Published in final edited form as:

Health Place. 2020 January ; 61: 102260. doi:10.1016/j.healthplace.2019.102260.

“Obesity Among U.S. Rural Adults: Assessing Selection and Causation with Prospective Cohort Data”

Mark Lee

Department of Sociology, University of Minnesota, 267 19th Ave S, Minneapolis, MN 55455, USA

Abstract

Rural adults in the U.S. have disproportionately high rates of obesity, but it is unclear whether this association exists because of selective migration or a contextual effect of the rural environment. Using nationally representative longitudinal data, this study investigates: (1) whether people with obesity select into rural counties, and (2) whether living in a rural area increases body weight after accounting for selection bias. Results indicate that people with obesity are less likely to move to a different county than people without obesity even after controlling for individual and household differences. Next, individual fixed effects regression models, which implicitly control for all time-constant variables, are used to produce a more robust estimate of the effect of rural residence on body weight. Rural residence predicts a significant increase in probability of obesity and body mass index. These results suggest that the association between rural residence and obesity in the United States is likely bidirectional.

Keywords

obesity; rural health; selective migration; obesogenic environment

1. Introduction

There is a disproportionate prevalence of obesity among U.S. rural adults (Hales et al. 2018; Lundeen et al. 2018). This trend increases the overall health burden in rural areas, which is exacerbated by relatively fewer resources and higher barriers to health care access compared to urban areas (Douthit et al. 2015). Identifying the reasons for this pattern is essential to crafting effective policy interventions. However, the factors that explain the relationship between rural residence and obesity are not well understood.

Compositional differences between rural and urban populations partially explain the link between rural residence and obesity. For example, rural residents are more likely to be older, less educated, and live in poverty (U.S. Department of Agriculture 2017, 2018). These

leex6611@umn.edu.

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

Declarations of interest: None.

factors are also linked to excess body weight. However, several studies have shown that living in a rural area continues to predict obesity after accounting for a host of demographic characteristics, including age, sex, race, education, income, employment, marital status, household structure, nativity, and language (Befort et al 2012, Bennett et al 2011, Hales et al 2018, Jackson et al 2005, Patterson et al 2004, Sobal et al 1996, Voss et al 2013, Wen et al 2018). The authors of these studies implicitly assume that the residual association between rural residence and obesity that is not explained by these individual-level control variables represents an unbiased estimate of the effect of the rural context on body weight.

However, it is dubious to infer a causal relationship between rural residence and obesity from previous studies. The county people live in is not randomly assigned; adjusting for demographic covariates in a multiple regression equation does not make it so. Distinguishing between selection and contextual effects on health is difficult in observational studies, especially when the researcher is limited by cross-sectional data (Oakes 2004).

Obese-selective migration could also explain disproportionate obesity prevalence in rural counties. Under the human capital view of migration, people move to seek greater economic returns to their labor. People with obesity tend to face greater discrimination in the work force (Morris 2007, Tunceli et al. 2006). Reduced job opportunities may subsequently reduce an obese person's odds of migrating relative to a similarly qualified person without obesity. This would represent a passive selection process in which people with obesity do not choose to move to rural areas, but they remain there because they lack opportunities for better employment elsewhere. Combined with a net migration flow of people without obesity from rural to urban counties (Johnson and Lichter 2019), this process would create a residual increase in obesity prevalence in rural areas over time even if no true contextual effect were operative.

People without obesity might also be actively selected into urban areas due to lifestyle preferences and local amenities. For example, rural counties are less likely to have accessible walking and biking paths, and a smaller proportion of commuters who use these physically-active modes of transportation (Hansen et al. 2015). A person who highly values these activities may choose to move to an area where the built environment better accommodates active commuting. This process may sort people without obesity into urban counties.

The relationship between rural residence and obesity could be bidirectional: people with obesity selecting into rural counties and the rural environment causing people's weight to increase (see Jokela et al. 2009). The current study used longitudinal data from a nationally representative cohort to improve on existing evidence in the United States. Longitudinal data make it possible to observe the temporal order of events—whether people with obesity are sorted into rural areas, or rural residents become obese. To investigate the possible bidirectional relationship between rural residence and obesity, this study aimed: (1) to demonstrate that individual socio-demographic characteristics do not entirely explain why rural residents have greater odds of obesity, (2) to test whether obesity predicts future inter-county migration, and (3) to test whether moving to a rural county predicts increased probability of obesity and increased BMI in an individual fixed effects analysis.

2. Methods

2.1 Data

The data for this study came from the National Longitudinal Study of Youth 1979 cohort (NLSY79). The NLSY79 began with 12,686 Americans aged 14 to 22 in 1979. Using the sampling weights, the NLSY79 is representative of the national population of those who were aged 14 to 22 on December 31, 1978. The Bureau of Labor Statistics collected data from the sample annually through 1994, then every other year through 2014. The study focuses on employment, though it also gathers data on education, family life, health, and many other topics. A thorough review of the study design has been published by Pergamit and colleagues (2001). This study used data from every even-numbered year between 1986 and 2014 (i.e., 1986, 1988, 1990, ..., 2014), since these are the years when body weight was measured. Restricted geocoded files were provided by the Bureau of Labor Statistics, which included the FIPS county code of every individual at each valid observation.

