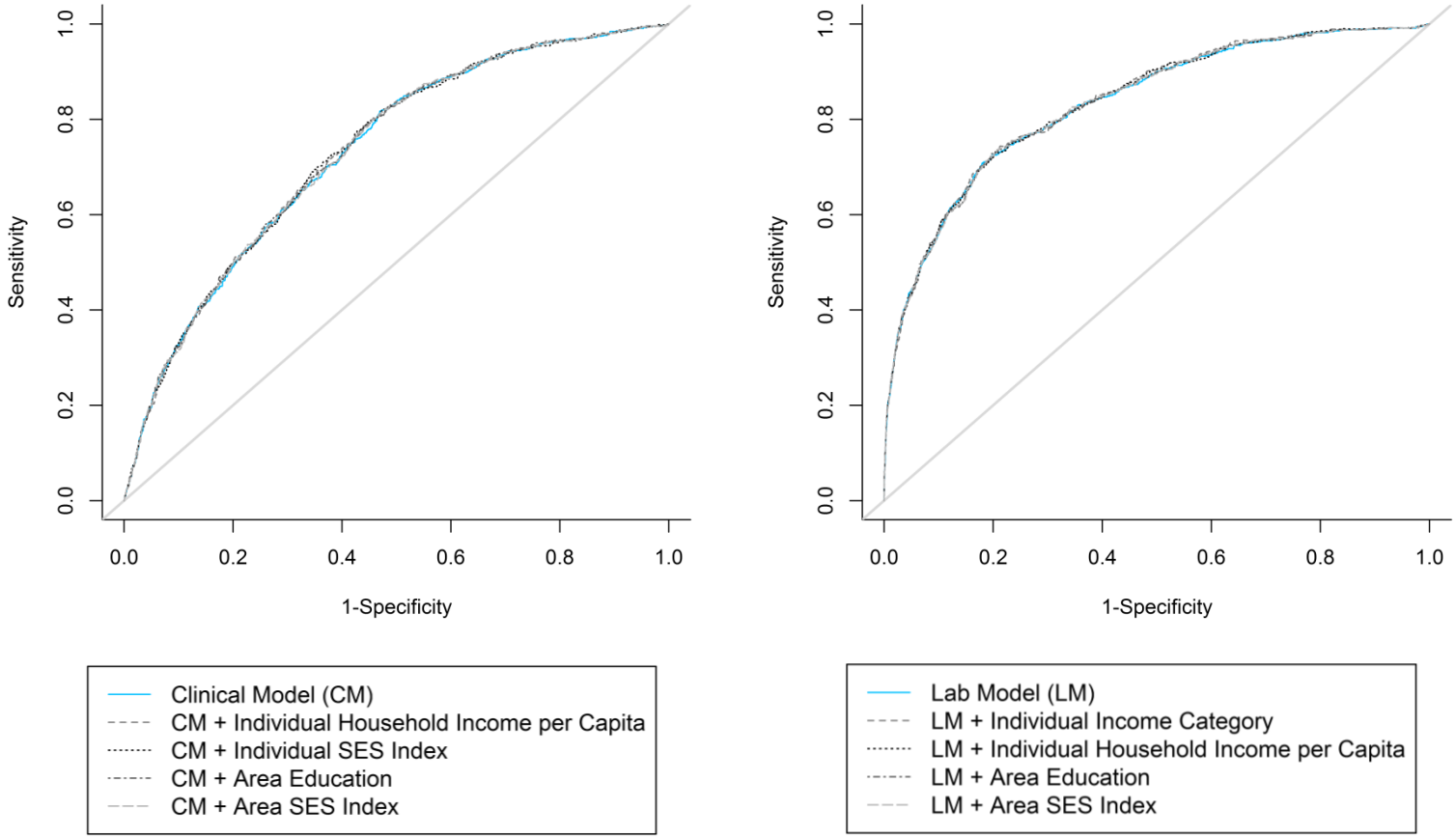


Table 4.A3 10-year risk of type 2 diabetes and risk reclassification in clinical prediction models with and without individual-level household income per capita and socioeconomic status index

