Journal of Rural Social Sciences

Volume 37
Issue 1 *Rural Population Health and Aging*

Article 4

February 2022

Exploring Rural-Urban Differences in Polygenic Associations for Health among Older Adults in the United States

Trent Davidson
University of Colorado, Boulder, trda6725@colorado.edu

Jason D. Boardman
University of Colorado at Boulder, boardman@colorado.edu

Lori M. Hunter
University of Colorado at Boulder, lori.hunter@colorado.edu

Exploring Rural-Urban Differences in Polygenic Associations for Health among Older Adults in the United States

Cover Page Footnote

Please address all correspondence to Trent Davidson (trda6725@colorado.edu).

Exploring Rural-Urban Differences in Polygenic Associations for Health among Older Adults in the United States

Trent Davidson
Jason D. Boardman
Lori M. Hunter
University of Colorado Boulder

ABSTRACT
A complex combination of genes and environment influence health and, as a result, both genes and environment can play a role in shaping health disparities. We consider distinctions in these influences across rural and urban settings, expanding upon work that shows lower genetic associations in rural compared to urban places by studying an older age group and examining more than the typical outcomes of alcohol/substance abuse. Using a sample of 14,994 adults from the 1992 through 2016 waves of the Health and Retirement Study, our results suggest genetic associations for BMI and heart conditions are significantly lower in rural compared to urban settings. We do not find evidence in support of this association for depression and smoking. In sum, the results suggest the gene-environment interaction may play a role in the well-documented disparities across rural and urban places within the United States, further highlighting the importance of the social, economic, and built environments for individual health.

KEYWORDS
Gene-environment interaction, Health and Retirement Study, health disparities, health outcomes, rural-urban

INTRODUCTION

Disparities in health behaviors, morbidities, and mortality pervade contemporary societies (Adler and Newman 2002; Dubay and Lebrun 2012; Singer 2012). These disparities have been examined in terms of

their socioeconomic, racial, and/or sex-gender delineations (e.g., Marmot 2005; Goosby & Heidbrink 2013; Krieger et al. 2003), and recent work has used individual and population-level genetic variation to provide important insights into the social determinants of health (Kashyap et al. 2015). In this article, we apply a Gene-Environment (GxE) framework to understanding health disparities in terms of the rural-urban divide in the United States. In doing so, we highlight the role of context in health while also challenging the oversimplified "nature versus nurture" framework that has structured much of the framing of the work on health outcomes and disparities for the past century. Specifically, *we examine whether health disparities in obesity, depression, cigarette smoking, and heart conditions between rural and urban residents are enhanced or reduced among individuals according to genetic risk.*

Overall, rural dwellers in the U.S. tend to have poorer health than their urban counterparts. For example, some research suggests rural residents are less likely to engage in a healthy lifestyle including regular physical activity, adequate sleep duration, and low to moderate rates of alcohol consumption (Matthews et al. 2017). Age-adjusted death rates among rural residents are higher for many causes (ca. 1999-2014; Moy et al. 2017) which combine to yield a "rural mortality penalty" that has been widening since the mid-1960s (Cosby et al. 2019; James 2014). By emphasizing rural-urban differentials in health, our work responds to the call for GxE scholarship to represent the full range of contexts in which Americans live (Boardman, Daw, and Freese 2013). Given the non-trivial representation of Americans living in rural areas—about 14 percent as of July 2019 (Cromartie et al. 2020)—understanding variation in rural-urban genetic risk for specific health problems is essential for a comprehensive understanding of rural health determinants. Here, we focus on cigarette smoking, obesity, depression, and heart conditions, all health issues that show disparities across the rural-urban divide. Further, emphasis on these issues expands prior GxE work in rural settings which has exclusively focused on alcohol and substance abuse, and younger age groups. Importantly, the rural United States is aging more rapidly relative to the rest of the country (Jensen et al. 2020; Sparks 2012). This in mind, the most urgent healthcare challenges faced by rural America in the near future will likely be most apparent for older rural residents, illustrating a need to understand the underlying health dynamics for this group in particular. Our work also captures key dimensions of health, namely physical health (heart conditions and obesity), mental health (depression),

and health behaviors (cigarette smoking and obesity), and introduces the most up-to-date methods in GxE research.

We might expect rural-urban health disparities as a function of genetic risk across these outcomes for a few reasons. First, prior, albeit limited, work has found that genes *do interact* with rural settings to influence complex health and behavioral outcomes (Davis, Natta, and Slutske 2017; Legrand et al. 2008; Rose et al. 2001; Taylor et al. 2011). Thus, there is a well-established rationale for studying rural health from a GxE perspective. Second, as our work captures the above-mentioned multidimensional nature of health (i.e., physical, mental, and behavior) this may lend insight to the relevant environmental mechanisms. For instance, a higher prevalence of obesity in rural places as a function of genetic risk could be driven by the phenomenon of food deserts (Whitley 2013). Likewise, a higher prevalence of heart conditions given genetic propensity for this outcome could be spurred by the problematic closing of rural hospitals and physical health services (Kaufman et al. 2016). Finally, many health problems are driven by stress exposure in that stress can "get under the skin" (McEwen 2012; Shields 2017). In other words, the adversity of a given environment goes beyond merely being correlated with poor health outcomes in that it may ***cause*** physical and psychosocial ailments through physiological, neurological, genetic, and epigenetic mechanisms (Galea, Uddin, and Koenen 2011). For example, prior work documents greater allostatic load (i.e., the cumulative wear and tear on the body due to adapting to adverse physical or psychosocial situations, see Geronimus et al. 2006) among impoverished children in rural compared to urban places (Evans et al. 2012). Further, stress exposure in rural places—both the source and frequency—may differ substantially from such exposure in urban areas (Dobis et al. 2020). Taken together, this confluence of factors motivates the current study.

BACKGROUND

Rural and Urban Health Disparities

Addressing each of our outcomes of interest in turn, recent CDC data suggest that 28.5 percent of adults in rural areas smoke regularly compared to 25.1 percent in urban areas and 18.3 percent in large metro areas (Center for Behavioral Health Statistics and Quality 2020). Moreover, rural smokers are much more likely to smoke heavily, defined as 15 or more cigarettes per day (U.S. Department of Health and Human Services 2014). This pattern persists across age groups—a national survey finds that in 2014-2016, rural youth were 50 percent more likely to

smoke than their urban peers (Ziller et al. 2019). On a national scale, prior work shows how individuals' genetic predisposition for cigarette smoking has shifted over time. Specifically, a study demonstrated that the first Surgeon General's Report advising against smoking coincides with an increase in the effect of genetic influences on regular smoking, while later legislation prohibiting smoking in public places significantly attenuated these influences (Boardman, Blalock, and Pampel 2010). Put differently, this implies that an environmental shock (i.e., the Report) convinced enough individuals to quit smoking such that those who continued had among the highest genetic risk for smoking, thereby "increasing" the importance of genetic factors in influencing whether one will smoke. With this dynamic in mind, understanding if and how rural residence acts as an environmental influence within a GxE framework could shed light on what is driving rural-urban disparities.

Obesity is another health issue that deeply impacts Americans as a whole but is substantially worse among the rural population. In 1998, 20.4 percent of rural adults compared to 17.8 percent of urban adults were obese (Patterson et al. 2004). By 2008, those figures had grown to 39.6 percent and 33.4 percent, respectively (Befort, Nazir, and Perri 2012), further revealing a widening of the rural-urban disparity. More recent data, from 2013-16, show continued increases with obesity among rural and urban dwellers reaching 43.1 percent and 35.1 percent, respectively (Hales et al. 2018). As with cigarette smoking, a large body of GxE work has examined obesity. For instance, research by Guo et al. (2015) found the influence of the genome on obesity is significantly influenced by historical period as well as physical activity, suggesting that genetic influences vary as a function of broader context as well as behaviors within an individual's lifetime. Specifically, this work showed the heritability of BMI to be substantially larger after the mid-1980s than in the preceding decades across multiple age groups, and apparently smaller among physically active individuals than those not active. All of this is to say that there is precedent for studying obesity in terms of a GxE framework as well as from a rural-urban health disparities perspective. Ours is the first project to combine the two.

Related to obesity is the comorbidity of heart conditions and, again, rural residents suffer relatively more than their urban counterparts. Specifically, crude prevalence of coronary heart disease is 38.8 percent higher among respondents living in rural areas compared with urban areas (O'Connor and Wellenius 2012). Similarly, to our knowledge, no prior work

has applied a GxE understanding to heart conditions in rural contexts. Thus, our work makes important inroads in this area.

Finally, GxE literature has also linked depression to social environments. One of the pioneering papers of GxE research with respect to stress found that individuals possessing one variant of a specific gene suffered the fewest depressive symptoms in salubrious (i.e., non-stressful) environments but by far the most severe depression in more stressful environments (Caspi et al. 2003). Although this particular study has a mixed replication history, the broader perspective—that certain genotypes are differentially susceptible to some outcomes (e.g., depression) as a function of their surrounding environment—is a cornerstone of the GxE framework. More recent work has accordingly found that the effects of genetic risk for depression are significantly moderated by overall environmental stress exposure (Gonda et al. 2018). This is important to rural-urban health disparities because stress exposure in rural areas—both the source and frequency—may differ substantially from such exposure in urban areas (Dobis et al. 2020). As examples, some rural areas have less diversified local economies, riskier occupations (e.g., mining), more restricted access to healthcare, lower social capital, and higher rates of mortality. Coping strategies could also vary. Potentially related, residents of rural areas suffer from higher rates of drug and alcohol abuse, as well as higher rates of self-harm and interpersonal violence (Dwyer-Lindgren et al. 2016).

Gene-Environment Interactions and Health

As suggested by the above studies, GXE research on social phenomena has most typically been concerned with explaining the “nurture” aspect of health outcomes. With the increasing availability of physiological and genetic markers, “nature” can be added to investigations to shed light on the interactions between genetics and human environments as they shape health behaviors and outcomes. In the context of this study, the GxE framework anticipates that rural-urban health disparities may be related to different, place-related, cumulative genetic risk for a specific morbidity. Thus, we contend that if context (i.e., rural residence) is not adequately considered, what may appear to be a purely biological or purely social process could in fact be, and likely is, a complex relationship between the two. The importance of introducing a GxE understanding to this body of literature, then, is that it can elucidate social-contextual effects likely to be misunderstood or altogether overlooked if examining behaviors from “nature” or “nurture” perspectives alone.

