Add Health

The National Longitudinal Study of Adolescent to Adult Health

Add Health as a Resource for the Science of the Exposome:

Rural-Urban Commuting Area (RUCA) Codes

Report prepared by

Andrea N. Goodwin

James D. Stewart

Robert A. Hummer

Eric A. Whitsel

This user guide is one in a set of user guides focusing on the built, environmental, and natural features of geopositioned/geocoded Add Health respondent locations over Waves I-VI. Collectively, they describe exposomic measures in the following three domains:

Built Domain	**Environmental Domain**	**Natural Domain**
Commuting Area	Ambient Air	Altitude
Land Use	Indoor Air	Meteorology
Roadway Proximity/Density	Noise	Green space
	Waterborne Lead	Blue space
	Nighttime Light Pollution	
	Solar Irradiation	

Under the Built Domain, this particular user guide summarizes the rationale for the latest construction and assignment of rural-urban commuting area (RUCA) codes. It also documents how the RUCA source data were acquired, as well as the protocol for quality controlling their assignment and classification across waves. Whenever possible, construction, assignment, and classification were harmonized to ensure temporal comparability, although important inter-wave differences exist and are grey-highlighted herein.

Acknowledgement

Citation for User Guide

Goodwin AN, Stewart JD, Hummer RA, Whitsel EA. Add Health as a Resource for the Science of the Exposome**:** *Rural-Urban Commuting Area (RUCA) Codes,* Wave VI Data Documentation*.* Chapel Hill, NC: Carolina Population Center, University of North Carolina at Chapel Hill. Available from: https://doi.org/10.17615/9ygq-g064

Table of Contents

1. Introduction

The National Longitudinal Study of Adolescent to Adult Health (Add Health) is a nationally representative sample of U.S. adolescents who were in grades 7-12 during the 1994-1995 school year. Using a complex, school-based cluster-sampling frame, researchers selected high school and feeder school pairs from 80 communities across the United States and drew a sex- and grade-stratified random sample of 20,745 adolescents for inclusion in the study. This sample has been followed from adolescence into early midlife across six waves of data collection to date, with the most recent wave of data collection (Wave VI) taking place between 2022 and 2025 when respondents were ages 39 to 49.

Over the years, Add Health has collected a wealth of information from respondents and their parents about demographic characteristics, familial structures, social relationships, health behaviors, cognition, physical and mental health status, medication usage, and health care access. Add Health also has collected anthropometric, cardiovascular, metabolic, renal, hepatic, inflammatory/immune, infectious, neurodegenerative, and multi-omic biomarkers from respondents. In addition, Add Health has merged multilevel contextual data about the economic, school, neighborhood, policy, and environmental contexts in which the respondents are embedded to the core survey and biological data at each wave. The Add Health dataset thereby provides researchers with rich opportunities to explore the causes and consequences of health status across multiple contextual domains as individuals age across the life course.

This user guide is one in a series documenting the latest contextual and environmental data assembled under the exposome supplement introduced in the preceding acknowledgment. Collectively, the supplemental data and documentation enable researchers to examine a broader array of built, environmental, and natural exposures linked to accurately geopositioned/geocoded Add Health respondent residences from Wave I through Wave VI. Because Wave VI data are not ready for geocoding or dissemination at present, this user guide and the associated data are focused on Wave I-V linkages. The Add Health Team will update this data set and user guide when Wave VI data are available for dissemination.

2. General Overview

Rural-urban commuting area (RUCA) codes classify U.S. census tracts using measures of population density, urbanization, and daily commuting. The data file including them is based on RUCA codes for census years 1990, 2000, and 2010. The rationale for and utility of acquiring RUCA codes, assigning them to census geographies in which Add Health respondents have resided over three decades, and then classifying them is described below.

2.1 Rationale

Since its inception, Add Health has continued amassing and disseminating contextual data files across multiple levels of geography, thus resulting in an increasingly comprehensive and diverse set of contextual measures in a nationally representative study spanning adolescence to mid adulthood. In general, these data have been provided to establish infrastructure for research addressing the role of diverse exposures across multiple levels and across the life course in the etiology and disparities of our most pressing health issues. The data collectively position Add Health as a central resource for scientists to more effectively operationalize and study the exposome and its consequences for population health across the life course, with particular attention to disparities across population subgroups.