10-year risk in clinical model without SES	10-year risk in clinical model with individual household income per capita					10-year risk in clinical model with individual SES index				
	0 to <10%	10-20%	>20%	% Reclassified	% Net Correctly Reclassified	0 to <10%	10-20%	>20%	% Reclassified	% Net Correctly Reclassified
0 to <10%										
No. of participants	2086	87	0	4.0	-3.4	2078	95	0	4.4	-3.8
No. of events	84	7	0	7.7	7.7	85	6	0	6.7	6.7
No. with no events	2002	80	0	3.9	-3.9	1993	89	0	4.3	-4.3
Kaplan-Meier 10-year estimate (95% CI)	4.55 (3.68, 5.62)	10.66 (5.18, 21.23)	NA			4.63 (3.75, 5.70)	8.63 (3.91, 18.49)	NA		
10-20%										
No. of participants	94	1792	59	7.9	1.6	96	1789	60	8.0	1.6
No. of events	11	250	9	7.4	-0.7	10	252	8	6.7	-0.7
No. with no events	83	1542	50	7.9	2.0	86	1537	52	8.2	2.0
Kaplan-Meier 10-year estimate (95% CI)	13.14 (7.45, 22.60)	15.68 (13.96, 17.58)	18.21 (9.81, 32.39)			11.02 (6.05, 19.61)	15.87 (14.14, 17.78)	14.96 (7.70, 27.94)		
>20%										
No. of participants	0	47	856	5.2	3.2	0	56	847	6.2	4.2
No. of events	0	9	245	3.5	-3.5	0	9	245	3.5	-3.5
No. with no events	0	38	611	5.9	5.9	0	47	602	7.2	7.2
Kaplan-Meier 10-year estimate (95% CI)	NA	19.36 (10.57, 33.91)	31.78 (28.55, 35.28)			NA	16.73 (9.06, 29.74)	32.05 (28.81, 35.57)		

Abbreviations: SES, socioeconomic status

Table 4.A4 10-year risk of type 2 diabetes and risk reclassification in clinical prediction models with and without area-level education and socioeconomic status index

10-year risk in clinical model without SES	10-year risk in clinical model with area-level education					10-year risk in clinical model with area-level SES index				
	0 to <10%	10-20%	>20%	% Reclassified	% Net Correctly Reclassified	0 to <10%	10-20%	>20%	% Reclassified	% Net Correctly Reclassified
0 to <10%										
No. of participants	2083	90	0	4.1	-3.6	2109	64	0	3.0	-2.5
No. of events	85	6	0	6.6	6.6	86	5	0	5.5	5.5
No. with no events	1998	84	0	4.0	-4.0	2023	59	0	2.8	-2.8
Kaplan-Meier 10-year estimate (95% CI)	4.62 (3.74, 5.69)	8.38 (3.84, 17.78)	NA			4.62 (3.75, 5.68)	10.51 (4.48, 23.61)	NA		
10-20%										
No. of participants	102	1788	55	8.1	2.1	74	1826	45	6.1	1.3
No. of events	10	253	7	6.3	-1.1	9	254	7	5.9	-0.7
No. with no events	92	1535	48	8.5	2.7	65	1572	38	6.1	1.6
Kaplan-Meier 10-year estimate (95% CI)	10.72 (5.87, 19.13)	15.90 (14.17, 17.82)	14.63 (7.12, 28.73)			13.47 (7.18, 24.49)	15.60 (13.91, 17.49)	20.69 (10.23, 39.23)		
>20%										
No. of participants	0	42	861	4.7	2.7	0	39	864	4.3	3.2
No. of events	0	9	245	3.5	-3.5	0	5	249	2.0	-2.0
No. with no events	0	33	616	5.1	5.1	0	34	615	5.2	5.2
Kaplan-Meier 10-year estimate (95% CI)	NA	21.94 (12.05, 37.98)	31.54 (28.33, 35.01)			NA	13.15 (5.68, 28.81)	31.92 (28.71, 35.39)		

Abbreviations: SES, socioeconomic status

Table 4.A5 10-year risk of type 2 diabetes and risk reclassification in laboratory prediction models with and without individual-level income category and household income per capita

