Concordance of Rural-Urban Self-identity and ZIP Code-derived Rural-Urban Commuting Area (RUCA) Designation

Tracy Onega, PhD, MA, MS[1,2,3,4], **Julie E. Weiss, MS**[1,3], **Jennifer Alford-Teaster, MPH, MA**[1,3,4], **Martha Goodrich, MS**[1,3], **M. Scottie Eliassen, MS**[5], **Sunny Jung Kim, PhD, MS, MA**[6,7]

[1]Department of Biomedical Data Science, Geisel School of Medicine at Dartmouth, Lebanon, New Hampshire

[2]Dartmouth Institute for Health Policy and Clinical Practice, Geisel School of Medicine at Dartmouth, Lebanon, New Hampshire

[3]Norris Cotton Cancer Center, Lebanon, New Hampshire

[4]Department of Epidemiology, Geisel School of Medicine at Dartmouth, Lebanon, New Hampshire

[5]Departments of Community and Family Medicine and Medical Education, Geisel School of Medicine at Dartmouth, Lebanon, New Hampshire

[6]Department of Health Behavior and Policy, School of Medicine, Virginia Commonwealth University, Richmond, Virginia

[7]Massey Cancer Center, Virginia Commonwealth University, Richmond, Virginia

Abstract

Purpose: This study examined the concordance between individuals' self-reported rural-urban category of their community and ZIP Code-derived Rural-Urban Commuting Area (RUCA) category.

Methods: An internet-based survey, administered from August 2017 through November 2017, was used to collect participants' sociodemographic characteristics, self-reported ZIP Code of residence, and perception of which RUCA category best describes the community in which they live. We calculated weighted kappa (κ) coefficients (95% confidence interval (CI)) to test for concordance between participants' ZIP Code-derived RUCA category and their selection of RUCA descriptor. Descriptive frequency distributions of participants' demographics are presented.

Findings: 622 survey participants, residents of NH (63%) and VT (37%), responded to the survey's self-reported rural-urban category. The overall κ was 0.33 (95% CI: 0.27–0.38). The highest concordance was found among those living in a small rural area (N=81, 13%): 62% of this group identified their communities as small rural. 65% (300/459) of participants residing in urban

or large rural areas reported their community as more rural (small rural or isolated). 68% (111/163) of participants living in small rural or isolated areas identified their community as more urban (large rural or urban).

Conclusions: Discordance was found between self-report of rural-urban category and ZIP Code-derived RUCA designation. Caution is warranted when attributing rural-urban designation to individuals based on geographic unit, since perceived rurality/urbanicity of their community that relates to health behaviors may not be reflected.

Keywords

concordance; health behaviors; rural communities; rural population; rural-urban

---

The construct of "rural" is multi-dimensional and theoretically complex.[1] Multi-disciplinary literature on rurality provides insight into the variation in the demography, economics, culture, and environmental characteristics of rural regions and populations. To address this variation, numerous taxonomies have been developed to categorize rural-urban in relation to the purpose and use of the data.[1]

The Rural-Urban Commuting Area Codes (RUCAs) taxonomy, developed at the University of Washington in conjunction with several state and federal agencies, is based on population density, urbanization, and commuting patterns at the census tract and ZIP Code levels.[2,3] A 4-tier classification scheme (urban, large rural, small rural, and isolated) has been developed that is frequently used to attribute a rural-urban context to individuals based on the RUCA of the census tract or ZIP Code in which their residence is located.

Three main issues arise from concordance of ZIP Code-based rural-urban designation with self-reported rural-urban category: 1) Variability in measurement across geographic scales; 2) Multi-dimensionality of rurality as a construct; and 3) How well current rurality measures capture potential health-related mechanisms.

Epidemiologic, behavioral, and other health-related studies often use a rural-urban contextual measure to either adjust for or directly estimate the effect of rurality/urbanicity. For contextual measures that are purely tied to physical location, such as environmental exposures, *perception* of one's own rurality/urbanicity may not influence health outcomes, compared to residence-based rural-urban classification, which would measure physical proximity. For health-related behaviors, however, individual perception versus residential categorization could be part of the causal pathway for a given health behavior.

