Health, Place, and Class

Level of rurality, demographic identity, socioeconomic status, or infrastructure: which is the most important predictor of self-reported health?

By: Grace Liu

Abstract

Background: It is well established that rural residents have worse health than non-rural residents, however the health outcomes of suburban residents and those residing in medium or small metropolitan areas have received less attention. The purpose of this study was to examine the association between level of rurality and self-reported poor health and to determine the impact of demographic identity, socioeconomic status, and infrastructure availability on this relationship.

Methods: This study utilized a nationally representative sample of US adults (age 18 – 64) from the National Health Interview Survey 2019 sample. The final analytic sample included 20,488 individuals (43.14 ± 13.19 years, 52.40% female). Level of rurality was classified as urban, suburban, medium and small metropolitan area, or rural. General health status was dichotomized into a binary indicator variable for poor health. Covariates of interest included region of the United States, demographic identity, socioeconomic status, and infrastructure availability.

Results: Our crude logistic regression model found that suburban respondents had 17% lower odds (95% CI: 26% lower – 6% lower) of reporting poor/fair health, those residing in medium or small metropolitan areas had 14% higher odds (95% CI: 3% higher – 28% higher) of reporting poor/fair health, and rural respondents had 73% higher odds (95% CI: 53% higher – 95% higher) of reporting poor/fair health, when compared to urban respondents. This trend remained consistent throughout all models and after adjustment for all covariates of interest.

Conclusions: Level of rurality was associated with self-reported poor health. Specifically, when compared to individuals residing in urban areas, increasing rurality was associated with increasing odds of poor health. This relationship remained significant after adjusting for covariates of interest but was attenuated.

Introduction

In the United States over 15% of the population lives in rural areas.[1] It has long been understood that people living in rural areas in the United States have worse health than those who reside in other areas, meaning that approximately 50 million Americans are at a higher risk of disability, disease, and death than their counterparts living in urban or suburban areas, or medium and small cities/towns. People living in rural communities are more likely to report poor health[2], and have a higher prevalence of chronic diseases (such as coronary heart disease[3,4], chronic obstructive pulmonary disease[5], and diabetes[4], among others) and disability.[6] Excluding deaths from the COVID-19 pandemic, rural areas also have a 20% higher rate of age-adjusted mortality[7], higher rates of death for all 10 of the leading causes of death in the United States[7], and higher percentages of excess death for the five leadings causes of death.[8] Additionally, increasing rurality is associated with decreasing life expectancy[9], with the exception of suburban areas, which have much better health than all other levels of rurality.[10]

Much of the research on urban/rural health disparities in the United States focuses on a specific outcome metric (i.e. prevalence of specific diseases, mortality, etc.). While this is a valid and important method for understanding the mechanisms of the relationship between level of rurality and a specific outcome metric, self-reported health serves as a more robust measure of health status and a proxy for health care utilization, physician rated health, morbidity, and mortality[11], making it a fitting measure for analyses interested in health status as an outcome. Additionally, most literature rarely mentions classifications in between urban and rural, such as suburban counties or medium and small metropolitan areas, leaving a large gap in knowledge about how different levels of rurality relate to health. Due to the complexities of place-based health disparities, it is difficult to understand whether these disparities are simply due to the rurality of an area, if they are a result of the demographic and socioeconomic composition of different levels of rurality, or if the varying level of infrastructure availability and accessibility of different levels of rurality is responsible. In order to craft effective interventions for narrowing gaps in health, health outcomes, and mortality, we must first understand the root cause of place-based health disparities.

The purpose of this study was to examine the association between level of rurality and self-reported poor health among adults using National Health Interview Survey (NHIS) data from 2019. In addition, we evaluated the impact that adjusting for demographic identity, socioeconomic status, and key infrastructure availability, such as food security, access to health care, and health insurance coverage had on this association. We hypothesized that increasing rurality would be associated with increasing odds of poor health, with the exception of suburban areas, which we hypothesized would have the lowest odds of poor health. Additionally, we hypothesized that after adjustment, the relationship between level of rurality and odds of poor health would be attenuated.