2.2 Utility

The RUCA codes described herein expand the contextual data available to Add Health researchers, enhancing their capacity to examine the social, environmental, and biological dimensions of the exposome and how they contribute to U.S. population health and disparities. They differ from the RUCA codes available in previously disseminated Add Health datasets (W1RUCA; W3RUCA; TRACT3; TRACT4; CONTEXT5) in that they cover all periods of respondent follow-up, from enrollment (Wave I) to right-censoring (currently, December 31, 2015) or respondent dates of death, whichever came first. The RUCA codes may be valuable to researchers who study disparities in disease risk, onset, duration, awareness, treatment, control, and / or sequelae including mortality. In addition to the RUCA-related contextualization of disease-specific and all-cause morbidity and mortality, RUCA codes may enhance research centered on disparities in access to health-promoting and health care resources.

3. Processing Details

The RUCA codes and their classifications in this file are based on the 1990, 2000, and 2010 decennial censuses because RUCA 2020 source data had not been released at the time of data construction. The 1990, 2000, and 2010 Excel files of Census tract-level RUCA codes were downloaded from the United States Department of Agriculture Economic Research Service (USDA ERS) website[1] (Tables 1-3), and extraneous codes were removed. For compatibility purposes, the U.S. Census tract-level Federal Information Processing Standard (FIPS) codes contained in the files were reformatted so that their format would match those of the 1990, 2000, and 2010 U.S. Census tract-level FIPS codes associated with geocoded respondent residences (1994-2019). Then the three RUCA Excel files were converted to SAS data sets.

TIGER/Line geographic boundary files were downloaded from the U.S. Census Bureau for decennial census years 1990, 2000, and 2010. These census geography files were compared to the tract-level FIPS codes in the RUCA source data files to verify that they matched each other for the corresponding decennial census years. Next, a spatial join was executed to attach the tract-level census geography identifiers (FIPS codes) to Add Health respondent residential locations for each decennial census from 1990 to 2015. The RUCA codes for all available decennial years (1990, 2000, and 2010) were merged onto geocoded respondent residential locations by tract-level FIPS codes. Lastly, the RUCA code was selected from the decennial census year that corresponded most closely to the time period for which geographic coordinates were available (Table 4), and those RUCA codes were classified using a common aggregation schema (Tables 5-7).[2] In 2000, one secondary RUCA code that only occurred in a small number of tracts within the same county was changed to an adjacent RUCA code to reduce the risk of deductive disclosure.

Note that the 2010 RUCA codes were revised July 3rd, 2019 by the USDA ERS to correct a programming error that affected secondary RUCA codes for 10,909 of 74,002 tracts. The net result was an increase in the number of 2010 census tracts classified as rural. Primary RUCA codes were not affected.

3.1 Rural-Urban Commuting Area Codes [RUCA]

Table 1. RUCA Codes [**RMERUCA001**], 1990

Code	Description
1.0	Metropolitan-area code: Primary flow within an urbanized area (UA)
1.1	Secondary flow 30% to 50% to a larger UA
2.0	Metropolitan-area high commuting: primary flow 30% or more to a UA
2.1	Primary flow to a 1.0 UA
2.2	Primary flow to a 1.1 UA
3.0	Metropolitan-area low commuting: primary flow 5% to 30% to a UA
4.0	Large town core: primary flow within a place of 10,000 to 49,999
4.1	Secondary flow 30% to 50% to a UA
5.0	Large town high commuting: primary flow 30% or more to a place of 10,000 to 49,999 (4.0 lg town)
5.1	Primary flow to a 4.1 large town
6.0	Large town low commuting: primary flow 5% to 30% to a place of 10,000 to 49,999
7.0	Small town core: primary flow within a place of 2,500 to 9,999
7.1	Secondary flow 30% to 50% to a UA
7.2	Secondary flow 30% to 50% to a large town
7.3	Secondary flow 5% to 30% to a UA
7.4	Secondary flow 5% to 30% to a large town
8.0	Small town high commuting: primary flow 30% or more to a place of 2,500 to 9,999 (7.0 sm town)
8.1	Primary flow to a 7.1 small town
8.2	Primary flow to a 7.2 small town
8.3	Primary flow to a 7.3 small town
8.4	Primary flow to a 7.4 small town
9.0	Small town low commuting: primary flow 5% to 30% to a place of 2,500 to 9,999
9.1	Secondary flow 5% to 30% to a UA
9.2	Secondary flow 5% to 30% to a large town
10.0	Rural areas: primary flow to a tract without a place of 2,500 or more
10.1	Secondary flow 30% to 50% to a UA
10.2	Secondary flow 30% to 50% to a large town
10.3	Secondary flow 30% to 50% to a small town
10.4	Secondary flow 5% to 30% to a UA
10.5	Secondary flow 5% to 30% to a large town
99.0	Not coded: Tracts with little or no population and no commuting flows