An increasing body of evidence demonstrates significant differences in health behaviors and outcomes along the rural-urban continuum, with more rural populations typically having higher rates of risk, disease, and mortality.[4–7] Using the Behavioral Risk Factor Surveillance System (BRFSS), rural-urban differences in health behaviors have been shown for: sufficient sleep, current nonsmoking, nondrinking or moderate drinking, maintaining normal body weight, and meeting aerobic leisure time physical activity recommendations.[5] These health behaviors are strongly implicated in cancer incidence, mortality, and other cancer-related burdens. Those health behaviors that are associated with rurality are key foci for cancer

prevention efforts and cancer survivorship activities, including in rural communities.[6,8] However, there is no literature on the relationship between people's perception of their rural-urban designation and their actual RUCA designation. Given the increasing role that rurality is playing in national health dialogues, especially in cancer,[9] it is important to assess how well attribution of the rural-urban context may or may not be capturing factors based on individual perceptions of rurality. Therefore, we examined the degree of concordance between self-report of community RUCA description and the RUCA category attributed to the ZIP Code of residence. This relationship between individual perception and residentially assigned RUCA category will inform the use of rural-urban attribution to measure context in epidemiological, behavioral, and health care delivery studies.

Methods

Survey and Cohort

This study was conducted at the Norris Cotton Cancer Center (NCCC) located in Lebanon, New Hampshire (NH), which serves a predominantly rural area covering Vermont (VT) and NH. A 139-item Internet-based survey, which led to 482 variables in the dataset, was implemented from August 2017 through November 2017 to assess sociodemographic, behavioral, health information seeking, and geographic characteristics. The survey involved selected questions from national surveys (Health Information National Trends Survey,[10] Behavioral Risk Factor Surveillance System,[11] and National Health Interview Survey[12]), as well as novel, study-specific questions generated by our research team. The sampling frame included residents of NH and VT as identified by Amazon Mechanical Turk (MTurk, Amazon Mechanical Turk, Inc., Seattle, Washington) and verified by self-report. MTurk is an online community in which requesters distribute various tasks and anonymous workers complete the tasks for compensation at a set rate. MTurk doesn't publicly release the demographic information of their workers or the size of their marketplace. A Pew Research Center report stated that there were 750,000 unique visitors in December 2015 alone, and 500,000 registered anonymous workers worldwide. Since its adoption among researchers in 2006 for research purposes, MTurk has been used for a wide range of human tasks related to research activities, such as psychological experiments, online surveys, coding for natural language processing, audio transcription, and image coding to name a few.[13–15] Researchers generated an online platform called "MTurk-tracker"[16] that tracks and live releases demographic information of workers (eg, year of birth, gender, household size). Authors of this work demonstrated heterogeneity of workers in demographic propensities over time.[17] Two studies report that the quality of the data collected from MTurk is comparable to that from other convenience samples such as undergraduate students or professional panels.[17,18]

We collected data from 702 respondents reached via MTurk. Eligibility (adults over 18 and full-time residents of NH or VT) was assessed via screening and basic demographic questions. Those who responded to our survey advertisements but were not deemed eligible by these metrics were not offered the opportunity to complete the survey. Those who were eligible to participate in the study were asked to report the ZIP Code of their residence. Participants were not asked to report their names, social security numbers, or any identification information. The collected information is not personally identifiable.

We conducted a pilot test to determine the time needed to complete the survey. The range was from 35 minutes to over 50 minutes depending on the survey logistics applied to each participant upon their responses. Based on the usual rate of compensation on MTurk ($0.10 per minute), we paid $5 per survey participant to provide reasonable compensation. However, we didn't overcompensate MTurk workers for our survey (eg, paying $15) to prevent a monetary incentive being a motivation for participation and to avoid potential deception by workers while answering eligibility questions.

Measures

Survey participants were asked to "Please select the category that you think best describes the community where you live" with the response values being the 4-tiered RUCA scale developed at the University of Washington (urban, large rural, small rural, and isolated).[3] These 4 categories are derived from 10 major categories, and they are defined in part by commuting patterns (30%, 10%–30%, or < 10% commuting flow to urbanized areas, along with population estimates: 50,000 [urban], 10,000–49,999 [large town], 2,500–9,999 [small town], and <2,500 [isolated rural].[19] The survey did not provide a definition of the 4 terms, largely because we wanted the respondents' perceptions of rurality-urbanicity, not an estimation of the number of people within their community. We derived the RUCA designation for each participant from their self-reported ZIP Code using the same 4-tiered RUCA scale provided to the participants.