2.2 Measures

Obesity.—Following Centers for Disease Control and Prevention guidelines, individuals were classified as having obesity for any observation when their BMI exceeded 30 kg/m². BMI was calculated from self-reported height and weight. Respondents' weight was recorded at each observation. Each respondent's height was assumed to be constant after 1985 (mean age 23.6), which was the last year this was reported. Height and weight were transformed into metric units to calculate BMI as kilograms per meter squared.

Rural Residence.—County of residence was determined for every participant at each observation using geocoded data available from the Bureau of Labor Statistics. Counties were linked to their 2013 Urban Influence Codes, which is a classification scheme developed by the US Department of Agriculture to distinguish metropolitan and non-metropolitan counties by their population size. These codes were then dichotomized to separate metropolitan counties (those with an urbanized population of at least 50,000 and/or adjacent counties with strong commuting ties) from non-metropolitan counties (including micropolitan and non-core counties). Hereafter, the terms urban and rural will be used to refer to metropolitan and non-metropolitan counties respectively.

Migration.—Migration was measured by comparing FIPS county codes between consecutive observations. Within-county migration was not of interest, since this type of migration could not confound rural-urban differences using a county level definition of urbanicity.

County-Level Controls.—To reduce bias in the estimated effect of rurality, dummy variables for census region (south, northeast, north central, and west) and an indicator of "persistent poverty" (20 percent or more of its residents classified as poor in consecutive Census Bureau estimates between 1980 and 2011) were included in the models.

Individual-Level Controls.—In addition to these variables, statistical models controlled for time-variant and time-invariant individual-level variables that may confound the

relationship between rural residence and obesity. These include age, age squared (to account with decelerating body weight increase with age), sex, race/ethnicity, education, family income, employment, marital status, family size, nativity, and language spoken in childhood home. Family income was transformed to represent 2014 dollars and logged at each observation. Time-constant variables were measured in 1979. Each of the time-dependent variables were measured consistently between 1986 and 2014.

2.3 Attrition and Missing Data

Observations during which an individual did not complete the survey were dropped from the analysis. This presents potential selection bias in the study design since people who migrate may be more likely to drop from the sample because they are harder for the NLSY investigators to locate for follow up. It is not possible to know whether those who did not complete the survey had migrated to a different county. However, by 2014, only 575 individuals (less than five percent of the original panel) did not participate because the investigators were unable to locate them. Between 1986 and 2014 nearly six thousand participants were observed moving at least once, and many several times. This level of retention makes the NLSY79 sample uniquely suited for studies on internal migration (e.g., Kennan and Walker 2011).

Among the retained cases, some were missing data for control variables. Eight individuals were missing data on nativity and five were missing data on language spoken in childhood. There were also 20 missing observations of marital status and 124 missing observations of educational attainment. Family income had the most non-response with 18,805 missing observations (about 15 percent of all retained observations). Imputing missing data is a preferred strategy over listwise deletion to reduce bias when data are not missing completely at random (Little and Rubin 2002). Multiple imputation with chained equations (MICE) is a flexible way to handle missing data across two or more variables of interest (White, Royston, and Wood 2011). In the first stage, MICE uses iterative equations to generate 10 simulated data sets with no missing observations. All control variables in this study were used to estimate the missing values. Next, a regression model is estimated separately for each simulated data set. Finally, the coefficient estimates across the 10 data sets are combined into a single estimate. This imputation method increases statistical efficiency by retaining all possible cases, while appropriately handling the uncertainty of missing data.

After restricting the sample to those who were not pregnant (199 observed pregnancies) and had valid measures of body weight and county code within the target years, there were 121,334 observations of 11,135 individuals. However, these observations were further restricted in each statistical model as described below.

2.4 Analytic Strategy

First, the cross-sectional association between rural residence and obesity was demonstrated using the 2014 survey wave of the NLSY79. A logistic regression equation was estimated on data from the 2014 survey wave (including the 6,868 individuals who participated in that wave) where the log odds of obesity were modeled as a function of rural residence and all demographic controls. The model was weighted using the 2014 sampling weights.

Next, the effect of obesity on inter-county migration behavior between consecutive survey rounds was estimated. For this analysis, all consecutive observations between 1986 and 2014 were pooled. First, a Generalized Estimating Equation (GEE) was estimated where any inter-county migration between t_i and t_{i+2} (the next observation, two years later) was modeled as a function of obesity plus the same set of demographic covariates included in the cross-sectional model measured at t_i . This model necessarily restricted the sample to those who have valid consecutive observations yielding 106,634 observations of 10,829 individuals. For example, if someone was observed in 1986, 1988, and 1992 only, then the 1992 observation would be dropped. This restriction maintained consistency in time scale between observations during which migration is predicted. The GEE model accounts for the correlation of observations within individuals by assuming a “working correlation structure” between responses in the same cluster and a robust variance estimator to adjust the standard errors (Zeger, Liang, and Albert 1988). The model was weighted using longitudinal weights to account for oversampling and attrition.