Methods

Sample Design & Study Population

The National Health Interview Survey (NHIS) is a cross-sectional household survey using a nationally representative, geographically clustered, multi-stage probability sample of the civilian noninstitutionalized population residing in the United States.[12] Excluded from sampling are those with no fixed household address (e.g. homeless and/or transient persons not residing in shelters), active-duty military personnel and civilians living on military bases, persons in long-term care institutions, persons in correctional facilities, and U.S. nationals living in foreign countries.[12]

The original 2019 NHIS sample included 41,190 individuals from ages 0 – 85 years old. This analysis included only individuals between the ages of 18 and 64 years old (N = 22,621). Individuals were excluded if they had missing values for general health status (n = 15), demographic characteristics (n = 1,358), socioeconomic characteristics (n = 90), infrastructure variables (n = 620), or insurance coverage (n = 50). Following all exclusions, the size of the final analytic sample was 20,488 individuals.

All data was collected by Field Representatives trained and directed by health survey supervisors in the U.S. Census Bureau Regional Offices to conduct interviews for NHIS.[12] The majority of the data is self-report data, all information is collected through face-to-face surveys in the respondents' home, with the option of a telephone interview.[12]

Outcome and Exposure Variables

Outcome of Interest: Self-Reported Health Status

The outcome of interest in this analysis was self-reported health status, which was given as a five-level ordinal variable ranging from excellent health to poor health. For the purposes of this study, general health status was dichotomized into a binary variable for poor health, with poor health and fair health grouped together (poor health), and good health, very good health, or excellent health grouped together (good health). Those with good health served as the reference level for this analysis.

Exposure of Interest: Level of Rurality

The primary exposure of interest in this analysis was level of rurality, as defined by the National Center for Health Statistics (NCHS) 2013 urban-rural classification scheme for counties.[2] The NHIS Field Representatives obtained the zip code of residence from survey respondents, then assigned one of the following classification levels: large central metropolitan, large fringe metropolitan, medium and small metropolitan, and nonmetropolitan. The remainder of this analysis will refer to large central metropolitan areas as urban, large fringe metropolitan areas as suburban, and nonmetropolitan areas as rural. For the purposes of this study, large central metropolitan areas served as the reference level.

Additional Covariates

Region of the United States

Region of the United States was given as a four-level categorical variable, designating whether a respondent was from the northeast, north central/Midwest, south, or west portion of the United States.

Demographic Identity

The demographic variables of interest included age, sex, and race. Age was provided as a continuous variable with a range of 18-65. Sex was provided as a binary variable designating male or female, with female serving as the reference level. Race was provided as a six-level categorical variable with values of: White only, Black/African American only, American Indian/Alaska Native only, Asian only, other race and multiple races, and American Indian/Alaska Native and any other race. For this study, White only served as the reference group.

Socioeconomic Status

Socioeconomic variables of interest included educational attainment and poverty. Educational attainment was recoded into a five-level categorical variable with potential values of less than high school, high school degree or equivalent, some college or associate's degree, bachelor's degree, and more than a bachelor's degree, with those with some college or an associate's degree serving as the reference level.

Poverty was recoded to a three-level categorical variable with values of under 1.0 (under the federal poverty threshold (FPT)), between 1.0 and 4.0 (one to four times the FPT), and 4.0 and over (four or more times the FPT), with individuals living at one to four times the FPT serving as the reference level.

Infrastructure

Infrastructure variables of interest included a food security score, having a usual place for medical care as a proxy for access to care, and health insurance coverage. The food security score was given as a continuous value that was a composite indicator reflecting the number of affirmative responses to each of the ten food security questions asked in the NHIS. These include

questions such as if an individual worried that food would run out before getting to buy more, or if they couldn't afford to eat balanced meals, among others. A higher score indicated a higher level of food insecurity.

Having a usual place for care was dichotomized into an indicator variable with responses of: there is no place or no and yes, has a usual place or yes and there is more than one place grouped together, with those that have a usual place for health care serving as the reference level.

Three NHIS variables were merged to create a composite health insurance coverage variable. The composite variable was a five-level categorical variable with values of not covered, covered with private health insurance, covered with Medicaid, and other, with those covered with private insurance serving as the reference level.

