

National Longitudinal Study of Adolescent Health

Wave I Crime Data Codebook

Penny Gordon-Larsen



Carolina Population Center
University of North Carolina at Chapel Hill

October 2009

This research was funded by National Institutes of Health (National Institute of Child Health and Human Development grants R01 HD39183, R01 HD041375, and K01 HD044263; National Institute of Diabetes and Digestive and Kidney Diseases grant DK56350; and National Institute on Environmental Health Sciences grant P30ES10126), National Institute of Aging NIA: K07 AG001015 and P30 AG024376; and a cooperative agreement with the Centers for Disease Control and Prevention (CDC SIP No. 5-00).

Persons interested in obtaining data files from Add Health should contact Add Health, Carolina Population Center, 123 W. Franklin Street, Chapel Hill, NC 27516-2524 (addhealth@unc.edu.)

Add Health Wave I Crime Data

Filename: w1crime.xpt

Description

W1CRIME contains county-level crime data based on the location of Wave I respondents. More specifically, it includes crime rates per 100,000 people based on crimes reported (as opposed to arrests), as well as some variables evaluating the quality of the crime rates calculated. The variables comprising this file resemble the reported crime measures found in the crime component of the existing Add Health Wave I contextual database, which contains arrest and reported crime data based on 1993 statistics.

Data are based primarily on the 1995 Uniform Crime Reporting (UCR) data (part four, "DS4: Crimes Reported," study number 6850), obtained from the National Archive of Criminal Justice Data. In certain instances, data from 1990 through 1994 were used, when a county associated with an Add Health respondent lacked 1995 crime statistics. For details, regard the following data quality section.

Data Quality

Joining the raw 1995 UCR data to the Wave I respondent locations generated 461 respondents with a county population count equaling zero. This resulted in division by zero when calculating the crime rates, which produced missing values for the 461 respondents. Also, only two of the 461 respondents were located in a county that had crime data for 1995. To fill in the missing population and crime values, data were extracted from the most recent complete UCR data for the years 1990 through 1994.

For 361 of the 461 respondents, missing values were filled in using data from 1990 through 1993, which did not include a variable for the geographic coverage indicator (W1CR4), as this field did not become available in the UCR data until 1994. For these 361 respondents, the W1CR4 field was populated with the replacement code for "not available" (999 to cover a maximum value of 100).

For the remaining 100 respondents, five of them were given values from 1994, and received the geographic coverage indicator value present in the data for 1994. The other 95 had no data for 1990 through 1995, and had to be assigned crime data by averaging the crime rates for the closest adjacent counties that had UCR data for 1995. For these 95 respondents, the value of the dummy variable W1CR6 (crime rate method) equals one, and the value for W1CR4 is an average of the geographic coverage indicator values for the counties used.

To isolate respondents for which data were not available in the raw 1995 UCR dataset, one should select the following: $W1CR5 < 1995$ or $(W1CR5 = 1995 \text{ and } W1CR6 = 1)$.

For a more detailed description of the data quality indicator variables created to help one understand the crime rate calculation methods used, please see the Data Quality Indicators section.

Variable Listing

Crime Rate Data

Variable Name	Variable Label and Description
W1CR1	Total reported crimes per 100,000 <ul style="list-style-type: none">• Formula = $\text{round}((\text{MODINDEX} / \text{CPOPCRIM}) * 100000)$.• MODINDEX = Modified Index, the total number of UCR Index crimes, including arson (murder, rape, robbery, aggravated assault, burglary, larceny, motor vehicle theft and arson). The sum of reported murders through cases of arson.• CPOPCRIM = the county population of the agencies providing the crime statistics.
W1CR2	Reported violent crimes per 100,000 <ul style="list-style-type: none">• Formula = $\text{round}(((\text{murder, rape, robbery, aggravated assault}) / \text{CPOPCRIM}) * 100000)$.
W1CR3	Reported non-violent crimes per 100,000 <ul style="list-style-type: none">• Formula = $\text{round}(((\text{MODINDEX} - (\text{murder, rape, robbery, aggravated assault})) / \text{CPOPCRIM}) * 100000)$.