Table 2. RUCA Codes [RMERUCA001], 2000

Code	Description
1.0	Metropolitan-area code: Primary flow within an urbanized area (UA)
1.1	Secondary flow 30% to 50% to a larger UA
2.0	Metropolitan-area high commuting: primary flow 30% or more to a UA
2.1	Secondary flow 30% to 50% to a larger UA
3.0	Metropolitan-area low commuting: primary flow 5% to 30% to a UA
4.0	Micropolitan area core: primary flow within an Urban Cluster of 10,000 to 49,999 (large UC)
4.1	Secondary flow 30% to 50% to a UA
4.2	Secondary flow 10% to 30% to a UA
5.0	Micropolitan high commuting: primary flow 30% or more to a large UC
5.1	Secondary flow 30% to 50% to a UA
5.2	Secondary flow 10% to 30% to a UA
6.0	Micropolitan low commuting: primary flow 10% to 30% to a large UC
6.1	Secondary flow 10% to 30% to a UA
7.0	Small town core: primary flow within an Urban Cluster of 2,500 to 9,999 (small UC)
7.1	Secondary flow 30% to 50% to a UA
7.2	Secondary flow 30% to 50% to a large UC
7.3	Secondary flow 10% to 30% to a UA
7.4	Secondary flow 10% to 30% to a large UC
8.0	Small town high commuting: primary flow 30% or more to a small UC
8.1	Secondary flow 30% to 50% to a UA
8.2	Secondary flow 30% to 50% to a large UC
8.3	Secondary flow 10% to 30% to a UA
8.4	Secondary flow 10% to 30% to a large UC
9.0	Small town low commuting: primary flow 10% to 30% to a small UC
9.1	Secondary flow 10% to 30% to a UA
9.2	Secondary flow 10% to 30% to a large UC
10.0	Rural areas: primary flow to a tract outside a UA or UC
10.1	Secondary flow 30% to 50% to a UA
10.2	Secondary flow 30% to 50% to a large UC
10.3	Secondary flow 30% to 50% to a small UC
10.4	Secondary flow 10% to 30% to a UA
10.5	Secondary flow 10% to 30% to a large UC
10.6	Secondary flow 10% to 30% to a small UC

Table 3. RUCA Codes [RMERUCA001], 2010

Code	Description
1.0	Metropolitan-area code: Primary flow within an urbanized area (UA)
1.1	Secondary flow 30% to 50% to a larger UA
2.0	Metropolitan-area high commuting: primary flow 30% or more to a UA
2.1	Secondary flow 30% to 50% to a larger UA
3.0	Metropolitan-area low commuting: primary flow 10% to 30% to a UA
4.0	Micropolitan area core: primary flow within an Urban Cluster of 10,000 to 49,999 (large UC)
4.1	Secondary flow 30% to 50% to a UA
5.0	Micropolitan high commuting: primary flow 30% or more to a large UC
5.1	Secondary flow 30% to 50% to a UA
6.0	Micropolitan low commuting: primary flow 10% to 30% to a large UC
7.0	Small town core: primary flow within an Urban Cluster of 2,500 to 9,999 (small UC)
7.1	Secondary flow 30% to 50% to a UA
7.2	Secondary flow 30% to 50% to a large UC
8.0	Small town high commuting: primary flow 30% or more to a small UC
8.1	Secondary flow 30% to 50% to a UA
8.2	Secondary flow 30% to 50% to a large UC
9.0	Small town low commuting: primary flow 10% to 30% to a small UC
10.0	Rural areas: primary flow to a tract outside a UA or UC
10.1	Secondary flow 30% to 50% to a UA
10.2	Secondary flow 30% to 50% to a large UC
10.3	Secondary flow 30% to 50% to a small UC
99.0	Census tract has zero population and no rural-urban identifier information

3.2 Table 4 RUCA Source Data Year [RMERUCAYR] Crosswalk

Table 4. RUCA Source Data Year [RMERUCAYR] Crosswalk

Time Period	RUCA Source Data Year
Up to Dec 1995	1990
Jan 1996 to Dec 2005	2000
Jan 2006 to Dec 2015	2010
Jan 2016 to Dec 2025	2020[*]

[*]RUCA 2020 source data had not been released as of May 2023. For details on replacement code used for these missing values, see the Missing Codes discussion, below.