Statistical Analysis Plan

This analysis utilized cross-sectional data to examine the association of level of rurality on odds of poor health in adults aged 18-64. Descriptive analyses were performed first to determine if there were any differences between exposure groups in demographic and socioeconomic makeup using median or mean values for continuous variables (age, food security score) and percentages for categorical variables (sex, race, poverty, educational attainment, health insurance coverage status, having a usual place for care).

All models used logistic regression for modeling the odds of poor health. Model 1 was a crude model including only the main exposure of interest: NCHS urban-rural classification. Model 2 expanded on the first by adjusting for region of the United States. Model 3 expanded on Model 2 by including adjustment for demographic characteristics. Model 4 expanded on Model 3 by including adjustment for socioeconomic identity. Lastly, Model 5 expanded on Model 4 by including adjustment for infrastructure variables. Before adding additional adjustments to each model, non-statistically significant covariates were removed. A significance value of $\alpha = 0.05$ was used for all analyses. All statistical analyses were performed in SAS 9.4. Institutional Review Board exemption for this study was granted by the University of Minnesota.

Results

Descriptive Analysis

Descriptive statistics of the total sample population and by level of rurality are provided in Table 1. The final analytic sample for this study included 20,488 individuals with a median age of 43 years. Overall, the sample was 52.4% female and 47.6% male. White individuals made up 77.49% of the sample, with the remainder identifying as Black/African American (12.70%); American Indian/Alaska Native (0.98%); Asian (6.29%); other or multiple races (1.55%); or American Indian/Alaska Native and any other race (0.99%). Of the study population, 29.36% reported living in an urban area, 23.29% in a suburban area, 32.24% in a medium or small metropolitan area, and 15.11% in a rural area. Out of the total sample, 12.29% reported poor or fair health, with respondents residing in rural areas reporting the highest proportion of individuals with poor or fair health (17.90%), and respondents residing in suburban areas reporting the lowest proportion of individuals with poor or fair health (9.53%).

Logistic Regression – Odds of Poor Health

Table 2 provides the odds of reporting poor or fair health (vs. good, very good, or excellent health) and associated 95% confidence intervals as calculated by five nested models. Model 1 estimated the unadjusted odds of poor/fair health by level of rurality. Models 2-5 added

hierarchical adjustments for region of the United States, demographic identity, socioeconomic status, and infrastructure availability, respectively. Model 2 found that the addition of region of the United States did not alter the relationship between level of rurality and odds of poor health. Subsequent models did not include adjustment for region of the United States.

Our crude model found that suburban residents had the lowest odds of reporting poor/fair health (0.83, 95% CI: 0.74 – 0.94) relative to urban respondents. Throughout all adjusted models, suburban respondents continued to have the lowest odds of reporting poor/fair health, with the weakest association with self-reported poor health for suburban respondents present in model 4 (OR = 0.88, 95% CI: 0.77 – 1.00). This odds ratio stayed the same in model 5, indicating that infrastructure availability had little impact on the association between level of rurality and self-reported poor health after adjusting for demographic identity and socioeconomic status.

The association between self-reported poor health and level of rurality for respondents living in a medium or small metropolitan area was found to be rather weak. Our crude model found that respondents residing in a medium or small metropolitan area had a 14% higher odds of reporting poor/fair health when compared to urban respondents (95% CI: 3% higher – 28% higher). Throughout all models, the odds of poor health among respondents residing in a medium or small metropolitan area was higher than suburban respondents and lower than rural residents. Model 4 found that this group had 6% lower odds of reporting poor/fair health when compared to urban respondents (95% CI: 17% lower – 6% higher). Similar to suburban respondents, this odds ratio remained the same in model 5, indicating that infrastructure availability had little impact on the association between level of rurality and self-reported poor health, after adjusting for demographic identity and socioeconomic status.