For the subsample of consecutive observations where a migration occurred, a GEE model estimated the effect of obesity at t_i on whether a migrant ends up in an urban county at t_{i+2} after controlling demographic characteristics at t_i . This model was restricted to the 13,203 observations of 5,795 individuals who migrated between consecutive survey waves. The model was also weighted using longitudinal weights.

Next, individual fixed effects regression models were used to estimate the effect of rural residence on probability of obesity and BMI. Individual fixed effects regression reduces bias by modeling change in a dependent variable within people (Allison 2009). Therefore, all time-constant variables are implicitly controlled for even if they are not observable. These models will produce a less-biased estimate of the effect of rural residence on body weight if any confounding selection effect (such as lifestyle preference) is relatively constant over time. Jokela (2014) used a similar design to estimate causal effects of neighborhood on health.

This study estimated a linear probability model of obesity and a linear regression model of BMI as functions of rural residence plus time-varying covariates. Because individual fixed effects models cannot estimate the effect of something that does not change within-person, the sample for these models was restricted to the 1,723 individuals who are observed living in both an urban and rural county at least once between 1986 and 2014. Clustered robust standard errors were used to account for correlation and the models are weighted using longitudinal weights.

Finally, quantile regression models with individual fixed effects were estimated to examine whether exposure to a rural environment had different effects on body weight across the BMI distribution. The *xtqreg* module (Machados and Santos Silva 2018) in Stata was used to estimate individual fixed effects regression models of BMI at every decile. These models were identical to the fixed effects linear regression model of BMI in terms of covariates. The *xtqreg* module does not estimate with imputed data, so the sample size is somewhat reduced by listwise deletion. The models were weighted using longitudinal weights.

3. Results

3.1 Cross-Sectional Analysis

Table 1 shows demographic characteristics of the NLSY79 sample by rural-urban status using the 2014 wave of data. NLSY79 participants in rural counties were more likely to have obesity than their urban peers (42.1% versus 35.7%). Rural and urban individuals also differed significantly by race/ethnicity (fewer Black and Hispanic residents in rural counties), education (lower in rural counties), income (lower in rural counties), employment (fewer working in rural counties), marital status (fewer never married in rural counties), family size (smaller in rural counties), nativity (fewer foreign-born in rural counties), non-English language exposure (fewer in rural counties), region (more rural residents live in North Central and South) and county-level persistent poverty (more common in rural counties).

Despite the compositional differences, rural residence continued to show an independent association with obesity after adjusting for all demographic covariates. Table 2 displays the results of the logistic regression equation. People living in a rural county in 2014 had greater odds (OR 1.325; 95% CI 1.126-1.559) of having obesity than otherwise similar peers living in an urban county. This demonstrates that the association between rural residence and obesity is not entirely explained by compositional differences between rural and urban populations. The county's region and persistent poverty status also do not explain the relationship between rural residence and obesity odds.

3.2 Selective Migration Analysis

The first migration GEE model (Table 3) showed that people with obesity are less likely to migrate to a different county within the next two years than people without obesity. This model controlled for many demographic characteristics measured at t_i that also predicted migration by t_{i+2} , including region, county-level poverty, age, sex, race/ethnicity, education, employment, marital status, and family size. Even after adjusting for these controls, having obesity predicted reduced odds (OR 0.919; 95% CI 0.850-0.993) of moving to a different county in the next two years. This suggests that people with obesity are less migratory than people without obesity, which is a possible source of selection bias in the cross-sectional estimate.

The second migration GEE model (Table 4) showed that, among inter-county migrants, having obesity at t_i did not significantly predict whether an individual would end up in a rural or urban county at t_{i+2} after adjusting for all other factors. While the point estimate is in the anticipated direction (reduced odds of moving to an urban versus rural county if obese), the wide confidence interval indicates no true effect in the population. This suggests that obesity does not influence selection into a rural or urban county after the decision to migrate has been made.

3.3 Individual Fixed Effects Regression

The individual fixed effects models (Table 5) indicated that exposure to a rural environment predicts a significant within-person increase in body weight. These models did not include

covariates for sex, race/ethnicity, nativity, and exposure to non-English language in childhood because these were time-constant characteristics. The first column displays estimated coefficients from the linear probability model of obesity. Moving from an urban to a rural county predicted a 2.0 percentage point increase in risk of having obesity (95% CI 0.4-3.5 percentage point greater risk). This within-person change in obesity risk is not explained by any time-constant individual-level characteristics (controlled with fixed effects) nor by any coinciding changes in age, education, income, employment, marital status, or family size.

The second column measures mean changes in BMI, which is a meaningful outcome in addition to probability of obesity. Weight gain may fall below the threshold of obesity but still result in adverse health effects. Similarly, weight gain could occur among people who already have obesity, which would not be detected in the previous linear probability model. The model shows that living in a rural instead of an urban county produces an estimated 0.275 kg/m² increase (95% CI 0.121-0.429 kg/m² increase) in BMI. This represents approximately a two-pound increase in body weight for an average adult.

3.4 Quantile Regression Models

Quantile regression models show how exposure to the rural environment effects different points of the BMI distribution. Figure 1 plots the range of estimated effects of rural residence on BMI across quantiles. The coefficient for rural residence is positive and significant in each model. Although the estimated effect at the 90th percentile (0.319) is slightly higher than the estimated effect at the 10th percentile (0.232), no estimate is significantly different from the OLS estimate of 0.275 kg/m². This supports the idea that living in a rural area generates a uniform upward shift in the BMI distribution.