Analysis

Of the 702 NH and VT residents who completed the online survey, 622 participants (63% NH, 37% VT) responded to the self-reported RUCA descriptor question. Frequency distributions (N, %) of the survey participant demographics, specifically, age group (18–34, 35–54, 55+), gender (male, female), race/ethnicity (white, black, Hispanic, other), level of education (less than high school to high school graduate, some college, college, post graduate), type of insurance (private, Medicare, Medicaid, other), and income level (less than $50,000 and $50,000 or more) by RUCA designation and self-report of RUCA category descriptor are reported (Table 1). The overall weighted (Fleiss-Cohen) kappa coefficient (κ) and 95% confidence interval (CI) for the measure of agreement are presented. The analyses were performed in SAS (SAS 9.4 System Options: Reference, 2nd ed; 2011. SAS Institute Inc., Cary, North Carolina).

Results

More than half of the 622 participants were aged 35 years and older (58%), white (91%), female (69%), had college or post graduate education (51%), had private insurance (64%), and reported an income of less than $50,000 (54%) [Table 1]. According to the ZIP Code-derived RUCA category, the majority (74%) of the participants lived in an urban area (N=342) or large rural area (N=117), while 26% lived in a small rural area (N=81) or an isolated area (N=82). Almost half of the participants (48%) described the community in which they lived as small rural (N=300), 21% (N=132) as urban, 29% (N=179) as large rural, and only 2% (N=11) as isolated [Table 1].

The overall κ was 0.33 (95% CI: 0.27–0.38), indicating low agreement between ZIP Code-derived RUCA category and perceived rural-urban designation. To further assess the measure of agreement, an n x n (n= 622 participants) agreement plot[20] was created [Figure 1]. The 4 rectangles within the outer square represent the marginal totals (row x column) of the concordance table for each response category (urban, large rural, small rural, isolated). For example, the urban rectangle dimensions (width by height) are 342 by 132 (bottom left corner of Figure 1). The degrees of shading, dark to white, within each response rectangle depict exact and partial agreements. The darkest shaded squares represent exact agreement (the concordance table's diagonal cell frequencies). Rectangles of decreasing shades indicate partial agreement of cells further away from the diagonal (exact agreement) cells of the concordance table. Deviation of the rectangles away from the 45-degree diagonal line indicates differences in response preferences of the self-report of community identity to the ZIP Code-derived RUCA category.

Those living in a small rural area were more likely to self-describe their community concordantly with the RUCA category for their community (62%) than those living in the other RUCA categories (urban (33%), large rural (27%), and isolated (1%)) [Figure 1 table - exact agreement]. Among survey participants who lived in an urban or large rural area, 65% described their community as more rural [Figure 1 table – partial agreement, which is defined as one-category of discordance on an ordinal scale]. Of those who resided in small rural or isolated areas, 68% described their community as more urban. Only 1% of those living in an isolated RUCA derived from ZIP Code self-reported their community as isolated [Figure 1 table].

A large portion (78%) of the urban rectangle and the entire large rural rectangle located below and to the right of the 45-degree diagonal line in Figure 1 indicate a shift of self-described community identity towards more rural, and the isolated rectangle's position above and to the left of the diagonal indicates a shift towards more urban.

Discussion

This study provides some of the first empirical evidence that residential-based rural-urban attribution to individuals is not well aligned with individuals' perceptions of their own rural-urban context. Only a third of individuals perceived their rural-urban context concordantly with the RUCA category of their ZIP Code of residence. The tendency to perceive one's community differently from its classification was similar for both urban and rural residents: urban residents tended to view their communities as less urban, while rural residents tended to view their communities as less rural. Notably, the most rural residents (living in a ZIP Code designated as isolated) composed 13% of the sample, but only 1% of the sample reported their community as isolated. Thus, many residents who would be classified as isolated by ZIP Code-based measures did not view themselves as living in an isolated community. These findings are important given the prevalence of attributing rural-urban designation based on ZIP Code or county of residence in epidemiologic, behavioral, and health care delivery research. Such attribution is sometimes intended as a proxy measure for individual-level contextual effects as potential factors, such as accessibility, social isolation, timely information/current knowledge, and inactivity in causal pathways related to such

domains as health behaviors, access to health care services, socio-psychological needs, and psychosocial environment. Understanding the concordance of measured attributes (ZIP Code-based rural-urban category) with individual perceptions that may drive actual mechanisms of effect is crucial to improving how we measure context.