Potential Rural-Urban Health Disparities in the GxE Framework

We propose four different hypothetical associations to illustrate the possible outcomes of a statistical interaction analysis as it pertains to GxE where E is conceptualized as rural residence (see Figure 1). GxE models in Figure 1 are unique from the existing GxE typology because they are designed specifically to emphasize existing rural-urban health disparities. Broadly, these hypothetical models illustrate the possible outcomes of a statistical interaction analysis as it pertains to GxE where E is conceptualized as rural residence. Descriptions specific to each hypothesized model are detailed below. Namely, we identify four potential outcomes: a "Constant Disparity," "Reduced Disparity," "Emergent Disparity," and "Crossover." The null hypothesis is described by the "Constant Disparity" model in the upper left corner of Figure 1. Here, residents of rural areas (dark lines) have a higher prevalence of a particular morbidity compared to urban residents but this disparity is *consistent across levels of genetic risk* (i.e. the values on the x-axis). In this model, genetic risk may certainly affect the morbidity, but the magnitude of the genetic association is similar for rural and urban residents and, therefore, bringing genotype to bear does little to shed light on existing disparities in health outcomes.

The "Reduced Disparity" model suggests that existing health disparities are only evident among those with the lowest overall genetic risk but among those with high genetic risk, there is no health risk associated with area of residence. In this case, the association between genotype and phenotype is *weaker* among rural residents. This model is important because it suggests that genetic sensitivity is not something that is enhanced or triggered among rural residents. Indeed, quite the opposite, among those with similarly low levels of genetic risk, large differences in health emerge which point to factors unique to rural settings that may be linked to this specific health outcome. The opposite is shown in the "Emergent Disparity" model in which there is only a rural-urban disparity among those with the highest genetic risk. This focuses on individual-level (i.e., genetic) differences as a determinant of larger rural-urban health disparities. Thus, inclusion of both the Reduced and Emergent Disparity models captures the possibility that there is less heterogeneity among rural than urban residents in the relevant variables. In other words, given more similar characteristics within rural areas, the effects of genotype on our outcomes could be either under- (Reduced) or over- (Emergent) determined. More specifically, the effects of a rural environment on one's health could be so uniform that in one instance, they

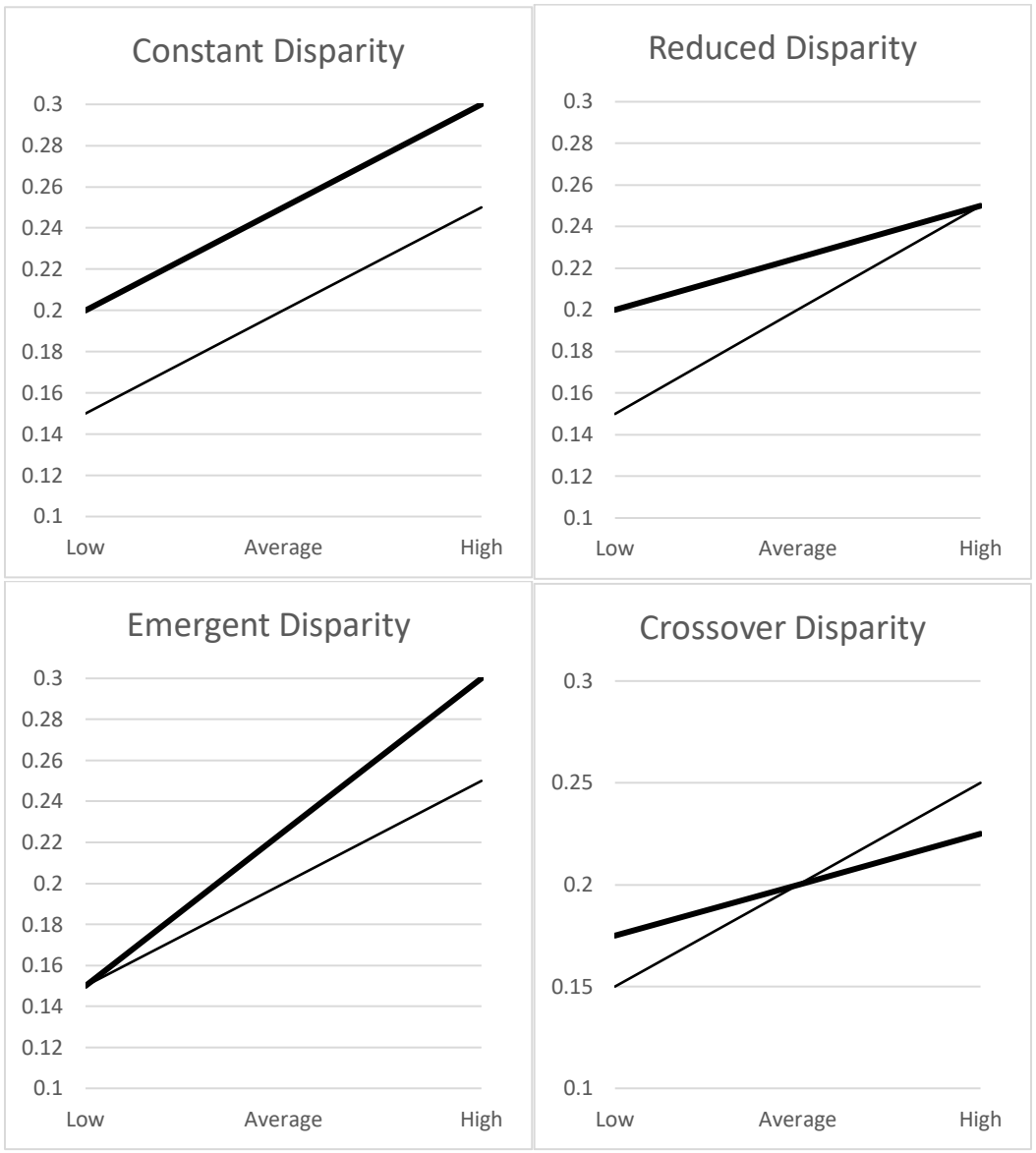
may overpower what protective individual factors one brings to the table, or they may allow for the genetic signal to be artificially strong relative to the noise. Separate work (Domingue et al. 2015) has shown this dynamic with respect to academic performance among Black versus White students (e.g., the "environmental" effects of structural racism suppress genetically oriented academic talent, leading to lower coefficients for genetic markers for educational attainment among Black students). In the context of rural-urban residence, this may manifest as rural places being so similar for individuals that their genetic propensities are "washed out," or alternatively, that this rural similarity leaves room for only genetics to affect certain health outcomes.

Finally, the "Crossover Disparity" model is similar to the "Reduced Disparity" model in that both anticipate that genetic associations will be weaker in magnitude among rural residents because of environmental differences, but the models differ with respect to the intercept. Accordingly, the overall risk of a specific morbidity among rural residents will be enhanced among those with the lowest genetic risk but reduced among those with the highest genetic risk. This is akin to the "strongest version" of the GxE framework, wherein the effects of the genotype are entirely contingent upon environment, or vice versa. We include this specification, then, in keeping with the history of GxE literature. Specifically, see the above description of Caspi and colleagues' (2003) study.

To our knowledge, only a handful of studies have applied a GxE approach to understanding rural health (Davis et al. 2017; Legrand et al. 2008; Rose et al. 2001; Taylor et al. 2011). Broadly, this body of research has found that genes *do interact* with rural settings to influence complex health and behavioral outcomes, such as obesity, drinking behavior, and substance use. Further, these studies indicate that rural environment may attenuate the heritable components of these outcomes by "overshadowing" the extent to which the genetic effects are expressed.

Our work expands upon this collection of research in three key ways. First, extant studies focus exclusively on alcohol use. While alcohol poses health challenges for rural America, it is far from the only issue. Second, of the four studies reviewed above, three (Davis et al. 2017; Legrand et al. 2008; Rose et al. 2001) derived their estimates from twin study design while Taylor et al. (2011) use a candidate gene approach. To be clear, we are not claiming that twin and candidate gene studies are inherently limited. Rather, we seek to expand upon these lines of inquiry by introducing genome-wide measures to the GxE study of rural health

Figure 1: Hypothetical Models of Rural Health Disparities as a Function of Genetic Risk



Note: The thicker lines represent residents of rural areas, and the thinner lines are residents of urban areas. The values on the x-axis denote low, average, and high levels of genetic risk. The values on the y-axis capture the probability of a particular morbidity or risky health behavior.

disparities. Candidate gene studies emphasize only one or a few genes whereas health outcomes are multifactorial, etiologically complex, and often influenced by multiple (sometimes hundreds of) genes (Lambert et al. 2021). Moreover, candidate gene work may yield biased estimates as a function of publication bias of novel work; publication bias of positive replication attempts; and insufficient statistical power, suggesting that a

substantial portion of candidate GxE work may represent Type 1 errors (Duncan and Keller 2011). Regarding twin studies, although historically a standard of GxE work, genome-wide association studies (GWAS) have the advantage of measuring the genome directly, rather than inferring genetic effects. Further, genome-wide studies have the advantage of being hypothesis-free and can assuage some of the strong assumptions classic twin studies make, including gene-environment correlation (which we address later in the Discussion), additive genetic components, and equal environments. Thus, in their search for relevant polymorphisms, GWAS may represent a more accurate picture of genetic influences on various outcomes.

Finally, the aforementioned studies have focused largely on adolescents, yet the rural United States is aging more rapidly relative to the rest of the country (Jensen et al. 2020; Sparks 2012). Thus, the most pressing healthcare challenges faced by rural America in the near-term will likely fall most heavily on older rural residents, illustrating a need to understand the underlying health dynamics for this group in particular.