4. Discussion

Previous studies describing the link between rurality and obesity have acknowledged that cross-sectional observational data are inadequate to infer causal effects because of potential selection bias. This is the first study to date examining the longitudinal relationship between rural residence and obesity in the United States. Results confirm previous studies that there is a cross-sectional association between rural residence and obesity after adjusting compositional differences between rural and urban populations. However, this cross-sectional association is biased by obese-selective inter-county migration behavior.

People with obesity are somewhat less likely to migrate than otherwise similar people without obesity. This effect of obesity on migration remained significant after adjusting for all other individual demographic characteristics that previous studies have used to estimate the contextual effect of rural residence on obesity. Testing the mechanisms producing this relative immobility among Americans with obesity is beyond the scope of this study, but it could be related to poorer job prospects for them compared with their peers who are not obese, or some other unmeasured quality. As people with obesity in rural and urban counties remain in place, a substantial proportion of rural counties are losing population as people (typically without obesity) move to urban centers (Johnson and Lichter 2019). This represents a passive form of selection bias. Over time, the less-migratory obese population

will take up a larger share of the rural population and smaller share of the urban population as people without obesity move from rural to urban counties. There was not, however, evidence for a more active selection mechanism. Once the decision to migrate has occurred, people are not less likely to end up in an urban county if they have obesity. This indicates that the cross-sectional association between rurality and obesity is confounded by people with obesity choosing not to migrate, but not by internal migrants choosing to live in counties that accommodate their pre-set lifestyle preferences.

Does selective migration totally account for the observed association between rurality and obesity, or is a contextual effect also at work? The fixed effects regression model indicates that people are more likely to become obese while living in a rural compared with an urban county. Rural residence also predicts a slight increase in body weight. These fixed effects models are stringent tests of the contextual effect of rural residence since they implicitly control for all time-constant characteristics that might influence body weight. They also controlled for measured time-variant factors that could influence someone's decision to live in a rural or urban area, such as age, marital status, and employment. All these factors did not entirely explain the significant positive relationship between rural residence and body weight increase. Furthermore, the quantile regression models show that exposure to the rural environment predicts a uniform upward shift in the distribution of body weight. In other words, everyone living in a rural environment becomes heavier, not just those with already extreme body weights. Since all rural residents endure this health burden, it follows that all would benefit if some policy intervention could modify the rural environment to be less obesogenic.

This study contributes to substantive debates regarding whether geographic health disparities are partially an artefact of selective migration. Outside the United States, there is considerable evidence for health-selective migration. In a study that closely parallels the current one, Jokela and colleagues (2009) demonstrated that both selective migration and contextual effects contribute to rural-urban body weight differences in Finland. Several studies in the United Kingdom show that selective migration is at least partially responsible for geographic disparities such as rural-urban differences in mortality (Riva, Curtis, and Norman 2011) and differences in long-term illness rates by area level of deprivation (Brown and Leyland 2009). However, current evidence in the United States is sparse. These findings contradict Geronimus and colleagues' (2014) study, which found that cross-sectional geographic disparities in Americans' health were not significantly confounded by migration. It could be that the current findings differ because this study focuses on rural to urban migration, while Geronimus and colleagues analyzed mobility between zip codes within large metropolitan areas. Rural areas with declining population may be more dramatically affected by internal migration patterns. Additional empirical evidence is needed to adjudicate between these disparate results.

This study also contributes to methodological discussion about how best to assess the joint influence of selective migration and place effects on geographic health disparities. Alternate approaches have been proposed. Norman and colleagues (2005, 2014) hypothetically "put people back" to imagine what geographic health disparities would have been had no one moved. This method requires observing people at two points in time and calculating

geographic differences in health outcomes at Time 2 based on where people had been living at Time 1. Comparing the observed geographic disparity (i.e., using Time 2 outcomes and Time 2 location) with the counterfactual geographic disparity (i.e., using Time 2 outcomes and Time 1 location) is a straightforward way to estimate the effect of health-selective migration. However, this method relies on the strong assumption that post-migration health outcomes are only determined by pre-migration factors. If, on the other hand, an individual's health improves because of moving from a deprived area to an affluent area, it seems unlikely that his or her health outcomes would have been the same under the counterfactual scenario of remaining in the deprived area. By contrast, this study shows that pre-migration obesity status predicts a person's risk of migrating in the future. This is stronger evidence for health-selective migration than the "put people back" method.

Another approach to estimating causal effects of place on health involves propensity score matching. Green, Arcaya, and Subramanian (2017) generated propensity scores for future migration based on a logistic regression of observable pre-migration characteristics, such as age, ethnicity, education, and certain health factors. Then, they matched individuals who had similar propensity scores but differed in migration outcomes. This method creates quasi treatment and control groups that appear similar in odds of receiving the treatment (i.e., migration in this case), thus ostensibly resolving the missing counterfactual problem. However, propensity scores are not immune to omitted variable bias, since only observed characteristics can be used to generate the scores. By contrast, modeling within-person change using individual fixed effects implicitly controls all characteristics that are time-constant even if they are not measured in the data. This makes individual fixed effects models robust for estimating causal effects of place (Jokela 2014).