Respondents residing in rural areas consistently had the highest odds of poor health. Our crude model found that rural respondents had 73% higher odds of reporting poor/fair health when compared to urban respondents (95% CI: 53% higher – 95% higher). The weakest association between self-reported poor health and level of rurality for rural respondents resulted from adjusting for demographic identity and socioeconomic status (model 4): the odds of reporting poor/fair health was only 13% higher for rural respondents when compared to urban respondents. Unlike for suburban respondents and those residing in medium or small metropolitan areas, infrastructure availability did impact the odds of reporting poor/fair health for rural residents: following this adjustment, rural respondents had a 16% higher odds of reporting poor/fair health when compared to urban respondents.

This analysis adjusted for several covariates, and thus produced important results past the impact of level of rurality on self-reported health. We found an approximately 4.5% increase in odds of reporting poor/fair health for each one-year increase in age throughout all models. When compared to White respondents, Asian respondents consistently had the lowest odds of poor health (unadjusted OR: 0.63, 95% CI: 0.50 – 0.80), while American Indian/Alaska Native and any other race respondents consistently had the highest odds of poor health (unadjusted OR: 2.50, 95% CI: 1.79 – 3.50). The relationship between race and self-reported poor health was greatly impacted by the addition of adjustment for infrastructure availability across all races, with odds ratios consistently biased strongly towards the null when considering this adjustment. When compared to respondents living at 1.0 – 4.0 times the poverty threshold, those living below the poverty threshold had 149% higher odds of reporting poor/fair health, while those living at 5.0 times or more above the poverty threshold had a 64% lower odds of reporting poor/fair health.

Lastly, individuals that did not have a usual place for care had a 27% lower odds of reporting poor/fair health when compared to individuals that had one or more usual places for care.

Table 1. Descriptive statistics (percent distributions and means) by level of rurality.

	Total N = 20,488	**Urban** N = 6,016	**Suburban** N = 4,772	**Medium & small metro** N = 6,605	**Rural** N = 3,095
	% or mean	% or mean	% or mean	% or mean	% or mean
Poor/fair health	12.29	11.22	9.53	12.61	17.90
Age (mean, (sd))	43.14 (13.19)	41.95 (12.98)	44.13 (13.00)	42.95 (13.33)	44.36 (13.35)
Female	52.4	51.36	51.7	53.88	52.34
Race					
White only	77.49	66.90	80.22	81.15	86.04
Black/African American only	12.70	18.37	10.62	11.46	7.53
American Indian/Alaska Native only	0.98	0.86	0.40	0.68	2.75
Asian only	6.29	11.35	6.81	3.89	0.81
Other and multiple race	1.55	1.89	1.28	1.65	1.07
American Indian/Alaska Native and any other race	0.99	0.62	0.67	1.17	1.81
Education					
Less than high school	6.79	5.93	4.88	6.95	11.08
High school degree or equivalent	24.41	19.03	22.23	26.57	33.63
Some college; Associate degree	30.40	26.83	29.59	32.78	33.51
Bachelors degree	24.09	29.84	26.80	21.59	12.09
More than a bachelors degree	14.30	18.37	16.49	12.11	7.69
Poverty					

1.0 or under	10.95	10.80	6.33	12.05	15.99
1.0 – 4.0	44.98	41.04	38.52	49.13	53.76
4.0 or over	44.07	48.15	55.16	38.82	30.24
Health Insurance Coverage Status					
Not covered	11.41	10.80	9.22	11.66	15.41
Covered, private	71.09	73.32	77.28	68.22	63.33
Covered, Medicaid	11.64	11.15	8.82	12.73	14.57
Other	5.86	4.70	4.67	7.39	6.69
Had a usual place for care					
No	11.30	13.00	10.67	10.70	10.27
Yes	88.68	87.00	89.33	89.30	89.72
Food security raw score (mean)	0.58	0.57	0.44	0.60	0.75

Table 2. Results of logistic regression models examining the odds of poor health among suburban, medium and small metro, and rural respondents compared to urban respondents.