Research Objectives

The present research addresses the limitations described above by (1) examining four different health indicators, (2) using state of the art polygenic scores (PGS) to estimate genetic influences (Ware et al. 2017; HRS Staff; HRS), and (3) extending the age range to older adults who have had greater exposure to different environments over time. Overall, the objective of this article is to apply a gene-environment framework to add an important dimension to understanding rural health in the United States. Specifically, *we examine whether health disparities in obesity, depression, cigarette smoking, and heart conditions between rural and urban residents are enhanced or reduced among individuals according to genetic risk.*

DATA AND METHODS

Sample

We use a sample of 14,994 older (i.e., 50+) adults in the HRS for the years 1994-2016 for whom genetic data were also available (see Supplemental Figure 3 for Study Flow Diagram). The HRS is sponsored by the National Institute on Aging (grant number NIA U01AG009740), focuses on older adults, and is conducted by the University of Michigan. Across all waves, the baseline response rate is 81.3 percent, and we use data from the 2016 RAND Longitudinal File (V2) for all analyses. Beyond

the biological issues faced by older adults, these individuals have also been exposed to a greater variety of environments and, consequently, stressors. Descriptive statistics are provided in Table 1.

Outcomes

We use four measures of health that are available for all waves of the HRS (the only exception is depression for which no data were available for the first wave). These four measures include: 1) body mass index (BMI); 2) depression; 3) current smoking; and 4) heart condition. BMI is calculated as self-reported weight divided by the square of self-reported height. Height is converted into meters and weight into kilograms. Beginning in Wave 3, height is only asked of new respondents, but weight is asked in every wave. For respondents being re-interviewed, height is carried forward from their first interview. Depression is measured with the eight-item Center for Epidemiologic Studies Depression (CESD) scale. The CESD score is the sum of six "negative" indicators minus two "positive" indicators. The negative indicators measure whether the respondent experienced the following sentiments all or most of the time: depression, everything is an effort, sleep is restless, felt alone, felt sad, and could not get going. The positive indicators measure whether the respondent felt happy and enjoyed life, all or most of the time. We use a threshold of 3 or higher to classify respondents as depressed (Kessler et al. 1994). Current smoking is measured by response to the question, “Do you smoke cigarettes now?” and was asked of all respondents at each wave. Finally, the presence of a heart condition is assessed with two questions. At the respondent’s first interview, they are asked, “Has a doctor ever told you that you have a heart condition?” and in subsequent interviews they are asked, “Since we last talked to you, that is since [last interview date], has a doctor told you that you have a heart condition?” Not all respondents who responded “yes” in an earlier wave respond “yes” in a later wave. Thus, there is intra-person variation on this variable over time.

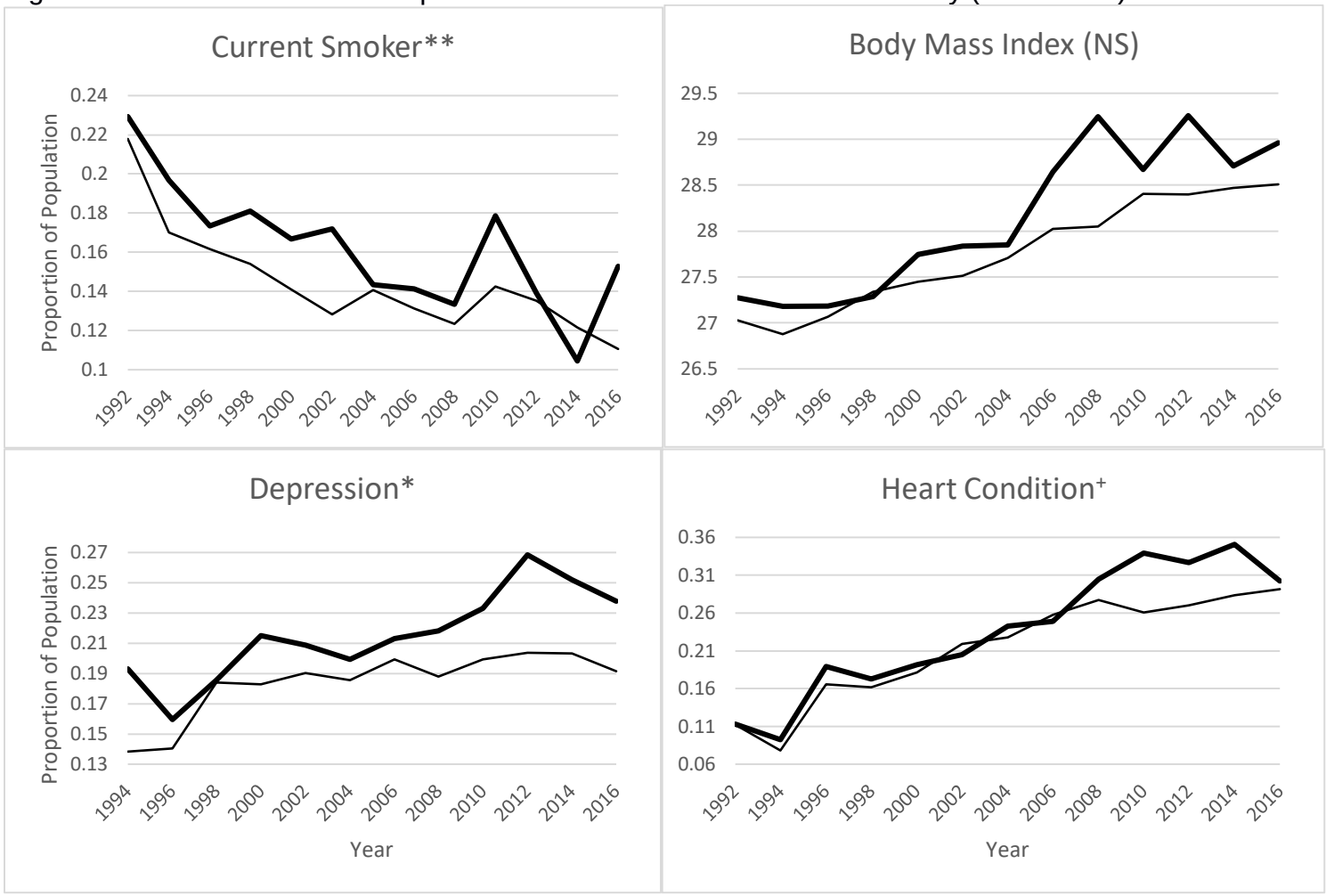
Figure 2 presents the prevalence of smoking, depression, heart conditions, and average BMI for rural and urban respondents in the Health and Retirement Study for the 13 waves of data collection from 1992-2016. The p-value statistical significance tests are of the rural-urban disparity across the full range of years, not differences across waves. Again, the goal of this figure is to demonstrate the general patterns with respect to health differences across the four domains in the two different contexts using HRS data.

Table 1: Descriptive Statistics for all Data Used in the Analyses: Health and Retirement Study (1992-2016)

Wave	Age	Female	Black	Educ	Rural	Heart	Depression	BMI	Smoke	N
1992	56.174	0.540	0.155	12.690	0.114	0.113	.	27.057	0.219	4,973
1994	60.875	0.588	0.143	12.615	0.092	0.079	0.143	26.906	0.173	6,343
1996	62.594	0.595	0.141	12.626	0.071	0.167	0.142	27.074	0.162	6,471
1998	63.557	0.584	0.139	12.756	0.060	0.162	0.184	27.334	0.156	8,954
2000	65.304	0.589	0.136	12.774	0.052	0.182	0.184	27.463	0.142	9,115
2002	67.080	0.589	0.138	12.782	0.040	0.219	0.191	27.529	0.130	9,369
2004	66.626	0.580	0.149	12.929	0.025	0.228	0.186	27.710	0.141	11,174
2006	68.425	0.583	0.147	12.935	0.016	0.258	0.199	28.031	0.131	11,439
2008	70.007	0.586	0.148	12.968	0.015	0.278	0.188	28.068	0.123	11,341
2010	67.940	0.579	0.204	13.112	0.013	0.262	0.199	28.399	0.143	13,260
2012	69.172	0.584	0.206	13.163	0.012	0.271	0.204	28.404	0.135	12,650
2014	70.374	0.595	0.209	13.223	0.012	0.285	0.204	28.463	0.121	11,531
2016	71.364	0.599	0.215	13.293	0.021	0.292	0.192	28.509	0.111	10,143
Total	67.021	0.585	0.168	12.957	0.034	0.229	0.189	27.889	0.140	
N (Obs)	126,763	126,763	126,763	126,763	126,763	126,699	118,092	125,556	126,123	126,763
N (Ind)	14,994	14,994	14,994	14,994	14,994	14,994	14,992	14,989	14,986	14,994

Note: Cell entries are the means for continuous variable and the proportion for binary variables. Depression was not measured in 1992.

Figure 2: Rural-Urban Health Disparities in the Health and Retirement Study (1992-2016)



Note: Thicker lines represent residents of rural counties and thinner lines are suburban and urban residents. All data from the 1992-2016 Health and Retirement Study (N= 15,306). NS Non-Significant, $^{+}p< .1$ *p< .05 **p< .01

Predictors

Genetic data. Genetic data for the HRS are based on samples collected in two phases. First, buccal swabs were taken in 2006 using the Qiagen Autopure method. Second, saliva samples were collected in 2008 and extended with Oragene. Genotype calls were then made with the Illumina HumanOmni2.5-4v1 array. Construction of various PGS for the HRS dataset is described in detail elsewhere (Ware et al. 2017; see especially page 10). As a brief primer, PGS derived from genome-wide association study (GWAS) summary statistics present a method by which complex health outcomes may be studied genetically. Others helpfully describe PGS as "a summary measure of a set of risk-associated genetic variants….[They] provide a quantitative measure of genetic predisposition that is calculated using information from multiple genetic variants" (Belsky and Israel 2014). Variation unexplained by the PGS is then attributed to social-behavioral factors, indirect genetic effects, and random error. While not free of controversy (e.g., PGS are assumed to be an additive measure of genetic effects, which contradicts understandings of [gene-gene interaction], see Belsky and Israel 2014), PGS provide the best current indicators of genetic risk across a variety of studies (Domingue et al. 2014). Crucially, PGS can be standardized to show gradations in individual genetic risk for a given outcome in terms of standard deviations.