10-year risk in laboratory model without SES	10-year risk in laboratory model with individual income category					10-year risk in laboratory model with individual household income per capita				
	0 to <10%	10-20%	>20%	% Reclassified	% Net Correctly Reclassified	0 to <10%	10-20%	>20%	% Reclassified	% Net Correctly Reclassified
0 to <10%										
No. of participants	3054	77	0	2.5	-1.8	3077	54	0	1.7	-1.3
No. of events	117	11	0	8.6	8.6	121	7	0	5.5	5.5
No. with no events	2937	66	0	2.2	-2.2	2956	47	0	1.6	-1.6
Kaplan-Meier 10-year estimate (95% CI)	4.45 (3.72, 5.32)	17.11 (9.74, 29.08)	NA			4.56 (3.82, 5.43)	15.56 (7.66, 30.16)	NA		
10-20%										
No. of participants	65	829	43	11.5	3.0	45	864	28	7.8	2.7
No. of events	6	92	9	14.0	2.8	2	99	6	7.5	3.7
No. with no events	59	737	34	11.2	3.0	43	765	22	7.8	2.5
Kaplan-Meier 10-year estimate (95% CI)	3.91 (4.60, 21.06)	12.77 (10.56, 15.46)	22.78 (12.41, 39.61)			4.80 (1.22, 17.92)	13.13 (10.89, 15.79)	25.63 (12.16, 49.16)		
>20%										
No. of participants	0	31	922	3.3	2.4	0	18	935	1.9	1.0
No. of events	0	4	376	1.1	-1.1	0	4	376	1.1	-1.1
No. with no events	0	27	546	4.7	4.7	0	14	559	2.4	2.4
Kaplan-Meier 10-year estimate (95% CI)	NA	16.04 (6.31, 37.45)	45.13 (41.72, 48.69)			NA	27.35 (11.14, 57.88)	44.47 (41.09, 48.00)		

Abbreviations: SES, socioeconomic status

Table 4.A6 10-year risk of type 2 diabetes and risk reclassification in laboratory prediction models with and without area-level education and socioeconomic status index

10-year risk in laboratory model without SES	10-year risk in laboratory model with area-level education					10-year risk in laboratory model with area-level SES index				
	0 to <10%	10-20%	>20%	% Reclassified	% Net Correctly Reclassified	0 to <10%	10-20%	>20%	% Reclassified	% Net Correctly Reclassified
0 to <10%										
No. of participants	3084	47	0	1.5	-1.1	3074	57	0	1.8	-1.3
No. of events	121	7	0	5.5	5.5	120	8	0	6.3	6.3
No. with no events	2963	40	0	1.3	-1.3	2954	49	0	1.6	-1.6
Kaplan-Meier 10-year estimate (95% CI)	4.56 (3.82, 5.43)	17.46 (8.61, 33.56)	NA			4.54 (3.80, 5.41)	16.53 (8.50, 30.75)	NA		
10-20%										
No. of participants	44	861	32	8.1	1.3	48	854	35	8.9	1.6
No. of events	3	101	3	5.6	0	3	100	4	6.5	0.9
No. with no events	41	760	29	8.4	1.4	45	754	31	9.2	1.7
Kaplan-Meier 10-year estimate (95% CI)	6.93 (2.29, 19.96)	13.38 (11.13, 16.06)	12.54 (4.16, 34.45)			6.34 (2.09, 18.39)	13.42 (11.15, 16.11)	12.88 (5.00, 30.98)		
>20%										
No. of participants	0	23	930	2.4	1.6	0	25	928	2.6	1.8
No. of events	0	4	376	-1.1	-1.1	0	4	376	1.1	-1.1
No. with no events	0	19	554	3.3	3.3	0	21	552	3.7	3.7
Kaplan-Meier 10-year estimate (95% CI)	NA	21.39 (8.56, 47.67)	44.74 (41.35, 48.29)			NA	21.08 (8.39, 47.26)	44.80 (41.40, 48.34)		

Abbreviations: SES, socioeconomic status

Table 4.A7 Distribution of predicted risk from clinical and laboratory prediction models by tertiles of individual income per capita and area-level socioeconomic status index[a]

	Low Tertile Individual SES			Middle Tertile Individual SES			High Tertile Individual SES			Total, n
Predicted 10-year risk score for diabetes	**Low Tertile Area SES**	**Middle Tertile Area SES**	**High Tertile Area SES**	**Low Tertile Area SES**	**Middle Tertile Area SES**	**High Tertile Area SES**	**Low Tertile Area SES**	**Middle Tertile Area SES**	**High Tertile Area SES**	
Total, n	865	516	268	478	500	455	329	653	957	5021
Clinical Model										
0 to <10%	30.52	34.88	78.73	34.73	45.20	50.11	37.69	44.10	60.19	2173
10-20%	45.09	47.67	37.69	41.84	34.80	35.82	36.78	38.59	31.14	1945
>20%	24.39	17.44	17.16	23.43	20.00	14.07	25.53	17.30	8.67	903
Laboratory Model										
0 to <10%	53.29	57.75	65.67	62.34	60.00	65.93	57.45	62.94	72.94	3131
10-20%	21.73	23.03	16.79	16.11	18.80	17.36	23.40	15.77	15.67	937
>20%	24.97	18.22	17.54	21.55	21.20	16.70	19.15	21.29	11.39	953

Abbreviations: SES, socioeconomic status

[a] Individual household income per capita used for individual SES, and census tract SES index used for area SES. Cell values are column percentages for Clinical and Laboratory models separately.