Three main issues arise from concordance of ZIP Code-based rural-urban designation with self-reported rural-urban category: 1) Variability in measurement across geographic scales; 2) Multi-dimensionality of rurality as a construct; and 3) How well current rurality measures capture potential health-related mechanisms. Measurement across geographic scales is fraught with the well-known challenge of the modifiable areal unit problem (MAUP) in which data aggregated at different geographic levels, such as county, ZIP Code, census tract, etc., will often have different results even when the same analysis is applied.[21,22] Rural-urban classification systems are typically defined at the county, ZIP Code, or census tract levels, which may obscure heterogeneity in population measures and may not accurately capture individual measures at all. For health-related studies, following the maxim of using the definition of rurality most suited to a given study[23] may be apt; however, given how little we know about mechanistic components of "rurality" related to health behaviors, this approach does not offer adequate specificity or guidance. The health literature has many examples of variation in health-related behaviors and outcomes by rural-urban designation. For example, risk behaviors such as smoking, obesity, and low physical activity have been shown to be higher in rural areas.[4,5] Higher rates of mortality and disease incidence, such as for cancer, have been well documented in rural areas,[4,7,9] and psychosocial differences in health-related beliefs have been shown.[24]

The findings from this study point to the need for caution when attributing a rural-urban designation to individuals based on a group-level measure, such as ZIP Code level RUCA category, particularly if trying to capture factors that may be influenced by individual perception. Additional measures of rurality should be developed to better capture dimensions of rurality that may only be able to be ascertained at the individual level. One such example is suggested by our finding of only 1% of the population reporting living in an isolated rural community, although according to ZIP Code-based measures, 13% did. Thus, a measure of rurality that accounts for perception of isolation could be important, particularly if social networks, support systems, or other aspects of connectedness are relevant to health behaviors, quality of life, or health services utilization. Few such measures exist, although at least one neighborhood and community rural perception scale is seen in some social science or health-related literature.[1,8,25–31] This scale used Likert responses to measure neighborhood cohesion (5 questions), community identity (6 questions), and rural identity (6 questions) to help understand factors related to health beliefs, cancer screening, obesity, and other outcomes among rural residents.[8,29–31] Developing and testing measures of rurality that better reflect individuals' attitudes, perceptions, and behaviors is likely to advance what we can learn about health influences in rural areas, which in turn can facilitate targeted interventions.

Limitations

This study is novel but has several limitations to note. First, we used a limited geographic region (NH and VT) that is almost 50% rural but may not represent other regions. Second, there may be response bias due to the study population being drawn from online platforms. Using any online survey method may exclude households that do not have access to a personal computer and/or access to the Internet. As in many other studies that used online recruitment, using MTurk to recruit study participants generated a crowd-sourced convenience sample based on non-probabilistic sampling. This method had considerable benefits compared to other types of convenience sampling methods, such as community/school-based approaches and recruiting participants from designated locations (eg, surveying customers at a local store). MTurk sampling was an expeditious and cost-effective method that was optimal for our catchment areas and study time frame. We note that, consistent with other online surveys, our respondents included more females than males, and younger rather than older participants, indicating a possible sampling error. Yet other demographic characteristics did not substantially deviate from those of our catchment area population (VT and NH residents). To date, there is no standardized method of probabilistic sampling specifically designed for online surveys. Given the gender disproportion in our online survey data but considerable benefits of using online environments for recruitment, future research may investigate sampling research designs that can produce a generalizable sample with more representative demographic characteristics as well as the application of assigning weights to the sample.