Future studies should attempt to identify the specific contextual mechanisms that mediate excess obesity in rural counties. Previous cross-sectional studies have pointed to potentially important factors, such as poverty, food environment, transportation options, and accessibility of recreation (Bennett, Probst, and Pumkam 2011; Wen, et al. 2018). However, robust causal evidence is still lacking. Future investigations should also adopt a life course approach to estimate how timing and duration of exposure to the rural environment influences body weight (see Patterson et al. 2017). These analyses would be useful in identifying policy interventions that could effectively reduce obesity prevalence in rural counties.

4.1 Limitations

While these findings represent a significant contribution to the literature, there are several limitations to the study. First, it relies on self-reported height and weight, rather than measured, to calculate BMI. Self-reports of weight are not always accurate (Rowland 1990). Another limitation is that 2013 county-level definitions are used to identify rural and urban areas across all observations from 1986 to 2014. Some counties may have switched between rural and urban status over the decades of observation, and this study is unable to detect such changes.

The fixed effects regression models are limited by excluding all individuals (representing the majority) who are only observed in an urban or rural county between 1986 and 2014. Fixed

effects models capitalize on internal validity by reducing omitted variable bias, but this comes at the expense of estimating effects on a restricted sample of migrants that may not represent the broader population well. Additionally, the fixed effects regression only controls for time-invariant confounders. It cannot help with unobserved confounders that change over time. However, the models in this study reduce this threat by controlling for many observed time-varying confounders, including age, income, education, employment, and marital status.

5. Conclusion

This study suggests that there is a bidirectional relationship between rural residence and obesity. Therefore, policy interventions aiming to modify the rural environment to reduce obesity will not entirely close the gap between rural and urban areas. Selective migration also plays a role in shaping rural health. Researchers have described the rural “brain drain”—the trend of bright young people leaving small towns for larger cities where there is a greater economic return for their labor (Carr and Kefalas 2009). It is an under-reported fact that those “brains” come with bodies, and those bodies tend to be physically healthier than the ones who stay behind. Efforts to attract and retain skilled and educated workers would not only boost local economies, it could also help to improve the health composition of rural populations.

Acknowledgments:

I thank John Robert Warren, Jeylan Mortimer, Carrie Henning-Smith, J. David Hacker, Audrey Dorelien, population studies trainees at the Minnesota Population Center, and attendants of the 2019 meeting of the Population Association of America for their valuable guidance. However, errors and omissions are my responsibility.

Funding: Research reported in this publication was supported by the Eunice Kennedy Shriver National Institute of Child Health and Human Development (NICHD) under Award Number 1T32HD095134-01A1. This project also benefited from support provided by the Minnesota Population Center, which receives core funding (Grant Number P2CHD041023) from the NICHD.

REFERENCES

- Allison PD. (2009). Fixed Effects Regression Models. SAGE Publications.
- Befort CA, Nazir N, and Perri MG. (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: 392–397. [PubMed: 23083085]
- Bennett KJ, Probst JC, and Pumkam C. (2011). “Obesity among Working Age Adults: The Role of County-Level Persistent Poverty in Rural Disparities.” *Health and Place* 17: 1174–1181. [PubMed: 21665515]
- Brown D and Leyland AH. (2009). “Population Mobility, Deprivation and Self-Reported Limiting Long-Term Illness in Small Areas Across Scotland.” *Health and Place* 15(1): 37–44. [PubMed: 18373945]
- Carr PJ and Kefalas MJ. (2009). *Hollowing Out the Middle: The Rural Brain Drain and What It Means for America*. Boston, MA: Beacon Press.
- Douthit N, Kiv S, Dwolatzky T, and Biswas S. (2015). “Exposing Some Important Barriers to Health Care Access in the Rural USA.” *Public Health* 129: 611–620. [PubMed: 26025176]
- Geronimus AT, Bound J, and Ro A. (2014). “Residential Mobility Across Local Areas in the United States and the Geographic Distribution of the Healthy Population.” *Demography* 51(3): 777–809. [PubMed: 24781651]