To illustrate further, PGS are summative measures of the individual effects of genetic variants, or single nucleotide polymorphisms (SNPs). PGS, thus, are computed by first genotyping an individual at the location of a given SNP in the genetic sequence. For example, suppose a particular gene is called "GA." One of the SNPs will be declared the reference allele. An allele is simply a specific version of a genetic variant. In this example, it does not matter which is declared the reference, as there is one G and one A. Consequently, the individual's genotype for this SNP will be a value of one. If the reference allele is G and a person's genotype is GG then their genotype will be a value of 2. This value is set as the number of G alleles at this particular SNP associated with a given outcome. Thus, if their genotype were AA, they would be assigned a value of 0. For the sake of simplicity, in this instance, we will say that the outcome is height. If the genetic effect size is computed to be $\beta = 0.4045$, then a one-unit increase in the number of G alleles increases an individual's height by 0.4045 inches. Subsequently, this process is iterated for every SNP in the genome (2.5 million regressions), thus estimating the

effect size of each SNP on that outcome. We then multiply that effect by the individual's genotype and sum these values, yielding the following:

$PGS_{ij} = \Sigma(gt_{ij} * b_j)$

Finally, the distribution of the PGS is standardized such that $\mu_{pgs} = 0, \sigma_{pgs} = 1$. We use PGS estimates that have been constructed by others to assess genetic risk for smoking initiation (ever/never) (Furberg et al. 2010), BMI (Locke et al. 2015), myocardial infarction (Nikpay et al. 2015), and depressive symptoms (Okbay et al. 2016).

An important consideration in the construction and use of PGS is the role of population stratification. This refers to small differences in allele frequencies that align with socially constructed racial and ethnic groups. In turn, these differences create spurious associations that are due to social-environmental differences between groups; in other words, genetic differences that appear to exist have no true causal association on the phenotype, per se. Accordingly, the HRS released the PGS estimates among individuals with European Ancestry and African Ancestry separately. Thus, the PGS values are standardized within each group and the principal components capture ethnic variation within each group separately. In this manner, it is not possible that population stratification can be driving the results, as the mean difference in PGS values are non-existent by design. In our models, we evaluate Black and White respondents together for three reasons: (1) substantively, we believe that the continued separation of Black and White respondents when examining genetic associations is a problematic practice that Duster (1990) predicted over 30 years ago in *Backdoor to Eugenics* and should not be the norm; (2) theoretically, we do not have any *a priori* reasons to believe that the genetic associations by rural status should differ as a function of one's racial identity and racialized experience; and (3) methodologically, we are not concerned with one specific causal biological pathway but rather an overall indicator of genetic associations akin to a narrowly sensed additive genetic variance component. To further assuage concerns regarding population stratification, we also evaluated our models separately by self-identified racial identity (Supplemental Table 1). In support of our decision to estimate models jointly, the direction and magnitude of the stratified parameter estimates were nearly identical for both groups. Changes in sample size, however, caused increases in the standard errors for Black respondents and an increase in the p-value above the traditional 0.05 level. Importantly, this does not compromise our logic for estimating both groups within one model, as it is a function of sample size, not underlying "true" differences between Black and White respondents. As such, the

results from these analyses further confirmed our decision to include all genotyped respondents in our analyses.

A final consideration is that the respondents in the HRS can only be in our study if they survived long enough to be included in the genetic sample which first occurred in 2006. Domingue et al. (2017) provided evidence that participation in the genotyping sample was not random with respect to health and sociodemographic background and more importantly, that genetic association results remained robust to this form of selection. To further ensure that our sample was not affected by this we performed analyses in which we regressed the number of observations that each individual contributed to our study (see Table 2) on their minimum age, their PGS, their average rural score (i.e., if someone resided in an urban area for all waves they would have a zero compared to one who resided in an urban area one-half of the time who would receive a .5), and an interaction between the two. The association between the PGS and the number of observations would provide evidence for the selection of cases based on health related to genotype, and the interaction would indicate if this selection were different for rural or urban residents. We found no evidence for either form of selection for our four health outcomes.

Rural-urban measures. Next, to examine the rural-urban divide, we employ the Rural-Urban Continuum Codes (RUCC) developed by the U.S. Department of Agriculture's (USDA) Economic Research Service and available in the Cross-Wave Geographic Information data file from the 1992-2016 HRS. Briefly, the continuum indicates the rural/urban nature of a county using a nine-point scale, with higher values representing more rural as determined by the population size of places within each county and the proximity to metropolitan centers of certain population thresholds. For instance, the "most rural" value of nine is assigned to counties that have no places with a population over 2,500 and are not adjacent to a metro area. Counties with a value of one reflect those with metro areas of at least one million residents. In the HRS, to protect confidentiality, the RUCC scale is truncated from its original nine categories into a binary variable where 0=*urban county-residing* and 1=*rural county-residing*. Rural counties are those with urban populations not exceeding 50,000. The proportion of rural dwelling respondents has declined over time (see Table 1) and overall, 3.4 percent of our sample is considered to live in a rural area.

Covariates

In all analyses, we control for respondent's age, sex, years of education, and race. Appropriately, each of these has been previously shown to be associated with the outcomes of interest, i.e., obesity (Cossrow and Falkner 2004; Lovejoy and Sainsbury 2009; Masters et al. 2013; Ogden et al. 2018), cigarette smoking (Boardman et al. 2010; Pampel 2006; Piper et al. 2010), heart conditions (Afzal et al. 1999; Cutler and Lleras-Muney 2010; Milner et al. 2004), and depression (Bailey, Mokonogho, and Kumar 2019; Barnes, Keyes, and Bates 2013; Eid, Gobinath, and Galea 2019; Kessler et al. 2010). Age and years of education are treated as continuous variables, while sex and race are dichotomized with female and Black as the respective reference groups. All models also control for the top five principal components of genetic ancestry to rule out the possibility that our results are confounded by population stratification (Price et al. 2006). Across all analyses, the only time-varying covariates are age and rural-urban residence.

Statistical Analysis

Because our data contain observations nested within individuals over time, we use a generalized linear and mixed modeling approach with the *mixed effects* suite of models in Stata 16. These are flexible models that allow for the analysis of continuous (e.g., BMI) and binary (e.g., depression, current smoker, heart condition) variables that are nested within individuals over time. In Table 2, the first and second columns describe the number of observations, and the third and fourth the number of individuals. For instance, there are 36,387 observations with 13 waves of data with 2,799 people having full data across all waves of the study. Only 151 people were observed just once. The median number of observations per person is nine and the most common number of observations is 13. The mixed effects models also allow for the use of sampling weights to adjust for the complex design effects inherent in the HRS.

RESULTS

Table 3 presents the results of the multilevel regression models described above. These models nest observations within individuals so that rural location is used to assess health and health behaviors contemporaneously. The first row presents the main effect of the PGS on its respective trait, while the interaction term indicates the differential effect of the PGS among residents of rural, relative to urban, counties. In all four

Table 2: Sample Sizes by Number of Observations in the Longitudinal HRS Data

Number of Waves	N (Obs)	%.	N (Ind)	%
1	151	0.12	151	1.01
2	736	0.58	368	2.45
3	1,692	1.33	564	3.76
4	9,472	7.47	2,368	15.79
5	1,945	1.53	389	2.59
6	3,516	2.77	586	3.91
7	11,914	9.40	1,702	11.35
8	7,376	5.82	922	6.15
9	9,612	7.58	1,068	7.12
10	20,470	16.15	2,047	13.65
11	9,548	7.53	868	5.79
12	13,944	11.00	1,162	7.75
13	36,387	28.70	2,799	18.67
Total	126,763	100.00	14,994	100.00

cases, a strong and statistically significant association emerges between the health indicator and its corresponding polygenic score.

In two of the four cases (BMI and heart condition) we observe moderate to strong, negative, statistically significant interactions between PGS and rural county, suggesting that polygenic risk is more weakly associated with its corresponding phenotype among residents of rural compared to urban areas.

While the effect of the smoking PGS ($b = .385$, $p<.001$ was strong and statistically significant, it does not differ across rural and urban residents as evidenced by the interaction term's lack of statistical significance, likewise for depression. For depression specifically, the observed non-significance of the interaction term may depend on sample composition of the HRS (see the Discussion section for further detail). Figure 3 presents plots of the interactions to better contrast them with our conceptual models. Heart condition is best summarized by the "Reduced Disparity," which suggests that rural-urban health disparities are most evident among those with the weakest genetic risk for heart conditions. In other words, rural environments do *not* trigger otherwise latent genetic risks for these health outcomes. In fact, the opposite is the case: factors of the built, social, or cultural rural environment may be the most salient risk for this morbidity.