	Suburban	Medium and Small Metro	Rural
Model 1 Level of rurality	0.83 (0.74 – 0.95)	1.14 (1.03 – 1.28)	1.73 (1.53 – 1.95)
Model 2 Rurality, region	0.84 (0.74 – 0.95)	1.14 (1.03 – 1.28)	1.73 (1.53 – 1.95)
Model 3 Rurality, demographics	0.80 (0.71 – 0.91)	1.13 (1.01 – 1.26)	1.64 (1.44 – 1.86)
Model 4 Rurality, demographics, socioeconomics	0.88 (0.77 – 1.00)	0.94 (0.84 – 1.06)	1.13 (0.99 – 1.29)
Model 5 Rurality, demographics, socioeconomics, infrastructure	0.88 (0.77 – 1.01)	0.94 (0.83 – 1.06)	1.16 (1.00 – 1.33)

Discussion

This analysis investigated the relationship between level of rurality and self-reported poor health, including adjustments for demographic identity, socioeconomic status, and infrastructure availability. We found that individuals residing in suburban areas had the lowest odds of poor health while individuals residing in rural areas had the highest odds of poor health. Overall, when compared to individuals residing in urban areas, increasing rurality was associated with increasing odds of poor health. This trend remained consistent after adjusting for all covariates.

Consistent with prior analyses, we found a statistically significantly higher odds of poor health in rural residents[2,11,13] and lower odds of poor health in suburban residents.[10] Generally, this analysis found a similar odds ratio of about 70% higher odds of poor health for rural residents when compared to urban residents that other studies with similar analysis methods found.[13] However, studies with more complex analysis methods found an attenuated odds ratio of about 40% higher odds of poor health for rural residents when compared to urban residents.[11]

Our adjusted models reveal some hints about the mechanisms underlying the relationship between level of rurality and self-reported poor health. We found that adjusting for demographic characteristics attenuated the relationship between level of rurality and self-reported poor health slightly, indicating that some portion of the effect size of level of rurality may be attributed to the impact demographic characteristics have on self-reported poor health. Similarly, we see that the relationship between self-reported poor health and level of rurality is greatly attenuated in model 4 suggesting that the majority of the association between self-reported poor health and

level of rurality may be explained by the relationship between socioeconomic status and self-reported poor health. Lastly, we can see that adjusting for infrastructure did not have a great impact on this relationship after adjusting for socioeconomic and demographic characteristics (i.e. model 5), leading us to infer that infrastructure may not be a major mechanism impacting the relationship between self-reported poor health and level of rurality if we have also considered socioeconomic characteristics. From this, we may understand that the greatest mechanism underlying the relationship between level of rurality and self-reported poor health are the societal mechanics in the United States that place importance on class, wealth, and economic status.

A major strength of this study is that it is generalizable to the United States population. The NHIS uses a sampling methodology that is nationally representative and trusted to be accurate, and thus the conclusions made here can be generalized to the population of the United States. Additionally, we used 2019 data, which is the most recent data that was not impacted by the COVID-19 pandemic. These results show a good picture of the relationship between level of rurality and self-reported poor health given the current environmental, social, and political landscape but not impacted by the economic and health impacts of the COVID-19 pandemic.

The first significant limitation of this analysis is that it used cross sectional data – because of this, we are not able to make any inferences on the causality of this relationship from this analysis. Additionally, the data utilized for this study did not include homeless individuals, which are an important population for this type of analysis. Due to the nature of surveying homeless individuals, there is not an agreed upon number of homeless individuals in the United States. However, it is estimated that on a given night there are between 500,000 and 600,000 people experiencing homelessness in the United States.[14] This is important to this analysis because the experience of homelessness varies greatly between an urban environment and a rural environment, and it is essential to capture this difference. Lastly, no sensitivity analyses were performed in regard to the level of rurality. There are 11 different commonly used classification schemes for level of rurality, each with their own pros and cons.[15] Ideally, we would have been able to perform a sensitivity analysis to determine if the relationship between level of rurality and self-reported poor health remains similar when using different classification schemes with different levels of specificity.

While this analysis echoed previous research on the relationship between level of rurality and self-reported poor health, it did add a significant component: inclusion of suburban residents. Very little literature on this topic in the United States includes classifications of rurality between urban and rural, instead most research focuses on disparities in self-reported health between rural and urban residents. Here, we showed a significant association between residing in a suburban area and self-reported poor health. Additionally, there are not many analyses using self-reported health as an outcome, and this analysis adds to the scant literature.