- Green MA, Arcaya M, and Subramanian SV. (2017). "Using Internal Migration to Estimate the Causal Effect of Neighborhood Socioeconomic Context on Health: A Longitudinal Analysis, England, 1995-2008." *Annals of the American Association of Geographers* 107(6): 1266–1278.
- Hales CM, Fryar CD, Carroll MD, Freedman DS, Aoki Y, and Ogden CL. (2018). "Differences in Obesity Prevalence by Demographic Characteristics and Urbanization Level Among Adults in the United States, 2013-2016." *JAMA* 319(23): 2419–2429. [PubMed: 29922829]
- Hansen AY, Umstadtd Meyer MR, Lenardson JD and Hartley D. (2015). "Built Environments and Active Living in Rural and Remote Areas: A Review of the Literature." *Current Obesity Reports* 4(4): 484–493. [PubMed: 26364307]
- Jackson JE, Doescher MP, Jerant AD, and Hart LG. (2005). "A National Study of Obesity Prevalence and Trends by Type of Rural County." *The Journal of Rural Health* 21(2): 140–148. [PubMed: 15859051]
- Johnson KM and Lichter DT. (2019). "Rural Depopulation: Growth and Decline Processes over the Past Century." *Rural Sociology* 84(1): 3–27.
- Jokela M, Kivimäki M, Elovainio M, Viikari J, Raitakari OT, and Kiltikangas-Järvinen L. (2009). "Urban/Rural Differences in Body Weight: Evidence for Social Selection and Causation Hypotheses in Finland." *Social Science and Medicine* 68: 867–875. [PubMed: 19147263]
- Jokela M (2014). "Are Neighborhood Health Associations Causal? A 10-Year Prospective Study with Repeated Measurements." *American Journal of Epidemiology* 27: 169–175.
- Kennan J and Walker JR. (2011). "The Effect of Expected Income on Individual Migration Decisions." *Econometrica* 79(1): 211–251.
- Little RJA and Rubin DB. (2002). *Statistical Analysis with Missing Data* (2nd edition). Wiley: Hoboken, NJ.
- Lundeen EA, Park S, Pan L, O'Toole T, Matthews K, and Blanck HM. (2018). "Obesity Prevalence Among Adults Living in Metropolitan and Nonmetropolitan Counties—United States, 2016." *MMWR Morbidity and Mortality Weekly Report* 67: 653–658. [PubMed: 29902166]
- Machados JAF and Santos Silva JMC. (2018). "XTQREG: Stata Module to Compute Quantile Regression with Fixed Effects" *Statistical Software Components S458523*, Boston College Department of Economics.
- Morris S (2007). "The Impact of Obesity on Employment." *Labour Economics* 14(3): 413–433.
- Norman P, Boyle P, and Rees P. (2005). "Selective Migration, Health and Deprivation: A Longitudinal Analysis." *Social Science and Medicine* 60(12): 2755–2771. [PubMed: 15820585]
- Norman P and Boyle P. (2014). "Are Health Inequalities Between Differently Deprived Areas Evident at Different Ages? A Longitudinal Study of Census Records in England & Wales, 1991-2001." *Health and Place* 26: 88–93. [PubMed: 24412656]
- Oakes MJ. (2004). "The (Mis)estimation of Neighborhood Effects: Causal Inference for a Practicable Social Epidemiology." *Social Science and Medicine* 58(10): 1929–1952. [PubMed: 15020009]
- Patterson KAE, Gall SL, Venn AJ, Otahal P, Blizzard L, Dwyer T, and Cleland VJ. (2017). "Accumulated Exposure to Rural Areas of Residence Over the Life Course Is Associated With Overweight and Obesity in Adulthood: A 25-year Prospective Cohort Study." *Annals of Epidemiology* 27: 169–175. [PubMed: 28317611]
- Patterson PD, Moore CG, Probst JC, and Shinogle JA. (2004). "Obesity and Physical Inactivity in Rural America." *The Journal of Rural Health* 20(2): 151–159. [PubMed: 15085629]
- Pergamit MR, Pierret CR, Rothstein DS, and Veum JR. (2001). "The National Longitudinal Surveys." *Journal of Economic Perspectives* 15(2): 239–253.
- Riva M, Curtis S, and Normap P. (2011). "Residential Mobility Within England and Urban-Rural Inequalities in Mortality." *Social Science and Medicine* 73: 1698–1706. [PubMed: 22075259]
- Rowland ML. (1990). "Self-reported Weight and Height." *The American Journal of Clinical Nutrition* 52(6): 1125–1133. [PubMed: 2239790]
- Sobal J, Troiano RP, and Frongillo EA Jr.. (1996). "Rural-Urban Differences in Obesity." *Rural Sociology* 61(2): 289–305.
- Tunceli K, Li K, and Williams LK. (2006). "Long-Term Effects of Obesity on Employment and Work Limitations Among U.S. Adults, 1986 to 1999." *Obesity* 14(9): 1637–1646. [PubMed: 17030975]

- United States Department of Agriculture. (2017). "Rural Education at a Glance, 2017 Edition." Economic Information Bulletin 171 4 2017.
- United States Department of Agriculture. (2018). "Rural America at a Glance, 2018 Edition." Economic Information Bulletin 200 11 2018.
- Voss JD, Masuoka P, Webber BJ, Scher AI, and Atkinson RL. (2013). "Association of Elevation, Urbanization and Ambient Temperature with Obesity Prevalence in the United States." International Journal of Obesity 37(10): 1407–1412. [PubMed: 23357956]
- Wen M, Fan JX, Kowaleski-Jones L, and Wan N. (2018). "Rural-Urban Disparities in Obesity Prevalence Among Working Age Adults in the United States: Exploring the Mechanisms." American Journal of Health Promotion 32(2): 400–408. [PubMed: 29214811]
- White IR, Royston P, and Wood AM. (2011). "Multiple Imputation Using Chained Equations: Issues and Guidance for Practice." Statistics in Medicine 30(4): 377–399. [PubMed: 21225900]
- Zeger SL, Liang K, and Albert PS. (1988). "Models for Longitudinal Data: A Generalized Estimating Equation Approach" Biometrics 44(4): 1049–1060. [PubMed: 3233245]

Highlights

- Rural residents were more likely to have obesity than urban residents in a cross-sectional analysis
- People with obesity were less likely to move to a new county than people without obesity
- Moving to a rural county predicted a within-person increase in body weight.