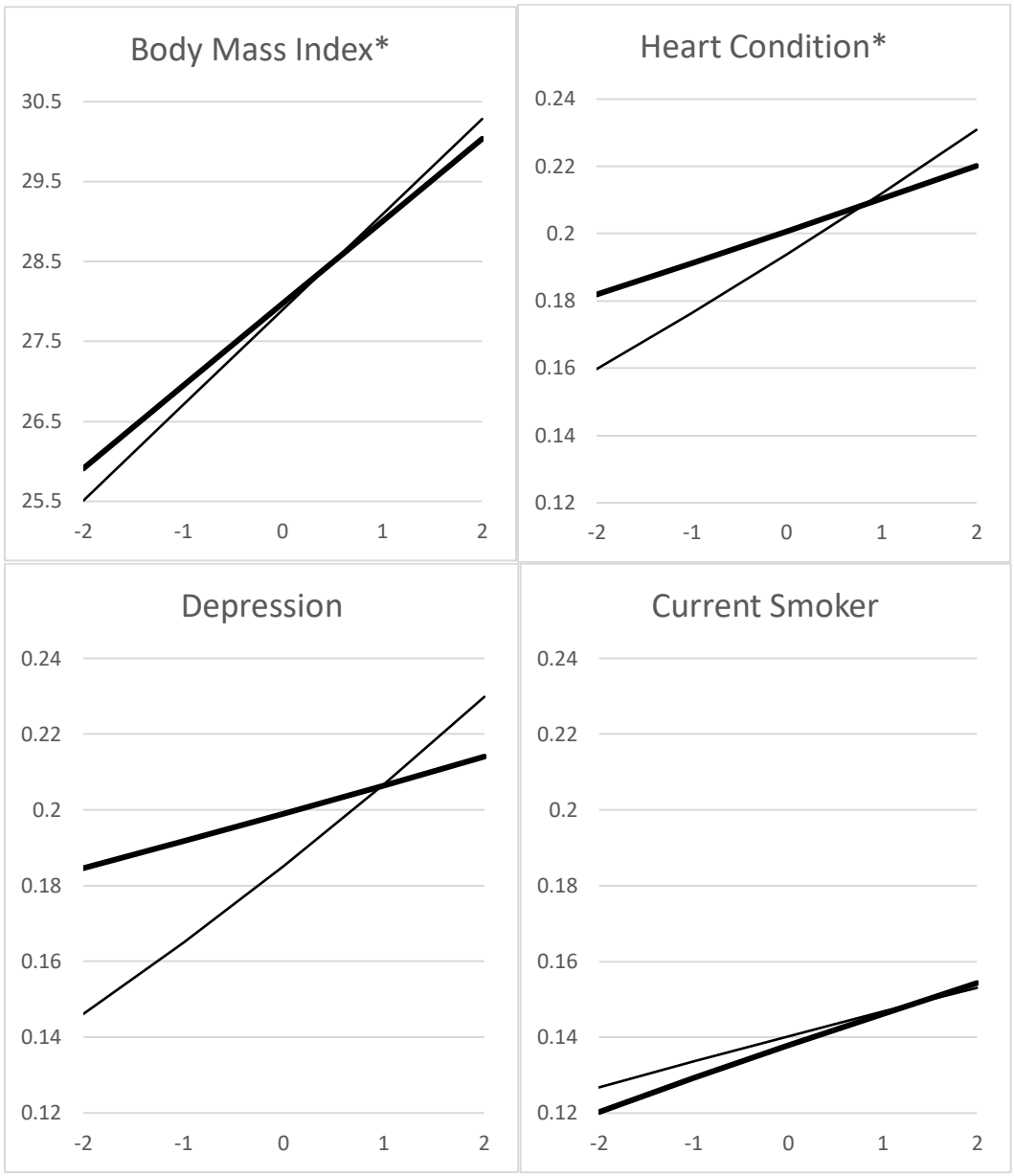
Table 3: Multilevel Regression Estimates: Polygenic Risk by Rural Residence

	BMI		**Heart Condition**		**Depression**		**Current Smoker**	
	b	p	b	p	b	p	b	p
Intercept	35.167	.000	-18.298	.000	-0.036	.848	11.516	.000
PGS	1.241	.000	0.519	.000	0.231	.000	0.385	.000
Rural	-0.160	.042	0.197	.095	0.150	.093	-0.139	.325
PGS*Rural	**-0.185**	**.018**	**-0.244**	**.023**	**-0.153**	**.065**	**0.108**	**.461**
Age (years)	-0.100	.000	0.206	.000	-0.005	.033	-0.236	.000
Female	0.005	.954	-1.082	.000	0.708	.000	-0.611	.000
Black	1.672	.000	-0.052	.731	0.672	.000	0.993	.000
Education (yrs)	-0.138	.000	-0.236	.000	-0.230	.000	-0.565	.000
Wave	0.248	.000	0.226	.000	0.038	.000	-0.001	.976
σ_u^2	26.991		7.119		4.859		10.023	
σ_e^2	4.255		-		-		-	
N (Obs)	119,651		125,473		113,250		125,958	
N (Ind)	14,882		14,895		14,884		14,888	

Note: All data come from the 1994-2016 Health and Retirement Study (HRS). Parameter estimates are unstandardized regression estimates from the model described above. All models include controls for the top 5 principal components of genetic ancestry and are weighted to reflect the design effects of the HRS study. p-values represent two-tailed tests of significance. For brevity, calculated 95% confidence intervals are not shown, but do not change substantive interpretation of the results. The σ^2 values refer to the individual (level-2) and observation (level-1) level error variances, respectively. Coefficients for all variables in the model are unstandardized. While the PGS variables are standardized to the dataset, the reported betas are not.

With regard to BMI, the association is best summarized by "Crossover Disparity" in that there is more evidence of a crossover effect just above the average genetic risk. Specifically, for respondents with low genetic risk for high BMI, those in rural areas are more at risk of this negative health outcome as compared to urban dwellers. In other words, rural contexts may trigger high BMI for those with low genetic risk. But the opposite is true for respondents with greater genetic risk of high BMI. In these cases, rural residents exhibit a lower risk of high BMI as compared to urban. As the triggering "Emergent Disparity" is not evident, this points to features of the built and social environment as the key mechanisms responsible for BMI disparities among rural and urban dwellers.

Figure 3: Observed Genetic Differences in Rural-Urban Health Disparities



Note: Results of Multilevel Regression Models. Thicker lines are rural residents and thinner lines represent urban residents. The values on the x-axis denote standard deviations of genetic risk. The values on the y-axis capture the proportion of a particular morbidity in our sample. For BMI, the y-axis captures measured BMI. *p<.05.

Importantly, all of the above results are robust to the issue of selective mortality. In ancillary analyses, we test whether the effect of a given phenotype (i.e., obesity, heart condition, depression, or cigarette smoking) has a relatively stronger effect on mortality for rural residents.

Thus, we are able to rule out the concern that individuals with one of these conditions had higher mortality risk and were, thus, contributing less data to our sample.

DISCUSSION AND CONCLUSION

In sum, the gene-environment interaction approach provides new insights into the study of rural-urban health disparities in the United States. Specifically, our findings suggest that latent genetic risks for heart conditions are not triggered by rural versus urban residence, although the risk of BMI is associated with genetic predisposition as it interacts with rural-urban residence. Applying the GxE approach allows researchers to consider the notion that health disparities may not be observable among those with the greatest genetic risk for a specific health outcome. In this manner, bringing genetic information to bear shines a light on disparities that may have otherwise been masked among individuals living in rural places.

In the current study, we build upon previous GxE work examining rural health in three primary ways: first, we examine health outcomes beyond alcohol abuse; next, we expand beyond younger age groups to include an elder population, which better reflects the age composition of rural America; finally, we introduce genome-wide research to this area, thus extending prior twin design and candidate gene studies.

The Reduced Disparity models we show highlight the centrality of environmental differences across the two residential contexts which may include access to nutritious food outlets in the case of obesity or access to physical health services related to heart conditions. The phenomena of food deserts and swamps, and the problematic closing of many rural hospitals and medical practices may be some of the effects driving this dynamic (Befort et al. 2012; Kaufman et al. 2016; Whitley 2013).

As to why the GxE effect was non-significant for cigarette smoking and depression, there is room for speculation. To smoking, prior work has shown that heritability of smoking has steadily increased over time (Boardman et al. 2010; Wedow et al. 2018). Indeed, in the present study, we see a strong, significant effect of one’s PGS for smoking on whether they do, in fact, smoke. In this way, environment simply may not matter in predicting cigarette smoking. Regarding depression, the observed non-significance of the GxE term may be a function of sample composition. Specifically, note that 11.4 percent of our sample resided in rural areas in the first wave compared to 2.1 percent in the most recent wave. This rural exodus may account for the fact rural residence itself is non-significant in

predicting depression to say nothing of the interaction term. Relatedly, the rural-urban sample ratio may be insufficient to detect this effect if it does exist. The beta for rural residence on its own has an associated p-value of .09; thus, it is possible, but speculative, that this may reach nominal significance with a larger sample.

We have several concluding comments and discussion points in light of our findings. Specifically, we highlight limitations of the present study and future directions. First, there are compositional differences in rural and urban areas that are important to consider, including age, sex, education, and birth cohort. While we control for these factors, it is nonetheless essential to keep in mind the large sociodemographic differences among urban and rural dwellers in our study and the fact that these are changing over time. Most importantly, we do not explicitly focus on the selection into or out of rural areas as a function of age, race, education, or gender. These factors are independent of genotype but the *effect* of genotype on its related phenotype is often different across sociodemographic groups (Boardman et al. 2013). Thus, while beyond the scope of this article, we encourage future researchers to focus explicitly on these selection processes as they unfold over time. For example, in ancillary analyses, we find that the most recent arrivals to rural areas are the oldest and are likely to be relocated retirees selected out of urban and into rural areas.

This point is particularly important because it is possible that rural and urban residents differ in their average polygenic risk for the health outcomes of interest. This is referred to as gene-environment correlation (rGE) and it is particularly problematic because it can bias GxE parameter estimates (Jaffee and Price 2007). We examined this possibility by comparing the average PGS among rural and urban residents and we find no evidence of rGE with respect to BMI, depression, or smoking behaviors. However, we do find a significant difference in genetic risk for myocardial infarction (b = .098; p<.005) in which rural residents have a higher genetic risk for a heart condition. This association was reduced with controls for self-reported race and the top five principal components (b = *.068; p<.047*) but remained statistically significant. This bears importantly on the point above if, for example, those who select *into* rural areas are also more likely to have genetic risk for heart conditions compared to comparable individuals who do not. To evaluate this possibility, we examined the distribution of the PGS for myocardial infarction among individuals as a function of where they lived when they were 64 and where they lived when they were 70. This was a simple exploratory model and it

should be noted that this drops respondents who did not have a record at either ages because of the timing of the survey. But in our comparison, those who were in an urban setting at both times had an average heart condition PGS of .028, and those in rural settings at both ages had an average PGS of -.069. We observed selection on both rural movers (those who were in urban settings when they were 64 and rural settings when they were 70) who had an average heart condition PGS of .270 *and* the urban movers (from rural to urban settings) who had an average PGS of -.134. That is, those with the highest genetic risk of heart conditions selected *into* rural areas after retirement age *and* those with the lowest genetic risk of heart conditions selected *out of* rural areas after retirement. Clearly, this is an exploratory and descriptive exercise, but it highlights the complexity of the questions that we pose, and we encourage future researchers to evaluate these questions in a more detailed manner.