Table 4.A8 Calibration of established diabetes risk scores by tertiles of socioeconomic status variables

Risk Prediction Model	Observed – Predicted, Mean Difference (95% CI)	
	ARIC Risk Score	**Framingham Risk Score**
Individual income per capita		
Model variables only		
Lowest Tertile	-0.29 (-1.19, 0.60)	1.63 (1.03, 2.22)
Middle Tertile	1.40 (0.65, 2.15)	1.17 (0.58, 1.76)
Highest Tertile	-1.18 (-1.82, -0.54)	-1.64 (-2.18, -1.17)
Model variables + household income per capita		
Lowest Tertile	-1.21 (-2.04, -0.38)	-0.26 (-0.89, 0.36)
Middle Tertile	1.05 (0.30, 1.79)	0.67 (0.07, 1.27)
Highest Tertile	-0.22 (-0.92, 0.48)	0.17 (-0.28, 0.62)
Individual SES index		
Model variables only		
Lowest Tertile	-0.48 (-1.38, 0.43)	1.71 (1.09, 2.32)
Middle Tertile	1.10 (0.39, 1.82)	0.82 (0.26, 1.38)
Highest Tertile	-0.93 (-1.56, -0.30)	-1.68 (-2.13, -1.22)
Model variables + individual SES index		
Lowest Tertile	-1.17 (-1.98, -0.36)	-0.52 (-1.14, 0.10)
Middle Tertile	0.90 (0.16, 1.63)	0.86 (0.30, 1.41)
Highest Tertile	-0.11 (-0.82, 0.61)	0.24 (-0.20, 0.69)
Area education		
Model variables only		
Lowest Tertile	0.79 (-0.09, 1.68)	2.60 (2.01, 3.18)
Middle Tertile	0.16 (-0.61, 0.92)	0.51 (-0.05, 1.06)
Highest Tertile	-1.35 (-1.97, -0.73)	-2.32 (-2.81, -1.83)
Model variables + area education		
Lowest Tertile	-0.23 (-1.10, 0.64)	0.32 (-0.34, 0.98)
Middle Tertile	-0.13 (-0.88, 0.63)	0.02 (-0.56, 0.60)
Highest Tertile	-0.14 (-0.77, 0.48)	0.16 (-0.27, 0.59)
Area SES index		
Model variables only		
Lowest Tertile	0.34 (-0.54, 1.22)	2.22 (1.64, 2.79)
Middle Tertile	0.27 (-0.49, 1.03)	0.44 (-0.14, 1.02)
Highest Tertile	-1.07 (-1.69, -0.45)	-1.96 (-2.43, -1.48)
Model variables + area SES index		
Lowest Tertile	-0.68 (-1.56, 0.20)	-0.16 (-0.82, 0.51)
Middle Tertile	0.25 (-0.51, 1.00)	0.46 (-0.13, 1.04)
Highest Tertile	-0.10 (-0.72, 0.52)	0.17 (-0.25, 0.58)

Abbreviations: SES, socioeconomic status

CHAPTER 5 :

DISCUSSION

Summary and Implications of Main Findings

Despite its demonstrated preventability, the prevalence and incidence of type 2 diabetes continue to rise. As such, attention has begun to shift from exclusively individual-based prevention strategies to population health-based approaches. One potential avenue for a population-based approach is altering residential environments to support healthy behaviors and promote wellbeing. This dissertation provides evidence that such an approach may indeed be helpful for preventing diabetes, and highlights particular ways in which individual and environmental factors may interact to inhibit or promote the development of disease.