Figure 1:

Estimated Effect of Rural Residence on BMI Across Quantiles of BMI Distribution with 95% CI

Note: Data come from National Longitudinal Study of Youth (1979), years 1986-2014. Estimates derived from quantile regression models with individual fixed effects. Models control for county region, county persistent poverty status, age, age squared, education, family income, employment, marital status, and family size. Models weighted with longitudinal weights.

Table 1:

Compositional Differences Between Rural and Urban Populations in 2014 Survey

<i>N</i> = 6,868			
Variable	Urban		Rural
	% or Mean (SD)		% or Mean (SD)
Obese	35.7%	***	42.1%
Age	53.6 (0.04)		53.8 (0.09)
Male	51.6%		50.1%
Race/Ethnicity		***	
Non-Black, Non-Hispanic	76.5%		91.9%
Black, Non-Hispanic	16.1%		5.8%
Hispanic	7.4%		2.3%
Education		***	
< 12 years	6.3%		9.7%
12 years	39.1%		50.8%
13-15 years	23.4%		22.2%
16 or more years	31.2%		17.2%
Logged income	11.0 (0.02)	*	10.7 (0.04)
Employment last year		***	
Full (50-52 weeks)	68.8%		63.6%
Partial (1 to 49 weeks)	11.4%		11.3%
None (0 weeks)	19.8%		25.0%
Marital Status		***	
Never married	12.6%		7.6%
Married	59.4%		62.8%
Formerly married	28.0%		29.5%
Family size	2.5 (0.02)	*	2.3 (0.04)
Born outside USA	4.3%	***	1.8%
Non-English spoken in home	15.4%	***	7.5%
Region		***	
Northeast	18.4%		6.6%
North Central	24.6%		42.9%
South	37.2%		40.6%
West	19.8%		9.9%
Persistent Poverty County	2.7%	***	9.7%

Note: Data come from National Longitudinal Survey of Youth 1979 Cohort, 2014 survey. Estimates weighted with 2014 cross-sectional sampling weights.

* $p < 0.05$,

** $p < 0.01$,

*** $p < 0.001$

Table 2:

Cross-Sectional Logistic Regression Model Predicting Obesity in 2014

N = 6,868

Independent Variables	OR	95% CI
Rural	1.325 ***	[1.126,1.559]
Persistent Poverty County	0.938	[0.723,1.218]
Region (South)	1.000	
Northeast	0.951	[0.793,1.141]
North Central	0.980	[0.842,1.141]
West	0.775 **	[0.649,0.926]
Age	0.972 *	[0.947,0.998]
Male	1.047	[0.929,1.180]
Race/Ethnicity (Non-Black, Non-Hispanic)	1.000	
Black, Non-Hispanic	1.624 ***	[1.415,1.863]
Hispanic	1.441 **	[1.137,1.826]
Education (12 years)	1.000	
< 12 years	0.890	[0.707,1.120]
13-15 years	0.940	[0.807,1.094]
16 or more years	0.651 ***	[0.556,0.761]
Logged family income	1.039 *	[1.006,1.073]
Employment last year (50-52 weeks)	1.000	
1 to 49 weeks	0.985	[0.814,1.192]
0 weeks	1.169	[0.997,1.369]
Marital status (Never married)	1.000	
Married	0.796 *	[0.649,0.976]
Formerly Married	0.923	[0.760,1.119]
Family Size	1.051	[0.997,1.106]
Born outside USA	0.793	[0.594,1.059]
Non-English in childhood home	1.037	[0.828,1.298]

Note: Data come from National Longitudinal Survey of Youth 1979 Cohort, 2014 survey. Missing data imputed. Model weighted with 2014 cross-sectional sampling weights. Reference group of categorical variables in parentheses.

* $p < 0.05$,

** $p < 0.01$,

*** $p < 0.001$

Table 3:**GEE Model Predicting Inter-County Migration**

Observations = 106,634; Individuals = 10,829

Independent Variables	OR	95% CI
Obese	0.919 *	[0.850,0.993]
Rural	0.951	[0.877,1.032]
Persistent Poverty County	1.175 **	[1.049,1.315]
Region (South)	1.000	
Northeast	0.711 ***	[0.651,0.777]
North Central	0.759 ***	[0.703,0.819]
West	0.894 *	[0.820,0.975]
Age	0.862 ***	[0.838,0.885]
Age Squared	1.001 ***	[1.001,1.002]
Male	1.212 ***	[1.136,1.293]
Race/Ethnicity (Non-Black, Non-Hispanic)	1.000	
Black, Non-Hispanic	0.754 ***	[0.697,0.816]
Hispanic	0.827 **	[0.728,0.939]
Education (12 years)	1.000	
< 12 years	1.021	[0.920,1.134]
13-15 years	1.238 ***	[1.141,1.344]
16 or more years	1.521 ***	[1.401,1.651]
Logged family income	0.980 *	[0.964,0.996]
Employment last year (50-52 weeks)	1.000	
1 to 49 weeks	1.481 ***	[1.398,1.569]
0 weeks	1.570 ***	[1.444,1.707]
Marital status (Never married)	1.000	
Married	0.936	[0.866,1.012]
Formerly Married	1.427 ***	[1.307,1.558]
Family Size	0.893 ***	[0.875,0.912]
Born outside USA	1.095	[0.942,1.273]
Non-English in childhood home	0.891	[0.787,1.008]

Note: Data come from National Longitudinal Survey of Youth 1979 Cohort, survey years 1986-2014. Missing data imputed. Analysis weighted using longitudinal weights. Reference group of categorical variables in parentheses. Working correlation structure was exchangeable.