Third, in examining the main effects of the PGS on their respective trait, we find strong, statistically significant associations across the board. As discussed in our Results section, the interaction term between PGS and rurality represents the effect of the PGS across residents of the different residential categories. For BMI and heart problems, we find significant, negative interactions between PGS and rurality in which *measured genotype is more weakly associated with its corresponding health outcome among residents of rural compared to urban areas*. Of note is that while the PGS effects are attenuated in rural areas, none of the relevant confidence intervals cross zero, suggesting that the strongest version of the GxE model is not supported here. Nevertheless, that these results indicate that genetic associations for BMI and heart problems are all significantly weaker among residents of rural compared to urban environments illustrates an interesting GxE relationship. Ultimately, this information has broader relevance insofar as it can be used to improve health policy by elucidating a major pathway towards health challenges in rural areas. By better comprehending the health environment vis-à-vis how individual-level variation (i.e., genetics) interacts with said environment, collectively, we may make important inroads in addressing rural-urban health disparities. Understanding for whom, when, why, and how health issues arise is in as many ways a social process as a purely biological one. Medical determinants of health do not exist in a vacuum, unaffected by social-behavioral elements. Rather, as we have shown with respect to rural environments, the two categories interact to contribute to further health disparities. Utilizing this information, as mentioned, may prove fruitful for policymakers seeking to address these disparities.

Future research may expand upon this work in several ways. Namely, we relied on the USDA's Rural-Urban Continuum Codes as our measure of rural-urban residence which was further dichotomized by HRS. While historically useful, we understand that other measures may be more suitable especially if our interests are in a more continuous characterization of rurality (Waldorf and Kim 2015). In the case of the present study, we were bound to what was available in the HRS. Nevertheless, we argue that this limitation is not detrimental to our overall results. As discussed, prior GxE work on rural health was largely underdeveloped, thus leaving an ample lacuna for important developmental work in the field. Next, additional work may control for within-ancestry principal components. In the case of the present study, sensitivity analyses (Supplemental Table 1) did not indicate significant differences in model estimates upon inclusion of principal components. Finally, future work may better address selection into rural areas. Given our reliance on a collapsed rural-urban indicator, our work may not reflect variation across the continuum of rural environments. For the reason stated above, we do not believe this compromises the current project, but rather, it may be an important component of the GxE relationship for rural health worth studying further. In all, the current study offers an important step towards better elucidating the multidimensionality of rural-urban health disparities.

Finally, there are two important considerations with respect to our model specification. First, as described above, the social demographic composition of rural and urban settings is very different and while we control for this, we do not specify an interaction between each social demographic indicator and the PGS. We explored this possibility and while the magnitude of the interaction coefficient changed from the initial model to the adjusted model, the interaction remained statistically significant, and the substantive conclusions remained the same. Specifically, the main interaction for BMI in Table 3 reports a slope of -.185 ($p < .018$) and the fully interactive estimate is slope is -.191 ($p < .016$). We also considered an additional socioeconomic factor in which the mean levels are different in rural and urban areas, wealth, that may be responsible for the results. In ancillary analyses, this additional control did not change our substantive findings, but we encourage future researchers to consider other potential confounding or mediating factors for our observed interactions.

It is also important to consider that our study contains repeated observations of individuals from different birth cohorts who are experiencing aging in ways that are unique to a specific historical period

and their specific birth cohort. It was not our goal to evaluate these different models in this article, but it is still important to consider that the composition of rural adults with respect to age and cohort may be changing over historical time. That is, what may appear to be a rural-PGS association may have more to do with who is in the rural context as a function of increasing time. We examined this possibility by (a) removing the control for Wave (our indicator of period), and (b) including an interaction between Wave and PGS. In both cases our results remained nearly identical, and we do not feel as though this form of selection may be driving our results.

In all, this study offers an intriguing first look at differential gene-environment interactions as related to four health outcomes and behaviors across rural and urban environments. The results are important in their suggestion that rural/urban context shapes the influence of genetic risk for BMI, heart conditions, and depression. Such findings should spur additional investigation to inform policy responses to the well-documented disparities in health across rural and urban America.

ACKNOWLEDGEMENTS

We would like to thank the administrative support from the University of Colorado Population Center (CUPC), most especially Marisa Seitz and Anni Magyary.

FUNDING

This research benefited from administrative, research, and/or computing support through the University of Colorado Population Center (CUPC) funded by Eunice Kennedy Shriver National Institute of Child Health & Human Development of the National Institutes of Health (P2CHD066613). The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH, CUPC, or the University of Colorado. Additionally, support for this project was provided by the Interdisciplinary Research Network on Rural Population Health and Aging, which is funded by the National Institute on Aging (grant R24-AG065159).

DISCLOSURE OF INTEREST

The authors declare no conflicts of interest.

REFERENCES

Adler, Nancy E., and Katherine Newman. 2002. “Socioeconomic Disparities In Health: Pathways And Policies.” *Health Affairs* 21(2):60–76. doi: 10.1377/hlthaff.21.2.60.

Afzal, Adnan, Karthik Ananthasubramaniam, Nagaraja Sharma, Qahtan Ai-Malki, Abbas S. Ali, Gordon Jacobsen, and Syed M. Jafri. 1999. “Racial Differences in Patients with Heart Failure.” *Clinical Cardiology* 22(12):791–94. doi: https://doi.org/10.1002/clc.4960221207.

Bailey, Rahn Kennedy, Josephine Mokonogho, and Alok Kumar. 2019. “Racial and Ethnic Differences in Depression: Current Perspectives.” *Neuropsychiatric Disease and Treatment* 15:603–9. doi: 10.2147/NDT.S128584.

Barnes, David M., Katherine M. Keyes, and Lisa M. Bates. 2013. “Racial Differences in Depression in the United States: How Do Subgroup Analyses Inform a Paradox?” *Social Psychiatry and Psychiatric Epidemiology* 48(12):1941–49. doi: 10.1007/s00127-013-0718-7.

Befort, Christie A., Niaman Nazir, and Michael G. Perri. 2012. “Prevalence of Obesity Among Adults From Rural and Urban Areas of the United States: Findings From NHANES (2005-2008).” *The Journal of Rural Health* 28(4):392–97. doi: https://doi.org/10.1111/j.1748-0361.2012.00411.x.

Belsky, Daniel W., and Salomon Israel. 2014. “Integrating Genetics and Social Science: Genetic Risk Scores.” *Biodemography and Social Biology* 60(2):137–55. doi: 10.1080/19485565.2014.946591.

Boardman, Jason D., Casey L. Blalock, and Fred C. Pampel. 2010. “Trends in the Genetic Influences on Smoking.” *Journal of Health and Social Behavior* 51(1):108–23. doi: 10.1177/0022146509361195.

Boardman, Jason D., Jonathan Daw, and Jeremy Freese. 2013. “Defining the Environment in Gene–Environment Research: Lessons From Social Epidemiology.” *American Journal of Public Health* 103(S1):S64–72. doi: 10.2105/AJPH.2013.301355.

Caspi, Avshalom, Karen Sugden, Terrie E. Moffitt, Alan Taylor, Ian W. Craig, HonaLee Harrington, Joseph McClay, Jonathan Mill, Judy Martin, Antony Braithwaite, and Richie Poulton. 2003. “Influence of Life Stress on Depression: Moderation by a Polymorphism in the 5-HTT Gene.” *Science* 301(5631):386–89. doi: 10.1126/science.1083968.

Center for Behavioral Health Statistics and Quality. 2020. Results from the 2019 National Survey on Drug Use and Health: Detailed tables. Rockville, MD: Substance Abuse and Mental Health Services Administration. Retrieved from https://www.samhsa.gov/data/

Cosby, Arthur G., M. Maya McDoom-Echebiri, Wesley James, Hasna Khandekar, Willie Brown, and Heather L. Hanna. 2019. “Growth and Persistence of Place-Based Mortality in the United States: The Rural Mortality Penalty.” *American Journal of Public Health* 109(1):155–62. doi: 10.2105/AJPH.2018.304787.

Cossrow, Nicole, and Bonita Falkner. 2004. “Race/Ethnic Issues in Obesity and Obesity-Related Comorbidities.” *The Journal of Clinical Endocrinology & Metabolism* 89(6):2590–94. doi: 10.1210/jc.2004-0339.

Cromartie, John, Elizabeth A. Dobis, Thomas P. Krumel, Jr., David McGranahan, and John Pender. 2020*. Rural America at a Glance: 2020 Edition*. Economic Information Bulletin 221, December. United States Department of Agriculture, Economic Research Service.

Cutler, David M., and Adriana Lleras-Muney. 2010. “Understanding Differences in Health Behaviors by Education.” *Journal of Health Economics* 29(1):1–28. doi: 10.1016/j.jhealeco.2009.10.003.

Davis, Christal N., Shanaliz S. Natta, and Wendy S. Slutske. 2017. “Moderation of Genetic Influences on Alcohol Involvement by Rural Residency among Adolescents: Results from the 1962 National Merit Twin Study.” *Behavior Genetics* 47(6):587–95. doi: 10.1007/s10519-017-9867-x.

Dobis, Elizabeth A., Heather M. Stephens, Mark Skidmore, and Stephan J. Goetz. 2020. “Explaining the Spatial Variation in American Life Expectancy.” *Social Science & Medicine* 246:112759. doi: 10.1016/j.socscimed.2019.112759.

Domingue, Benjamin W., Daniel W. Belsky, Kathleen Mullan Harris, Andrew Smolen, Matthew B. McQueen, and Jason D. Boardman. 2014. “Polygenic Risk Predicts Obesity in Both White and Black Young Adults.” *PLOS ONE* 9(7):e101596. doi: 10.1371/journal.pone.0101596.

Domingue, Benjamin W., Daniel W. Belsky, Dalton Conley, Kathleen Mullan Harris, and Jason D. Boardman. 2015. “Polygenic Influence on Educational Attainment: New Evidence From the National Longitudinal Study of Adolescent to Adult Health.” *AERA Open* 1(3):2332858415599972. doi: 10.1177/2332858415599972.

Domingue, Benjamin W., Daniel W. Belsky, Amal Harrati, Dalton Conley, David R. Weir, and Jason D. Boardman. 2017. “Mortality Selection in a Genetic Sample and Implications for Association Studies.” *International Journal of Epidemiology* 46(4):1285–94. doi: 10.1093/ije/dyx041.