In chapter 2, we found that long term exposure to neighborhood environments with more resources to support physical activity, and to a lesser extent healthy diets, was associated with lower risk for developing type 2 diabetes. These findings point to specific neighborhood characteristics that may partly explain the associations between more general neighborhood environments and diabetes observed in other studies that were unable to track specific neighborhood features, including the experimental MTO study.[1] Contrary to our hypothesis, levels of social cohesion and safety were largely not associated with diabetes incidence. This could be due to the true absence of a causal effect or inadequate measures of the neighborhood constructs (e.g. the use of survey-based safety instead of crime data). In analyses exploring effect modification of the

neighborhood exposures by select individual attributes, we found that high income individuals were more likely to benefit from the presence of healthy food stores and recreational establishments. They were, however, less likely to benefit from higher levels of neighborhood safety, suggesting that an overall association may have been masked by subgroup heterogeneity. These results mirror what has already been suggested in the literature: that the simple presence of health promoting resources may not be equally beneficial to all residents, and that factors like affordability (for food and physical activity resources) may be especially pertinent to low-income individuals.[2,3]

The results in chapter 2 represent an important contribution to the literature on neighborhood environments and diabetes. In contrast to prior work, we used longitudinal data on specific neighborhood exposures and incident diabetes, providing stronger causal evidence of the relationship between neighborhood exposures and diabetes risk. The results also provide evidence that is pertinent to policy questions regarding the health effects of neighborhood change on residents that continually live in the neighborhood. These questions are not easily answerable with experimental studies like the MTO, which focus on residential relocation as the "treatment" of interest. Evidence from the MTO study showing adverse behavioral and mental health effects on adolescent males also points towards the potential negative consequences of residential relocation as a mechanism for improving neighborhood environments.[4-6] As such, research examining the dynamics of neighborhood change, health behaviors, and diabetes risk is important and policy relevant, and additional studies are warranted to increase confidence in the results observed.

Chapter 3 built upon work from chapter 2 by demonstrating that genetic susceptibility to type 2 diabetes was significantly modified by the availability of healthy food stores and recreational establishments. Specifically, greater availability of healthy food stores and, to a lesser extent, recreational establishments weakened the association between genetic risk and type 2 diabetes. Neighborhood SES did not modify the genetic risk for diabetes, though strong correlations between race/ethnicity and neighborhood SES made these analyses difficult to interpret and raised the issue of possible structural confounding.[7] In analyses with dichotomized exposures, the effects of high genetic risk and decreased availability of healthy food and recreational establishments appeared to interact in a synergistic manner. Such results indicate that increased genetic risk for type 2 diabetes may be more pernicious (or may only be evident) in environments with few resources to support healthy behaviors.

While additional replication studies are needed, the preliminary results from chapter 3 may have implications for future research on genetic and neighborhood effects on diabetes, and for understanding disparities in diabetes burden. The fact that allelic penetrance may vary by neighborhood environments (or factors strongly correlated with these environments) raises important questions about the interpretation and stability of genetic risk estimates over space and time.[8] In a similar manner, neighborhood effect estimates like those from chapter 2 may be more heterogeneous than expected depending upon the distribution of genetic risk among residents (a point that has been made in previous gene-by-neighborhood environment studies).[9,10] Finally, regarding disparities in diabetes, there is an ongoing debate about whether disparities by race and SES are caused by genetic, behavioral, and/or environmental factors.[11,12] . To the extent that

neighborhood resources like healthy food availability are segregated by race and/or poverty status,[13] and genetic risk is not, the results from chapter 3 would suggest that disparities in type 2 diabetes observed by race and SES may arise primarily due to differences in environments and their related behaviors rather than genes.[14]

Given the findings that neighborhood features were associated with diabetes risk and modified genetic risk, chapter 4 evaluated the utility of including individual and area-level socioeconomic information into public health and clinical decision making through diabetes risk prediction models. While area-level SES may not be a perfect proxy for the specific neighborhood characteristics that drive diabetes risk, measures of area-level SES are widely available and could feasibly be linked to electronic medical records. Surprisingly, the inclusion of individual and area-level SES did not help to discriminate between who would and would not go on to develop diabetes. This suggests that the traditional risk factors included in risk scores likely capture much of the effect of SES on diabetes risk. The inclusion of SES variables, however, did have implications for the accuracy of the risk predictions. Consistent with similar research on cardiovascular risk models,[15-17] diabetes risk prediction models, which include only traditional, largely biological risk factors tended to have differential prediction accuracy by SES. We observed an underestimation of risk among individuals of low-SES and for those residing in low-SES neighborhoods but an overestimation of risk for high-SES individuals and those residing in high-SES neighborhoods. Adding individual and area-level SES measures to the prediction models, particularly individual household income and area-level education, eliminated these inaccuracies.