Conclusion

In conclusion, we found only limited concordance between self-report of rural-urban category and ZIP Code-derived RUCA category. Therefore, one should use caution when attributing rural-urban designation to individuals solely based on geographic unit since individual perceptions that relate to health behaviors may not be reflected. Creation of validated rural-specific measurement tools is a critical need for health-related studies, particularly as research into the complex effects of rurality on health, behaviors, and outcomes advances.
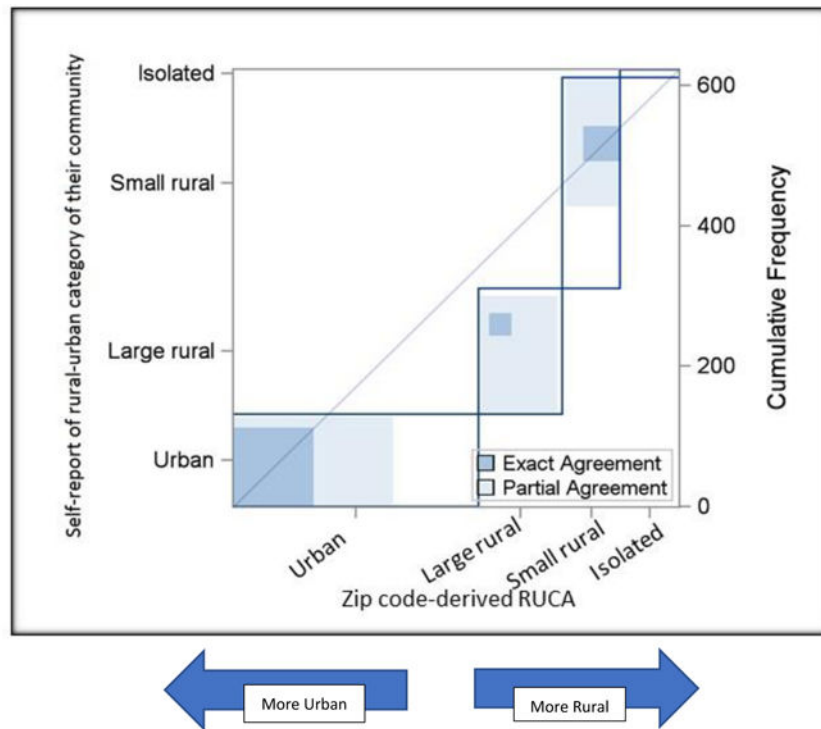
Acknowledgments

References

1. Hart LG, Larson EH, Lishner DM. Rural definitions for health policy and research. Am J Public Health 2005;95(7):1149–1155. [PubMed: 15983270]

2. United States Department of Agriculture, Economic Research Service. "Rural-Urban Commuting Area Codes." https://www.ers.usda.gov/data-products/rural-urban-commuting-area-codes/. Last accessed 9/1/2018.

3. University of Washington. http://depts.washington.edu/uwruca/. Last accessed 8/2/2018.

4. Henley SJ, Anderson RN, Thomas CC, Massetti GM, Peaker B, Richardson LC. Invasive Cancer Incidence, 2004–2013, and Deaths, 2006–2015, in Nonmetropolitan and Metropolitan Counties - United States. MMWR Surveill Summ 2017;66(14):1–13.

5. Matthews KA, Croft JB, Liu Y, et al. Health-Related Behaviors by Urban-Rural County Classification - United States, 2013. MMWR Surveill Summ 2017;66(5):1–8.

6. Weaver KE, Palmer N, Lu L, Case LD, Geiger AM. Rural-urban differences in health behaviors and implications for health status among US cancer survivors. Cancer Causes Control 2013;24(8):1481–1490. [PubMed: 23677333]

7. Zahnd WE, James AS, Jenkins WD, et al. Rural-Urban Differences in Cancer Incidence and Trends in the United States. Cancer Epidemiol Biomarkers Prev 2017;27(11):1265–1274. [PubMed: 28751476]

8. Rice EL, Patel M, Serrano KJ, Thai CL, Blake KD, Vanderpool RC. Beliefs About Behavioral Determinants of Obesity in Appalachia, 2011–2014. Public Health Rep 2018;133(4):379–384. [PubMed: 29791807]

9. Blake KD, Moss JL, Gaysynsky A, Srinivasan S, Croyle RT. Making the Case for Investment in Rural Cancer Control: An Analysis of Rural Cancer Incidence, Mortality, and Funding Trends. Cancer Epidemiol Biomarkers Prev 2017;26(7):992–997. [PubMed: 28600296]

10. National Cancer Institute. Health Information National Trends Survey (HINTS) http://hints.cancer.gov/default.aspx. Last accessed 8/2/2018.

11. Centers for Disease Control and Prevention. Behavioral Risk Factor Surveillance System (BRFSS). http://www.cdc.gov/brfss/index.html. Last accessed 8/2/2018.

12. Centers for Disease Control and Prevention. National Health Interview Survey (NHIS). http://www.cdc.gov/nchs/nhis/index.htm. Last accessed 8/2/2018.

13. Clery D Galaxy Zoo Volunteers Share Pain and Glory of Research. Science (New York, NY) 2011;333:173–175.

14. Khatib F, Cooper S, Tyka MD, et al. Algorithm discovery by protein folding game players. Proceedings of the National Academy of Sciences 2011;108:18949–18953.

15. von Ahn L, Maurer B, McMillen C, Abraham D, Blum M. reCAPTCHA: human-based character recognition via Web security measures. Science (New York, NY) 2008;321(5895):1465–1468.