Dubay, Lisa C., and Lydie A. Lebrun. 2012. “Health, Behavior, and Health Care Disparities: Disentangling the Effects of Income and Race in the United States.” *International Journal of Health Services* 42(4):607–25. doi: 10.2190/HS.42.4.c.

Duncan, Laramie E., and Matthew C. Keller. 2011. “A Critical Review of the First 10 Years of Candidate Gene-by-Environment Interaction Research in Psychiatry.” *American Journal of Psychiatry* 168(10):1041–49. doi: 10.1176/appi.ajp.2011.11020191.

Duster, Troy. 1990. *Backdoor to Eugenics*, 1st Ed. Routledge.

Dwyer-Lindgren, Laura, Amelia Bertozzi-Villa, Rebecca W. Stubbs, Chloe Morozoff, Michael J. Kutz, Chantal Huynh, Ryan M. Barber, Katya A. Shackelford, Johan P. Mackenbach, Frank J. van Lenthe, Abraham D. Flaxman, Mohsen Naghavi, Ali H. Mokdad, and Christopher J. L. Murray. 2016. “US County-Level Trends in Mortality Rates for Major Causes of Death, 1980-2014.” *JAMA* 316(22):2385–2401. doi: 10.1001/jama.2016.13645.

Eid, Rand S., Aarthi R. Gobinath, and Liisa A. M. Galea. 2019. “Sex Differences in Depression: Insights from Clinical and Preclinical Studies.” *Progress in Neurobiology* 176:86–102. doi: 10.1016/j.pneurobio.2019.01.006.

Evans, Gary W., Edith Chen, Gregory Miller, and Teresa Seeman. 2012. “How Poverty Gets under the Skin: A Life Course Perspective.” Pp. 13–36 in *The Oxford Handbook of Poverty and Child Development*, Oxford Library of Psychology. New York: Oxford University Press.

Furberg, Helena, YunJung Kim, Jennifer Dackor, Eric Boerwinkle, …. Patrick F. Sullivan, and The Tobacco and Genetics Consortium. 2010. “Genome-Wide Meta-Analyses Identify Multiple Loci Associated with Smoking Behavior.” *Nature Genetics* 42(5):441–47. doi: 10.1038/ng.571.

Galea, Sandro, Monica Uddin, and Karestan Koenen. 2011. “The Urban Environment and Mental Disorders.” *Epigenetics* 6(4):400–404. doi: 10.4161/epi.6.4.14944.

Geronimus, Arline T., Margaret Hicken, Danya Keene, and John Bound. 2006. “‘Weathering’ and Age Patterns of Allostatic Load Scores

Among Blacks and Whites in the United States." *American Journal of Public Health* 96(5):826–33. doi: 10.2105/AJPH.2004.060749.

Gonda, Xenia, Gabor Hullam, Peter Antal, Nora Eszlari, Peter Petschner, Tomas GM Hökfelt, Ian Muir Anderson, John Francis William Deakin, Gabriella Juhasz, and Gyorgy Bagdy. 2018. "Significance of Risk Polymorphisms for Depression Depends on Stress Exposure." *Scientific Reports* 8(1):3946. doi: 10.1038/s41598-018-22221-z.

Goosby, Bridget J., and Chelsea Heidbrink. 2013. "The Transgenerational Consequences of Discrimination on African-American Health Outcomes." *Sociology Compass* 7(8):630–43. doi: 10.1111/soc4.12054.

Guo, Guang, Hexuan Liu, Ling Wang, Haipeng Shen, and Wen Hu. 2015. "The Genome-Wide Influence on Human BMI Depends on Physical Activity, Life Course, and Historical Period." *Demography* 52(5):1651–70. doi: 10.1007/s13524-015-0421-2.

Hales, Craig M., Cheryl D. Fryar, Margaret D. Carroll, David S. Freedman, Yutaka Aoki, and Cynthia L. Ogden. 2018. "Differences in Obesity Prevalence by Demographic Characteristics and Urbanization Level Among Adults in the United States, 2013-2016." *JAMA* 319(23):2419–29. doi: 10.1001/jama.2018.7270.

Jaffee, S. R., and T. S. Price. 2007. "Gene–Environment Correlations: A Review of the Evidence and Implications for Prevention of Mental Illness." *Molecular Psychiatry* 12(5):432–42. doi: 10.1038/sj.mp.4001950.

James, Wesley L. 2014. "All Rural Places Are Not Created Equal: Revisiting the Rural Mortality Penalty in the United States." *American Journal of Public Health* 104(11):2122–29. doi: 10.2105/AJPH.2014.301989.

Jensen, Leif, Shannon M. Monnat, John J. Green, Lori M. Hunter, and Martin J. Sliwinski. 2020. "Rural Population Health and Aging: Toward a Multilevel and Multidimensional Research Agenda for the 2020s." *American Journal of Public Health* 110(9):1328–31. doi: 10.2105/AJPH.2020.305782.

Kashyap, Meghana V., Michael Nolan, Marc Sprouse, Ranajit Chakraborty, Deanna Cross, Rhonda Roby, and Jamboor K. Vishwanatha. 2015. "Role of Genomics in Eliminating Health Disparities." *Journal of Carcinogenesis* 14. doi: 10.4103/1477-3163.165158.

Kaufman, Brystana G., Sharita R. Thomas, Randy K. Randolph, Julie R. Perry, Kristie W. Thompson, George M. Holmes, and George H. Pink. 2016. “The Rising Rate of Rural Hospital Closures.” *The Journal of Rural Health* 32(1):35–43. doi: 10.1111/jrh.12128.

Kessler, Ronald C., Katherine A. McGonagle, Shanyang Zhao, Christopher B. Nelson, Michael Hughes, Suzann Eshleman, Hans-Ulrich Wittchen, and Kenneth S. Kendler. 1994. “Lifetime and 12-Month Prevalence of DSM-III-R Psychiatric Disorders in the United States: Results From the National Comorbidity Survey.” *Archives of General Psychiatry* 51(1):8–19. doi: 10.1001/archpsyc.1994.03950010008002.

Kessler, Ronald C., Howard Birnbaum, Evelyn Bromet, Irving Hwang, Nancy Sampson, and Victoria Shahly. 2010. “Age Differences in Major Depression: Results from the National Comorbidity Surveys Replication (NCS-R).” *Psychological Medicine* 40(2):225. doi: 10.1017/S0033291709990213.

Krieger, Nancy, Jarvis T. Chen, Pamela D. Waterman, David H. Rehkopf, and S. V. Subramanian. 2003. “Race/Ethnicity, Gender, and Monitoring Socioeconomic Gradients in Health: A Comparison of Area-Based Socioeconomic Measures—The Public Health Disparities Geocoding Project.” *American Journal of Public Health* 93(10):1655–71. doi: 10.2105/AJPH.93.10.1655.

Lambert, Samuel A., Laurent Gil, Simon Jupp, Scott C. Ritchie, Yu Xu, Annalisa Buniello, Aoife McMahon, Gad Abraham, Michael Chapman, Helen Parkinson, John Danesh, Jacqueline A. L. MacArthur, and Michael Inouye. 2021. “The Polygenic Score Catalog as an Open Database for Reproducibility and Systematic Evaluation.” *Nature Genetics* 53(4):420–25. doi: 10.1038/s41588-021-00783-5.

Legrand, Lisa N., Margaret Keyes, Matt McGue, William G. Iacono, and Robert F. Krueger. 2008. “Rural Environments Reduce the Genetic Influence on Adolescent Substance Use and Rule-Breaking Behavior.” *Psychological Medicine* 38(9):1341–50. doi: 10.1017/S0033291707001596.

Locke, A.E., Kahali, B., Berndt, S.I., Justice, A.E., Pers, … Speliotes, E.K., Epidemiology and biostatistics, NCA - Neurobiology of mental health, and EMGO - Lifestyle, overweight and diabetes. 2015. “Genetic studies of body mass index yield new insights for obesity biology.” *Nature* 518(7538):197–206. doi: 10.1038/nature14177.

Lovejoy, J. C., and A. Sainsbury. 2009. “Sex Differences in Obesity and the Regulation of Energy Homeostasis.” *Obesity Reviews* 10(2):154–67. doi: https://doi.org/10.1111/j.1467-789X.2008.00529.x.

Marmot, Michael. 2005. “Social Determinants of Health Inequalities.” *Public Health* 365:6.

Masters, Ryan K., Eric N. Reither, Daniel A. Powers, Y. Claire Yang, Andrew E. Burger, and Bruce G. Link. 2013. “The Impact of Obesity on US Mortality Levels: The Importance of Age and Cohort Factors in Population Estimates.” *American Journal of Public Health* 103(10):1895–1901. doi: 10.2105/AJPH.2013.301379.

Matthews, Kevin A., Janet B. Croft, Yong Liu, Hua Lu, Dafna Kanny, Anne G. Wheaton, Timothy J. Cunningham, Laura Kettel Khan, Ralph S. Caraballo, James B. Holt, Paul I. Eke, and Wayne H. Giles. 2017. “Health-Related Behaviors by Urban-Rural County Classification — United States, 2013.” *MMWR Surveillance Summaries* 66(5):1–8. doi: 10.15585/mmwr.ss6605a1.

McEwen, Bruce S. 2012. “Brain on Stress: How the Social Environment Gets under the Skin.” *Proceedings of the National Academy of Sciences* 109(Supplement 2):17180–85.

Milner, Kerry A., Viola Vaccarino, Amy L. Arnold, Marjorie Funk, and Robert J. Goldberg. 2004. “Gender and Age Differences in Chief Complaints of Acute Myocardial Infarction (Worcester Heart Attack Study).” *The American Journal of Cardiology* 93(5):606–8. doi: 10.1016/j.amjcard.2003.11.028.

Moy, Ernest, Macarena C. Garcia, Brigham Bastian, Lauren M. Rossen, Deborah D. Ingram, Mark Faul, Greta M. Massetti, Cheryll C. Thomas, Yuling Hong, Paula W. Yoon, and Michael F. Iademarco. 2017. “Leading Causes of Death in Nonmetropolitan and Metropolitan Areas—United States, 1999–2014.” *MMWR Surveillance Summaries* 66(1):1–8. doi: 10.15585/mmwr.ss6601a1.