*
 $p < 0.05$,

**
 $p < 0.01$,

 $p < 0.001$

Table 4:

GEE Model Predicting Urban Destination for Inter-county Migrants

Observations = 13,162; Individuals = 5,795

Independent Variables	OR	95% CI
Obese	0.898	[0.766,1.053]
Rural	0.267 ***	[0.228,0.313]
County Persistent Poverty	1.173	[0.905,1.522]
Region (South)	1.000	
Northeast	1.662 ***	[1.338,2.065]
North Central	0.802 **	[0.693,0.929]
West	1.162	[0.972,1.388]
Age	0.941 *	[0.886,0.999]
Age Squared	1.001	[1.000,1.002]
Male	0.983	[0.866,1.115]
Race/Ethnicity (Non-Black, Non-Hispanic)	1.000	
Black, Non-Hispanic	1.979 ***	[1.693,2.313]
Hispanic	1.600 ***	[1.228,2.083]
Education (12 years)	1.000	
< 12 years	0.816 *	[0.676,0.986]
13-15 years	1.224 *	[1.042,1.437]
16 or more years	1.772 ***	[1.495,2.100]
Logged family income	1.058 **	[1.022,1.094]
Employment last year (50-52 weeks)	1.000	
1 to 49 weeks	0.877	[0.767,1.002]
0 weeks	0.954	[0.790,1.152]
Marital status (Never married)	1.000	
Married	0.801 **	[0.679,0.946]
Formerly Married	0.822 *	[0.690,0.979]
Family Size	0.970	[0.929,1.013]
Born outside USA	1.239	[0.828,1.854]
Non-English in childhood home	1.158	[0.900,1.490]

Note: Data come from National Longitudinal Survey of Youth 1979 Cohort, survey years 1986-2014. Missing data imputed. Analysis weighted using longitudinal weights. Reference group of categorical variables in parentheses. Working correlation structure was exchangeable.

*
 $p < 0.05$

**
 $p < 0.01$

 $p < 0.001$

Table 5:

Individual Fixed Effects Models Predicting Obesity and Body Mass Index

Independent Variables	LPM of Obesity		OLS of Body Mass Index	
	<i>Obs = 20,684; Ind = 1,723</i>		<i>Obs = 20,684; Ind = 1,723</i>	
	Est.	95% CI	Est.	95% CI
Rural	0.020 [*]	[0.004,0.035]	0.275 ^{***}	[0.121,0.429]
County Persistent Poverty	-0.047 ^{**}	[-0.081,-0.012]	-0.156	[-0.579,0.267]
Region (South)	0.000		0.000	
Northeast	-0.019	[-0.082,0.043]	-0.106	[-0.565,0.353]
North Central	-0.003	[-0.041,0.034]	-0.043	[-0.472,0.386]
West	-0.009	[-0.052,0.035]	-0.143	[-0.615,0.329]
Age	0.017 ^{***}	[0.010,0.024]	0.466 ^{***}	[0.399,0.534]
Age Squared	-0.000 [*]	[-0.000,-0.000]	-0.004 ^{***}	[-0.005,-0.003]
Education (12 years)	0.000		0.000	
< 12 years	0.036	[-0.020,0.092]	0.396	[-0.130,0.921]
13-15 years	0.008	[-0.045,0.061]	0.339	[-0.163,0.842]
16 or more years	0.008	[-0.059,0.076]	0.281	[-0.417,0.979]
Logged family income	0.004 [*]	[0.000,0.007]	0.055 ^{**}	[0.019,0.090]
Employment last year (50+)	0.000		0.000	
1 to 49 weeks	0.003	[-0.010,0.015]	0.086	[-0.027,0.199]
0 weeks	0.025 [*]	[0.002,0.048]	0.511 ^{***}	[0.292,0.730]
Marital status (Never marr.)	0.000		0.000	
Married	0.041 [*]	[0.002,0.079]	0.330	[-0.005,0.664]
Formerly Married	0.011	[-0.031,0.053]	-0.339	[-0.714,0.036]
Family Size	-0.002	[-0.008,0.004]	-0.071 ^{**}	[-0.125,-0.017]
Constant	-0.343 ^{***}	[-0.470,-0.215]	14.246 ^{***}	[12.983,15.508]

Note: Data come from National Longitudinal Survey of Youth 1979 Cohort. Years 1986-2014. Missing data imputed. Reference group of categorical variables in parentheses. LPM = Linear Probability Model. OLS = Ordinary Least Squares model. Standard errors are adjusted to account for repeated observations of individuals.

*
 $p < 0.05$

**
 $p < 0.01$

 $p < 0.001$