16. Ipeirotis PG. Analyzing the Amazon Mechanical Turk marketplace. XRDS 2010;17(2):16–21.

17. Difallah D, Filatova E, Ipeirotis P. Demographics and Dynamics of Mechanical Turk Workers. Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining; 2018; Marina Del Rey, CA, USA.

18. Kees J, Berry C, Burton S, Sheehan K. An Analysis of Data Quality: Professional Panels, Student Subject Pools, and Amazon's Mechanical Turk. Journal of Advertising 2017;46(1):141–155.

19. Guidelines for Using Rural-Urban Classification Systems for Community Health Assessment https://www.doh.wa.gov/Portals/1/Documents/1500/RUCAGuide.pdf. Last accessed 1/8/2018.

20. Bangdiwala SI, Shankar V. The Agreement Chart. BMC Med Res Methodol 2013;13:97 10.1186/1471-2288-13-97. Last accessed 2/18/2019. [PubMed: 23890315]

21. Openshaw S The modifiable areal unit problem Norwick: GeoBooks; 1983 ISBN 0860941345. OCLC 12052482.

22. Kwan Mei-Po. The Uncertain Geographic Context Problem. Ann Assn Amer Geogr 2012;102(5):958–968.

23. Cloke P An index of rurality for England and Wales. Regional Studies B 1977;11(1):31–46. doi:10.1080/09595237700185041.

24. Vanderpool RC, Huang B. Cancer risk perceptions, beliefs, and physician avoidance in Appalachia: results from the 2008 HINTS Survey. J Health Commun 2010;15(Suppl 3):78–91. [PubMed: 21154085]

25. Gursoy D, Jurowski CA, Uysal M. Resident Attitudes: A Structural Modeling Approach. Annals of Tourism Research 2002;29 (1):79–105.

26. Chavis DM, Lee KS, Acosta JD. The sense of community (SCI) revised: The reliability and validity of the SCI-2. Paper presented at the 2nd International Community Psychology Conference, Lisboa, Portugal; 2008.

27. Chavis D, Hogge, McMillan D, Wandersman A Sense of community through Brunswick's lens: A first look. J Community Psych 1986;14(1):44–40.

28. Sampson RJ, Raudenbush SW, Earls F. Neighborhoods and violent crime: A multilevel study of collective efficacy. Science 1997;277(5328):918–924. doi: 10.1126/science.277.5328.918. [PubMed: 9252316]

29. Paskett ED, Hiatt RA. Catchment Areas and Community Outreach and Engagement: The New Mandate for NCI-Designated Cancer Centers. Cancer Epidemiol Biomarkers Prev 2018;27(5): 517–519. [PubMed: 29716925]

30. VanDyke SD, Shell MD. Health Beliefs and Breast Cancer Screening in Rural Appalachia: An Evaluation of the Health Belief Model. J Rural Health 2017;33(4):350–360. [PubMed: 27545099]

31. Vanderpool RC, Huang B, Shelton B. Seeking Cancer Information: An Appalachian Perspective. Journal of Health Disparities Research and Practice: 2008;2(1):Article 7. Available at: https:// digitalscholarship.unlv.edu/jhdrp/vol2/iss1/7

Figure 1.
Agreement of ZIP Code-derived RUCA and Self-report of Rural-Urban Category of Community

Table 1.

Distributions of Participants' Demographics by ZIP Code-derived RUCA and Self-report of Rural-Urban Category of Community (N = 622)

Participants' Demographics[a]	N	%
Age Group (years)		
18 – 34	259	41.6
35 – 54	198	31.8
55	165	26.5
Gender		
Male	193	31.0
Female	429	69.0
Race/Ethnicity		
White	565	90.8
Black	8	1.3
Hispanic	19	3.1
Other	30	4.8
Education		
< High school - HS	119	19.1
Some college	183	29.4
College	236	37.9
Post graduate	84	13.5
Insurance		
Private	360	64.5
Medicare	92	16.5
Medicaid	87	15.6
Other	19	3.4
Income		
< $50,000	338	54.3
$50,000	284	45.7
ZIP Code-derived RUCA category		
Urban	342	55.0
Large rural	117	18.8
Small rural	81	13.0
Isolated	82	13.2
Self-report of rural-urban category of community		
Urban	132	21.2
Large rural	179	28.8
Small rural	300	48.2
Isolated	11	1.8

[a]Missing (N): Insurance (64).

Abbreviations: RUCA: Rural-Urban Commuting Area; HS: high school.