While the magnitude of under/overestimation using only traditional risk factors was on average no more than 1 to 2 percent, the significance of this model "miscalibration" would depend upon where in the risk distribution an individual falls. An under- or overestimate near a treatment threshold, for instance, would be far more problematic than near the low or high ends of the risk distribution. Though no treatment thresholds currently exist for diabetes care as they do for cardiovascular disease interventions, studies which promote treatments based upon individual predicted risk are increasingly common.[18] Ensuring that the risk models used to guide such decisions perform equally well across the spectrum of social advantage/disadvantage is thus important to prevent inadvertently widening disparities based upon inaccurate projected risks.[16]

Collectively, the results of this dissertation support the legitimacy of a population-based approach to diabetes prevention. They furthermore identify specific neighborhood features that could feasibly be altered to help prevent diabetes development and modify inherited risks for disease. Chapter 2 demonstrated that healthy food and physical activity resources likely shape health behaviors and diabetes risk, if only to a small degree in some cases. Yet even if environments have only small effects on behaviors, shifting the entire population towards slightly healthier behaviors may have a large influence on the population burden of diabetes. The results from chapter 3 illustrate the potential of such an approach: if the population can be thought of as having a distribution of genetic risk, then modifying neighborhood environments may effectively shift this distribution in a way that ultimately prevents many cases. And as research on neighborhood environments and diabetes risk moves forward, it remains important to apply these insights to public

health and clinical decision-making. Ensuring that prevention strategies account for and act on social and environmental causes of diabetes may not only help prevent inadvertently widening disparities, but may also represent a viable avenue for ultimately decreasing them.[14]

Strengths and Limitations

The work presented in this dissertation has several strengths. The use of longitudinal data with detailed information about specific neighborhood exposures, health behaviors, and disease outcomes is exceedingly rare in the literature on neighborhoods and health. Having numerous measures of specific neighborhood features allowed us to explore which characteristics of neighborhoods may independently, or jointly, influence diabetes risk. It also enabled the use of more theoretically appropriate cumulative measures of neighborhood environments that reflect the long term nature of the pertinent behavioral and disease processes. Linking neighborhood environment data to genetic risk was a novel approach to gene-environment interaction. Conceptualizing gene-environment interaction in this manner provides a needed expansion beyond "gene-behavior" interaction studies that fail to place individual behaviors in context.[10] Finally, chapter 4 employed an innovative approach to incorporating individual and area-level social information into public health and clinical decision processes via risk prediction modeling.

This work is not without limitations as well. First, the results are based on a single cohort of middle- to older-aged adults for whom we have relevant individual and neighborhood exposure data for only a small portion of their lifecourse. To the extent that individual and neighborhood exposures in childhood and young adulthood may affect

health behaviors and diabetes risk in ways not captured by data in later adulthood, we may be missing critical windows of exposure. With regard to gene-environment interaction in chapter 3, it is important to note that interaction results based upon single studies may be due to chance and need to be replicated. As in all studies of neighborhood effects on health, chapters 2 and 3 are susceptible to bias due to residential selection (i.e. endogeneity).[19] If individuals with certain health behaviors which influence their risk for diabetes elect to live in neighborhoods that are equipped with resources to promote those very behaviors, then the associations observed may actually reflect individual preferences and behaviors rather than true neighborhood effects. Finally, this research is based entirely upon observational data with well-known limitations. The causal nature of the associations should therefore not be over interpreted. Instead, this work should be interpreted within the context of the broader literature on neighborhoods and health, both quantitative and qualitative, and be viewed as a small contribution to a much larger body of work.

Future Directions

The analyses presented in this dissertation highlight the complexity of the causal pathways linking individual and area-level attributes to diabetes risk. In general, they support the notion that modification of neighborhood environments may provide a complementary, population-based approach to preventing diabetes. Nonetheless, there are several directions for future research that would strengthen our confidence in the causal nature of the associations observed, and guide both policy efforts and clinical and public health decisions to support diabetes prevention.

To strengthen causal inference regarding neighborhood changes and health (including diabetes), the field would benefit from more well-designed, quasi-experimental studies. Using exogenous changes in neighborhood resources can partly sidestep the most intractable forms of bias, including residential selection. For instance, supermarkets and recreational facilities are commonly opened and closed in communities throughout the US, but rarely are formal evaluations performed to assess their potential health effects on the surrounding community. The few quasi-experimental studies that do exist have shown mixed results, indicating that the addition of health promoting resources may not have straightforward effects on the health of residents.[20] Similar evaluation problems have been discussed with respect to larger social policies regarding education, immigration, and work, and their potential health effects.[21] As such, the public health community interested in the health effects of neighborhood changes and social policies should advocate for, plan and execute formal health evaluations of these processes.