Nikpay, Majid, Anuj Goel, Hong-Hee Won, Leanne M. Hall, …. Martin Farrall, and the CARDIoGRAMplusC4D Consortium. 2015. “A Comprehensive 1000 Genomes–Based Genome-Wide Association Meta-Analysis of Coronary Artery Disease.” *Nature Genetics* 47(10):1121–30. doi: 10.1038/ng.3396.

O’Connor, A., and G. Wellenius. 2012. “Rural–Urban Disparities in the Prevalence of Diabetes and Coronary Heart Disease.” *Public Health* 126(10):813–20. doi: 10.1016/j.puhe.2012.05.029.

Ogden, Cynthia L., Cheryl D. Fryar, Craig M. Hales, Margaret D. Carroll, Yutaka Aoki, and David S. Freedman. 2018. “Differences in Obesity Prevalence by Demographics and Urbanization in US Children and Adolescents, 2013-2016.” *JAMA* 319(23):2410–18. doi: 10.1001/jama.2018.5158.

Okbay, Aysu, Bart M. L. Baselmans, Jan-Emmanuel De Neve,… and David Cesarini. 2016. “Genetic Variants Associated with Subjective Well-Being, Depressive Symptoms, and Neuroticism Identified through Genome-Wide Analyses.” *Nature Genetics* 48(6):624–33. doi: 10.1038/ng.3552.

Pampel, Fred C. 2006. “Global Patterns and Determinants of Sex Differences in Smoking.” *International Journal of Comparative Sociology* 47(6):466–87. doi: 10.1177/0020715206070267.

Patterson, Paul Daniel, Charity G. Moore, Janice C. Probst, and Judith Ann Shinogle. 2004. “Obesity and Physical Inactivity in Rural America.” *The Journal of Rural Health* 20(2):151–59. doi: https://doi.org/10.1111/j.1748-0361.2004.tb00022.x.

Piper, Megan E., Jessica W. Cook, Tanya R. Schlam, Douglas E. Jorenby, Stevens S. Smith, Daniel M. Bolt, and Wei-Yin Loh. 2010. “Gender, Race, and Education Differences in Abstinence Rates among Participants in Two Randomized Smoking Cessation Trials.” *Nicotine & Tobacco Research* 12(6):647–57. doi: 10.1093/ntr/ntq067.

Price, Alkes L., Nick J. Patterson, Robert M. Plenge, Michael E. Weinblatt, Nancy A. Shadick, and David Reich. 2006. “Principal Components Analysis Corrects for Stratification in Genome-Wide Association Studies.” *Nature Genetics* 38(8):904–9. doi: 10.1038/ng1847.

Rose, Richard J., Danielle M. Dick, Richard J. Viken, and Jaakko Kaprio. 2001. “Gene-Environment Interaction in Patterns of Adolescent Drinking: Regional Residency Moderates Longitudinal Influences on Alcohol Use.” *Alcoholism: Clinical and Experimental Research* 25(5):637–43. doi: 10.1111/j.1530-0277.2001.tb02261.x.

Shields, Alexandra E. 2017. “Epigenetic Signals of How Social Disadvantage ‘Gets under the Skin’: A Challenge to the Public Health Community.” *Epigenomics* 9(3):223–29. doi: 10.2217/epi-2017-0013.

Singer, Marjorie. 2012. “Applying the Concept of Culture to Reduce Health Disparities through Health Behavior Research.” *Preventive Medicine* 55(5):356–61. doi: 10.1016/j.ypmed.2012.02.011.

Sparks, P. Johnelle. 2012. “Rural Health Disparities.” Pp. 255–71 in *International Handbook of Rural Demography*, *International Handbooks of Population*, edited by L. J. Kulcsár and K. J. Curtis. Dordrecht: Springer Netherlands.

Taylor, A. E., M. N. Sandeep, C. S. Janipalli, C. Giambartolomei, D. M. Evans, M. V. Kranthi Kumar, D. G. Vinay, P. Smitha, V. Gupta, M. Aruna, S. Kinra, R. M. Sullivan, L. Bowen, N. J. Timpson, G. Davey Smith, F. Dudbridge, D. Prabhakaran, Y. Ben-Shlomo, K. S. Reddy, S. Ebrahim, and G. R. Chandak. 2011. “Associations of FTO and MC4R Variants with Obesity Traits in Indians and the Role of Rural/Urban Environment as a Possible Effect Modifier.” *Journal of Obesity* 2011:e307542. doi: 10.1155/2011/307542.

U.S. Department of Health and Human Services. 2014. “The Health Consequences of Smoking—50 Years of Progress: A Report of the Surgeon General.” Atlanta: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion, Office on Smoking and Health.

Waldorf, Brigitte, and Ayoung Kim. 2015. “Defining and Measuring Rurality in the US: From Typologies to Continuous Indices.” In commissioned paper presented at the Workshop on Rationalizing Rural Area Classifications, Washington, DC.

Ware, Erin B., Lauren L. Schmitz, Jessica Faul, Arianna Gard, Colter Mitchell, Jennifer A. Smith, Wei Zhao, David Weir, and Sharon LR Kardia. 2017. “Heterogeneity in Polygenic Scores for Common Human Traits.” Preprint. Genetics. DOI: 10.1101/106062

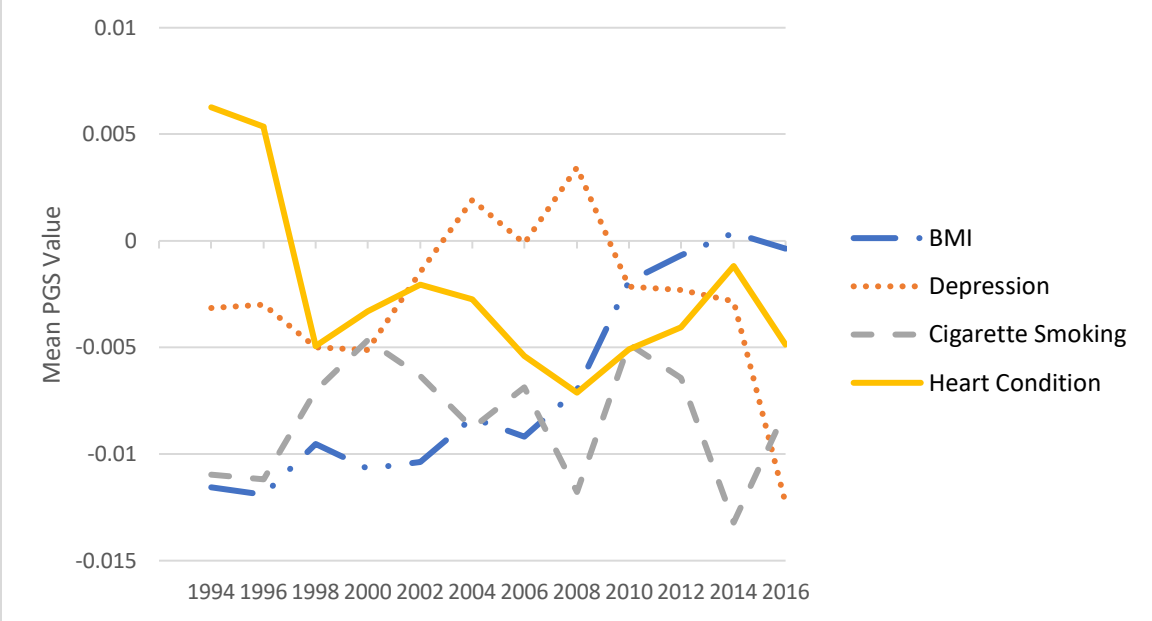
Wedow, Robbee, Meghan Zacher, Brooke M. Huibregtse, Kathleen Mullan Harris, Benjamin W. Domingue, and Jason D. Boardman. 2018. “Education, Smoking, and Cohort Change: Forwarding a Multidimensional Theory of the Environmental Moderation of Genetic Effects.” *American Sociological Review* 83(4):802–32. doi: 10.1177/0003122418785368.

Whitley, Sarah. 2013. “Changing Times in Rural America: Food Assistance and Food Insecurity in Food Deserts.” *Journal of Family Social Work* 16(1):36–52. doi: 10.1080/10522158.2012.736080.

Ziller, Erika C., Jennifer Dunbar Lenardson, Nathan C. Paluso, Jean A. Talbot, and Angela Daley. 2019. “Rural–Urban Differences in the Decline of Adolescent Cigarette Smoking.” *American Journal of Public Health* 109(5):771–73. doi: 10.2105/AJPH.2019.304995.

SUPPLEMENT

Figure S1. Average Polygenic Score by Wave



Note: We omit the year 1992, as depression was not measured at that wave.

Table S1: Multilevel Output from Race-Stratified Models

	White			Black			
	b	se	p	b	se	p	p B=W
				BMI			
PGS	1.442	0.049	.000	1.061	0.150	.000	.016
Rural	-0.138	0.082	.093	-0.272	0.262	.300	.626
Rural*PGS	-0.186	0.082	.024	-0.152	0.239	.526	.894
N (Obs)	99392			20259			
N (Ind)	11883			2999			
				Heart Condition			
PGS	0.556	0.050	.000	0.139	0.160	.387	.013
Rural	0.241	0.123	.050	-0.276	0.436	.526	.253
Rural*PGS	-0.242	0.113	.033	-0.393	0.341	.249	.674
N (Obs)	104403			21063			
N (Ind)	11888			3002			
				Depression			
PGS	0.266	0.024	.000	0.089	0.044	.044	.000
Rural	0.128	0.098	.189	0.485	0.209	.020	.122
Rural*PGS	-0.155	0.090	.086	-0.094	0.198	.633	.782
N (Obs)	94132			19118			
N (Ind)	11882			3002			
				Current Smoker			
PGS	0.486	0.079	.000	0.176	0.156	.260	.077
Rural	-0.116	0.148	.433	-0.352	0.450	.434	.619
Rural*PGS	0.084	0.154	.586	0.449	0.472	.341	.462
N (Obs)	104821			21130			
N (Ind)	11883			3000			

Note: Models are identical to those presented in Table 3 and include the same controls. Only the main and interactive effects for PGS and Rural are presented here.

Figure S2: Study Flow Diagram