

2021

Add Health Wave IV Documentation



Report prepared by

Elizabeth Tung, MD, MS

Add Health

Wave IV City Crime Rate Data User Guide



CAROLINA POPULATION CENTER | CAROLINA SQUARE - SUITE 210 | 123 WEST FRANKLIN STREET | CHAPEL HILL, NC 27516

Add Health is supported by grant P01-HD31921 from the Eunice Kennedy Shriver National Institute of Child Health and Human Development, with cooperative funding from 23 other federal agencies and foundations.

Acknowledgments

Construction of this data file was made possible with the fundamental support and contributions of the Add Health staff and the Carolina Population Center investigators of The National Longitudinal Study of Adolescent to Adult Health, especially Kathleen Mullan Harris. This work was funded by a K23 grant (PI: Tung, 1K23HL145090-01) provided by The National Heart, Lung, and Blood Institute (NHLBI) and a University of Chicago Data Discovery Grant.

Introduction

This Crime Rate Data file facilitates research examining the impact of community violence on the health trajectories of Add Health participants. Although Add Health disseminates crime statistics provided by the Uniform Crime Report (UCR); the UCR data are measured at the county-level. To add precision, the Crime Rate Data file uses data assembled by the Crime Open Database (CODE) which reports geocoded crime statistics at the point level (i.e., latitude and longitude). Based on police department crime data from 13 U.S. cities with high rates of crime, these source data enable the construction of average crime rates for these limited number of cities inhabited by Add Health participants at Wave IV. Merging Add Health data to geocoded crime rates by census tract permits tract-level analyses of a participant's exposure to community violence during years 2008-2010.

Data Structure and Form

These data contain one observation for each respondent in the Wave IV survey. The first variable is the respondent identifier (the AID), which permits merging these crime data with other Add Health data files. All other variables report average crime rates.

With the exception of AID, all variables in the city crime data adhere to the following nomenclature:

1st character – Summary level of the variable.

The character *C* refers to *cities' tracts*. Cities serve as the geographic area represented in the source data, but, note, these data are associated with Add Health participant locations by census tract.

2nd, and 3rd characters – Data file Content.

The characters *CR* denotes *crime rate*.

4th and 5th characters – Year.

The next two digits refer to the year of the crime data measures. All variables use 08 to denote 2008, even when averages also include estimates from 2009 and/or 2010.

6th and 7th characters – Crime classification.

The crime type or classification measured is identifiable by the next two characters:

01 = Total crimes

02 = Violent crimes

03 = Property crimes

8th character – Years represented.

The final character distinguishes variables by the number of years included in the average's construction.

1 = 2008

2 = 2008 and 2009

3 = 2008, 2009, and 2010

Data Dictionary

The variables in this file are clustered by the classification of crime, and within each cluster, the 3-year average, 2-year average, and 1-year average are reported, in that order.

Name	Description
CCR08013	City Tract-level average total crime rate 3-year estimate, W4
CCR08012	City Tract-level average total crime rate 2-year estimate, W4
CCR08011	City Tract-level average total crime rate 1-year estimate, W4
CCR08023	City Tract-level average violent crime rate 3-year estimate, W4
CCR08022	City Tract-level average violent crime rate 2-year estimate, W4
CCR08021	City Tract-level average violent crime rate 1-year estimate, W4
CCR08033	City Tract-level average property crime rate 3-year estimate, W4
CCR08032	City Tract-level average property crime rate 2-year estimate, W4
CCR08031	City Tract-level average property crime rate 1-year estimate, W4

The Crime Open Database (CODE) was used to develop crime rate estimates. Both, 2 and 3-year estimates were developed for Wave IV with the beginning of each time period corresponding to the beginning of data collection for Wave IV of Add Health. As such, CODE data from 2008-2010 were used in the construction of the Wave IV City Crime Rate Data.

Missing codes

These data include 3 types of missing codes. They are as follows:

Value	Reason for missing data
9/9992	city/tract not represented in source
9/9996	data missing in source/suppressed
9/9998	geocode missing

The replacement codes of 9992 and 99992 indicate that the tract inhabited by the Add Health participant did not comprise one of the cities measured by the source data. Codes of 9996 and 99996 report missing values for the tracts comprising represented cities. Census tracts that were officially within city boundaries may have had missing data for any of the following reasons: (1) Census tracts contained a total population of less than 100 residents (i.e., unable to calculate precise rates for these tracts), (2) Census tracts were outside of police department boundaries (e.g., census tract was policed by a nearby city police department), (3) Census tracts were only partially within police department boundaries (e.g., part of the census tract was policed by the city police department and part was policed by a neighboring police department), or (4) Census tracts were policed by a private police department (e.g., college campus police). Finally, values of 9998 and 99998 indicate that the Add Health participant's geocode is missing at Wave IV.

Source Description

The data used in the construction of Add Health's City Crime Rate Data file originates from The Crime Open Database (CODE), which compiles open-data license crime data from a limited number of police departments in the U.S. At the time of this file's production, CODE included data from 13 U.S. cities with high rates of crime and has added cities since. The police department data that CODE homogenizes and submits to quality assurance checks are available under a Creative Commons Attribution License. For inclusion, police departments must release incident-level crime data that meet the following requirements, as specified in CODE documentation:

1. Can be re-used under an open-data license,
2. Include the location of each crime,
3. Include the date on which each crime occurred,
4. Include a code or category that can be used to create a consistent set of crime types across cities,
5. Provide data for at least four consecutive years,
6. Pass basic checks on the quality of the data.

Notes

1. Ashby, M P J. 2019. Studying Crime and Place with the Crime Open Database. *Research Data Journal for the Humanities and Social Sciences*. Advanced online access <https://doi.org/10.1163/24523666-00401007>
2. <https://doi.org/10.17605/osf.io/zyaqn>