Aside from more quasi-experimental studies, the literature linking neighborhoods and health would also benefit from nesting neighborhoods within larger city, regional, and state policy environments. Numerous policies with suspected health implications are enacted at the local level. For instance, city and state policies related to urban design (e.g. complete streets[22]), public assistance (e.g. the use of SNAP benefits at farmers' markets[23]), policing (e.g. "stop and frisk"[24]), and immigration enforcement (e.g. secure communities[25]) likely have health effects that modify, or are modified by, the physical and social neighborhood environments in which individuals live. There are few studies that examine these interactions. Efforts to link longitudinal neighborhood studies to larger social and health policies are thus needed to further contextualize "area" health effects on

residents, and to understand the conditions under which neighborhood changes are likely to lead to health improvements.

In terms of using contextual information to guide diabetes prevention decisions, it remains unclear whether individual or area-level social information can improve the ability to predict who will develop disease. However, the notion that risk prediction models may perform differently across SES strata is troubling, especially given increasing efforts to use such models to guide prevention and treatment decisions. In light of the possibility that a miscalibrated risk model could inadvertently exacerbate disparities, future research testing the performance of risk prediction models across the spectrum of social advantage/disadvantage may be important. For instance, the newly developed ACC/AHA pooled cardiovascular risk prediction model designed to guide statin prescribing decisions could be checked to ensure that it is equally accurate in subgroups defined by their social standing.[26] While the ultimate decision regarding whether or not to include social information in risk prediction should involve both ethical and statistical considerations, understanding the performance of existing models across social categories merits further research.

Finally, as mentioned in the introduction, the physical and social characteristics of neighborhoods that are the subject of this dissertation are but simple, specific examples of a larger and more complex social structure through which residential environments shape the health of their residents. Factors like the availability of healthy food and physical activity resources are likely important for ensuring equitable opportunities for individuals to live healthy lives. They are also, however, more politically palatable ways to discuss the causes of health disparities, and should not obfuscate the broader social forces (e.g.

structural racism) that place marginalized peoples at higher risk. Indeed, it would be naïve to assume that the provisioning of such factors would, by itself, greatly affect the disproportionate burden of diabetes experienced by low-income communities and communities of color. In this regard, complementary research addressing the more fundamental causes of disparities, including the mechanisms that sort low-income individuals and racial/ethnic minorities into resource-poor communities, is needed.

Conclusion

This dissertation suggests that altering neighborhood environments may represent a viable, population-based approach to the prevention of type 2 diabetes. While the pathways linking individual and neighborhood factors to type 2 diabetes are dynamic and interact in ways that may defy simple causal explanations, focusing exclusively on individual-based approaches to diabetes prevention while ignoring context is inadequate. Our hope is that by altering neighborhood environments and explicitly recognizing the importance of context in clinical and public health decision making, we may expand the scope of diabetes prevention and reduce disparities in the burden of diabetes on a population-wide scale.

References:

1. Ludwig J, Sanbonmatsu L, Gennetian L, et al. Neighborhoods, obesity, and diabetes--a randomized social experiment. *N. Engl. J. Med.* Oct 20 2011;365(16):1509-1519.

2. Drewnowski A, Aggarwal A, Hurvitz PM, Monsivais P, Moudon AV. Obesity and supermarket access: proximity or price? *Am. J. Public Health.* Aug 2012;102(8):e74-80.

3. Breyer B, Voss-Andreae A. Food mirages: geographic and economic barriers to healthful food access in Portland, Oregon. *Health Place.* Nov 2013;24:131-139.

4. Osypuk TL, Schmidt NM, Bates LM, Tchetgen-Tchetgen EJ, Earls FJ, Glymour MM. Gender and crime victimization modify neighborhood effects on adolescent mental health. *Pediatrics.* Sep 2012;130(3):472-481.

5. Osypuk TL, Tchetgen EJ, Acevedo-Garcia D, et al. Differential mental health effects of neighborhood relocation among youth in vulnerable families: results from a randomized trial. *Arch. Gen. Psychiatry.* Dec 2012;69(12):1284-1294.