3.3 RUCA Code Classification [RMERUCA002]

Table 5. RUCA Code Classification [RMERUCA002], 1990

Class	Description	RUCA Codes
1	Urban	1.0, 1.1, 2.0, 2.1, 2.2, 3.0, 4.1, 5.1, 7.1, 8.1, and 10.1
2	Large, Rural City / Town	4.0, 5.0, and 6.0
3	Small, Rural Town	7.0, 7.2, 7.3, 7.4, 8.0, 8.2, 8.3, 8.4, 9.0, 9.1, and 9.2
4	Isolated, Small, Rural Town	10.0, 10.2, 10.3, 10.4 and 10.5

Table 6. RUCA Code Classification [RMERUCA002], 2000

Class	Description	RUCA Codes
1	Urban	1.0, 1.1, 2.0, 2.1, 3.0, 4.1, 5.1, 7.1, 8.1 and 10.1
2	Large, Rural City / Town	4.0, 4.2, 5.0, 5.2, 6.0 and 6.1
3	Small, Rural Town	7.0, 7.2, 7.3, 7.4, 8.0, 8.2, 8.3, 8.4, 9.0, 9.1 and 9.2
4	Isolated, Small, Rural Town	10.0, 10.2, 10.3, 10.4, 10.5 and 10.6

Table 7. RUCA Code Classification [**RMERUCA002**], 2010

Class	Description	RUCA Codes
1	Urban	1.0, 1.1, 2.0, 2.1, 3.0, 4.1, 5.1, 7.1, 8.1 and 10.1
2	Large, Rural City / Town	4.0, 5.0 and 6.0
3	Small, Rural Town	7.0, 7.2, 8.0, 8.2 and 9.0
4	Isolated, Small, Rural Town	10.0, 10.2 and 10.3

4. Missing codes

When respondent residential location coordinates were missing and otherwise there was no way to assign a RUCA code, *-9990* was used to indicate missing RUCA and class codes. In the instance that respondent residential location coordinates were available, but RUCA data were missing, a replacement code of *-9992* was assigned. This was the case for respondent residential locations with date ranges beginning in January 2016, since RUCA 2020 data had not been released at the time of publishing this user guide (see Table 4). Additionally, a small number of respondents with coordinates outside of the United States and three instances in which the RUCA code assigned was 99.0 (see Tables 1 and 3), also were assigned the *-9992* replacement code.

5. Usage Note

5.1 Potential for Longitudinal Misapplication

Although the 1990, 2000, and 2010 RUCA Codes use the same primary classification scheme (1–10), they are not directly comparable because many census tracts were reconfigured during each decade. Also, temporal changes in census methodologies significantly affected the RUCA classifications. Between 1990 and 2000, for example, changes to methods for defining urban areas decreased rural population and territory, and between 2000 and 2010, the data source for daily commuting patterns switched from the decennial census (measuring one point in time during 2000) to the American Community Survey (providing a 5-year average during 2006 - 2010).[1,3] For this reason, caution must be exercised when leveraging repeated measures per respondent as if they were longitudinally harmonized.

6. Data File

6.1 Structure

The RUCA codes are provided as a multiple-records-per-respondent long file comprised of six variables (Table 8). The file contains one record for each of the 20,745 Add Health Wave I sample members at every time period during their follow-up as identified by a masked respondent identifier (AID), the date from (RMERUCADFR) and date to (RMERUCADTO) identifying the start and end of each period, right-censored on December 31, 2015 or the date of death, whichever came first. Please consult the accompanying codebook for additional details.

6.2 Contents

The RUCA Codes data file includes the variables below, which are described in the corresponding codebook documentation that also contains frequencies.

Variable Name	Variable Description
AID	Add Health Respondent ID
RMERUCADFR	Date From
RMERUCADTO	Date To
RMERUCAYR	RUCA Source Data Year
RMERUCA001	RUCA Code
RMERUCA002	RUCA Code Class

Table 8. Data Contents

Variable	Description	Type	Format
AID	Add Health Respondent ID	character	L#####
RMERUCADFR	Date From	date	MM/DD/YYYY
RMERUCADTO	Date To	date	MM/DD/YYYY
RMERUCAYR	RUCA Source Data Year	numeric	YYYY
RMERUCA001	RUCA Code	numeric	NA
RMERUCA002	RUCA Code Class	numeric	NA

7. References

1. United States Department of Agriculture. Economic Research Service. Rural-Urban Commuting Area Codes. Overview. Available at https://www.ers.usda.gov/data-products/rural-urban-commuting-area-codes/.

2. WWAMI RUCA Rural Health Research Center. RUCA Version 1.11. Using RUCA codes: some aggregation schemes for health-related work. Categorization A. Available at https://depts.washington.edu/uwruca/ruca1/ruca-uses11.php.

3. United States Department of Agriculture. Economic Research Service. Rural-Urban Commuting Area Codes. Documentation. Available at https://www.ers.usda.gov/data-products/rural-urban-commuting-area-codes/documentation/.