Data Quality Indicators

Variable Name	Variable Label and Description
W1CR4	Coverage indicator <ul style="list-style-type: none">• Available in the UCR datasets beginning in 1994 (for greater detail see the 1995 UCR codebook).• If W1CR5 is before 1994, which means no coverage indicator value was available, the W1CR4 value was set to 999. Otherwise, non-missing values range from 0 to 100.• For the three counties for which the crime rate was calculated by averaging the 1995 crime rates for the closest adjacent counties (W1CR6 = 1), the W1CR4 value is an average of the coverage indicators for the counties used.
W1CR5	Year of UCR data <ul style="list-style-type: none">• Values range from 1990 to 1995, excluding 1991, which provided no useful data.
W1CR6	Crime rate calculation method <ul style="list-style-type: none">• 0 = extracted from values for a single county in a single year as identified in W1CR5.• 1 = generated via averaging crime rates for the closest adjacent counties. There were only three counties for which this crime rate calculation method was necessary.
W1CR7	Population data source <ul style="list-style-type: none">• 0 = UCR based on year identified in W1CR5.• 1 = intercensal population estimate from the U.S. Census Bureau. There were 2 counties for which this population data source was necessary.

W1CRIME: Wave I Crime Data

Wave I Crime Data

AID		Str 8	Respondent Identifier NOTE: Smallest 5 and largest 5 values are displayed.
Frequency	Percent	Value	Label
1	0%	10316654	
1	0%	10316952	
1	0%	10506342	
1	0%	10570810	
1	0%	10606128	
20735	100%	10676355-99886994	NOTE: Range of values omitted from display
1	0%	99886995	
1	0%	99886996	
1	0%	99886997	
1	0%	99886998	
1	0%	99886999	

W1CR1		Num 2	Total crimes per 100,000 NOTE: Smallest 5 and largest 5 values are displayed.
Frequency	Percent	Value	Label
1	0%	108	
2	0%	417	
5	0%	571	
148	1%	612	
1	0%	655	
19604	94%	666-10090	NOTE: Range of values omitted from display

2	0%	10166	
5	0%	10982	
4	0%	11325	
713	3%	12342	
260	1%	13723	

W1CR2		Num 2	Violent crimes per 100,000 NOTE: Smallest 5 and largest 5 values are displayed.
Frequency	Percent	Value	Label
1	0%	7	
30	0%	16	
1	0%	18	
3	0%	19	
2	0%	34	
19632	95%	36-1887	NOTE: Range of values omitted from display
366	2%	1920	
4	0%	2128	
5	0%	2232	
441	2%	2550	
260	1%	3035	

W1CR3		Num 2	Non-violent crimes per 100,000 NOTE: Smallest 5 and largest 5 values are displayed.
Frequency	Percent	Value	Label
1	0%	72	
5	0%	239	
2	0%	397	
1	0%	448	

148	1%	480	
19222	93%	622-8750	NOTE: Range of values omitted from display
193	1%	8822	
196	1%	8965	
4	0%	9198	
713	3%	10456	
260	1%	10688	

W1CR4		Num 2	Coverage indicator
Frequency	Percent	Value	Label
5	0%	10	10
1	0%	28	28
1	0%	66	66
1	0%	75	75
119	1%	82	82
8	0%	86	86
15	0%	87	87
5	0%	88	88
209	1%	89	89
2	0%	90	90
13	0%	91	91
13	0%	92	92
315	2%	93	93
1	0%	94	94
16	0%	96	96
260	1%	97	97
753	4%	98	98

2532	12%	99	99
16115	78%	100	100
361	2%	999	Data Not Available

W1CR5		Num 2	Year of UCR data
Frequency	Percent	Value	Label
5	0%	1990	
354	2%	1992	
2	0%	1993	
5	0%	1994	
20379	98%	1995	

W1CR6		Num 1	Crime rate calculation method
Frequency	Percent	Value	Label
20650	100%	0	Single county and year
95	0%	1	Average of the closest adjacent counties

W1CR7		Num 1	Population data source
Frequency	Percent	Value	Label
20743	100%	0	Population based on year identified in W1CR5
2	0%	1	Intercensal population estimate