There are two future research directions that we would recommend based on the results of this analysis. The first is to repeat this analysis using a different classification scheme for rurality. The second is to repeat this analysis using data that has more specific rurality information: that is, instead of having individuals designated only as what level of rurality they reside in, a county or zip code is provided. While this analysis found that region of the United States did not impact the relationship between level of rurality and self-reported poor health, policies that impact health (i.e. health, economic, transportation, or education policy among others) and factors that impact health (i.e. environmental, social, etc.) vary greatly from state to state. We believe that a rural resident in the Pacific Northwest may have a very different experience than a rural resident in the Deep South. Thus, incorporating this varied experience into this type of analysis is essential to truly capture the extent of place-based health disparities.

References

1. About Rural Health | CSELS | OPHSS | CDC. Published March 25, 2020. Accessed February 12, 2023. https://www.cdc.gov/ruralhealth/about.html
2. Rothwell CJ, Madans JH, Arispe IE. 2013 NCHS Urban-Rural Classification Scheme for Counties.
3. Increasing Mortality From Premature Coronary Artery Disease in Women in the Rural United States. doi:10.1161/JAHA.119.015334
4. Coughlin SS, Clary C, Johnson JA, et al. Continuing Challenges in Rural Health in the United States. Published online 2020.
5. Croft JB, Wheaton AG, Liu Y, et al. Urban-Rural County and State Differences in Chronic Obstructive Pulmonary Disease — United States, 2015. *MMWR Morb Mortal Wkly Rep*. 2018;67(7):205-211. doi:10.15585/mmwr.mm6707a1
6. Zhao G, Okoro CA, Hsia J, Garvin WS, Town M. Prevalence of Disability and Disability Types by Urban–Rural County Classification—U.S., 2016. *Am J Prev Med*. 2019;57(6):749-756. doi:10.1016/j.amepre.2019.07.022
7. Curtin S, Spencer MR. *Trends in Death Rates in Urban and Rural Areas: United States, 1999–2019*. National Center for Health Statistics (U.S); 2021. doi:10.15620/cdc:109049
8. Garcia MC. Potentially Excess Deaths from the Five Leading Causes of Death in Metropolitan and Nonmetropolitan Counties — United States, 2010–2017. *MMWR Surveill Summ*. 2019;68. doi:10.15585/mmwr.ss6810a1
9. Singh GK, Siahpush M. Widening Rural–Urban Disparities in Life Expectancy, U.S., 1969–2009. *Am J Prev Med*. 2014;46(2):e19-e29. doi:10.1016/j.amepre.2013.10.017
10. The Importance of Place of Residence: Examining Health in Rural and Nonrural Areas. Accessed February 8, 2023. https://ajph.aphapublications.org/doi/epub/10.2105/AJPH.94.10.1682
11. Monnat SM, Beeler Pickett C. Rural/urban differences in self-rated health: examining the roles of county size and metropolitan adjacency. *Health Place*. 2011;17(1):311-319. doi:10.1016/j.healthplace.2010.11.008
12. NHIS - About the National Health Interview Survey. Published March 3, 2022. Accessed February 17, 2023. https://www.cdc.gov/nchs/nhis/about_nhis.htm
13. Bethea TN, Lopez RP, Cozier YC, White LF, McClean MD. The Relationship Between Rural Status, Individual Characteristics, and Self-Rated Health in the Behavioral Risk Factor Surveillance System. *J Rural Health*. 2012;28(4):327-338. doi:10.1111/j.1748-0361.2012.00414.x
14. The Size and Census Coverage of the US Homeless Population. BFI. Accessed March 29, 2023. https://bfi.uchicago.edu/insight/finding/the-size-and-census-coverage-of-the-us-homeless-population/
15. Bennett KJ, Borders TF, Holmes GM, Kozhimannil KB, Ziller E. What Is Rural? Challenges And Implications Of Definitions That Inadequately Encompass Rural People And Places. *Health Aff (Millwood)*. 2019;38(12):1985-1992. doi:10.1377/hlthaff.2019.00910