6. Kessler RC, Duncan GJ, Gennetian LA, et al. Associations of housing mobility interventions for children in high-poverty neighborhoods with subsequent mental disorders during adolescence. *JAMA.* Mar 5 2014;311(9):937-948.

7. Messer LC, Oakes JM, Mason S. Effects of socioeconomic and racial residential segregation on preterm birth: a cautionary tale of structural confounding. *Am. J. Epidemiol.* Mar 15 2010;171(6):664-673.

8. Rosenquist JN, Lehrer SF, O'Malley AJ, Zaslavsky AM, Smoller JW, Christakis NA. Cohort of birth modifies the association between FTO genotype and BMI. *Proc. Natl. Acad. Sci. U. S. A.* Jan 13 2015;112(2):354-359.

9. Boardman JD, Barnes LL, Wilson RS, Evans DA, Mendes de Leon CF. Social disorder, APOE-E4 genotype, and change in cognitive function among older adults living in Chicago. *Soc. Sci. Med.* May 2012;74(10):1584-1590.

10. Boardman JD, Daw J, Freese J. Defining the environment in gene-environment research: lessons from social epidemiology. *Am. J. Public Health.* Oct 2013;103 Suppl 1:S64-72.

11. Golden SH, Brown A, Cauley JA, et al. Health disparities in endocrine disorders: biological, clinical, and nonclinical factors--an Endocrine Society scientific statement. *J. Clin. Endocrinol. Metab.* Sep 2012;97(9):E1579-1639.

12. Keaton JM, Cooke Bailey JN, Palmer ND, et al. A comparison of type 2 diabetes risk allele load between African Americans and European Americans. *Hum. Genet.* Dec 2014;133(12):1487-1495.

13. Bower KM, Thorpe Jr RJ, Rohde C, Gaskin DJ. The intersection of neighborhood racial segregation, poverty, and urbanicity and its impact on food store availability in the United States. *Prev. Med.* 2014;58(0):33-39.

14. Diez Roux AV. Conceptual approaches to the study of health disparities. *Annu. Rev. Public Health.* Apr 2012;33:41-58.

15. Fiscella K, Tancredi D, Franks P. Adding socioeconomic status to Framingham scoring to reduce disparities in coronary risk assessment. *Am. Heart J.* Jun 2009;157(6):988-994.

16. Fiscella K, Tancredi D. Socioeconomic status and coronary heart disease risk prediction. *JAMA.* Dec 10 2008;300(22):2666-2668.

17. Pollack CE, Slaughter ME, Griffin BA, Dubowitz T, Bird CE. Neighborhood socioeconomic status and coronary heart disease risk prediction in a nationally representative sample. *Public Health.* Oct 2012;126(10):827-835.

18. 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.

19. Oakes JM. The (mis)estimation of neighborhood effects: causal inference for a practicable social epidemiology. *Soc. Sci. Med.* May 2004;58(10):1929-1952.

20. Cummins S, Flint E, Matthews SA. New neighborhood grocery store increased awareness of food access but did not alter dietary habits or obesity. *Health Aff. (Millwood).* Feb 2014;33(2):283-291.

21. Schoeni RF, Foundation RS. *Making Americans Healthier: Social and Economic Policy as Health Policy*: Russell Sage Foundation; 2008.

22. Moreland-Russell S, Eyler A, Barbero C, Hipp JA, Walsh H. Diffusion of Complete Streets policies Across US communities. *J. Public Health Manag. Pract.* May-Jun 2013;19(3 Suppl 1):S89-96.

23. Cole K, McNees M, Kinney K, Fisher K, Krieger JW. Increasing access to farmers markets for beneficiaries of nutrition assistance: evaluation of the farmers market access project. *Preventing chronic disease.* 2013;10:E168.

24. Geller A, Fagan J, Tyler T, Link BG. Aggressive policing and the mental health of young urban men. *Am. J. Public Health.* Dec 2014;104(12):2321-2327.

25. Rhodes SD, Mann L, Siman FM, et al. The impact of local immigration enforcement policies on the health of immigrant hispanics/latinos in the United States. *Am. J. Public Health.* Feb 2015;105(2):329-337.

26. Stone NJ, Robinson JG, Lichtenstein AH, et al. 2013 ACC/AHA guideline on the treatment of blood cholesterol to reduce atherosclerotic cardiovascular risk in adults: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines. *J. Am. Coll. Cardiol.* Jul 1 2014;63(25 Pt B):2889-2934.