

National Longitudinal Study of Adolescent Health

Wave III 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 III Crime Data

Filename: w3crime.xpt

Description

W3CRIME contains county-level crime data based on the location of Wave III respondents. More specifically, it includes crime rates per 100,000 people based on crimes reported (as opposed to arrests made), as well as some variables to help understand the quality of the crime rate data calculated for an individual. The variables comprising this file differ from the crime measures found in the crime component of the existing Add Health Wave III contextual database, which measures arrests rates, not reported crime rates, using 2001 crime statistics.

Data are based primarily on the 2001 Uniform Crime Reporting (UCR) data (part four, "DS4: Crimes Reported," study number 3721) obtained from the National Archive of Criminal Justice Data. In certain instances, data from 1999 and 2000 were used when a county associated with an Add Health respondent lacked 2001 crime statistics. For details, regard the following data quality section.

Data Quality

During the initial exploratory processing of these data, it was discovered that crime rate variables could not be extracted from the raw 2001 UCR data for all respondents. As a result, 2001 data were adjusted, and additional UCR data were downloaded and adjusted.

Records were identified in the UCR 2001 table where the population of the area for which crimes were reported equaled zero, but the coverage indicator and modified index (the total number of UCR Index crimes) were both greater than zero. In these records crimes were reported by at least one agency in a given county, but the population of that county was assigned a zero value in the source UCR data. In these cases, the population values for the county in question were extracted from the US Census Bureau Intercensal estimates for 2001 and used in data processing.

A total of 14,042 respondent records successfully joined to the updated 2001 UCR file. UCR files for 1996, 1997, 1998, 1999, and 2000 were then downloaded, and missing population values were updated using the associated intercensal population estimate. These tables were then joined to the respondent table, which yielded 199 records joined to the 2000 UCR table, and 1 record joined to the 1999 table. Zero records joined to the 1996, 1997, and 1998 UCR tables.

The remaining 647 records were then imputed. Of these, 333 records were imputed by averaging the crime rate and coverage indicator of adjacent counties. Three hundred and fourteen of the imputed records were not immediately adjacent to counties with data; consequently, these records were imputed by extracting the crime rate, coverage indicator, and population source variables from the closest county for which UCR data were available. All imputation was done using 2001 UCR data.

Respondents not associated with a county due to a lack of geocodes received the Add Health reserve code for geocode missing, which equals 8 proceeded by a series of 9s; the number of which varies depending upon variables' maximum values. A total of 308 respondents were ultimately assigned missing geocode values.

To isolate respondents for which data were not available in the raw 2001 UCR dataset, select the following: $W3CR5 < 2001$ or $0 < W3CR6 < 8$. For a more detailed description of the data quality indicator variables created to help one understand the crime rate calculation methods used for this dataset, please see the Data Quality Indicators section.

Variable Listing

Crime Rate Data

Variable Name	Variable Label and Description
W3CR1	Total 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.
W3CR2	Reported violent crimes per 100,000 <ul style="list-style-type: none">• Formula = $\text{round}(((\text{murder, rape, robbery, aggravated assault}) / \text{CPOPCRIM}) * 100000)$.
W3CR3	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
W3CR4	<p>Coverage indicator</p> <ul style="list-style-type: none">• Available in the UCR datasets beginning in 1994 (for greater detail see introduction, pages i through iv, in 1995 UCR codebook).• W3CR4 non-missing values range from 0 to 100, rounded from four decimal places down to two if less than 100.• For the 333 records for which the crime rate was calculated by averaging the 2001 crime rates for the closest adjacent counties (W3CR6= 1), the W3CR4value is an average of the coverage indicators for the counties used.• For the 314 records for which crime was extracted from the closest county the W3CR4is an indicator of the coverage of the closest county, and not the county in which the respondent is located.
W3CR5	<p>Year of UCR data</p> <ul style="list-style-type: none">• Non-missing values range from 1999 to 2001.
W3CR6	<p>Crime rate calculation method</p> <ul style="list-style-type: none">• 0 = extracted from values for a single year as identified in W3CR5.• 1 = generated via averaging crime rates for the closest adjacent counties.• 2 = generated via finding the nearest county with valid crime rates. This method was used when there were no adjacent counties with valid crime rate data.
W3CR7	<p>Population data source</p> <ul style="list-style-type: none">• 0 = UCR based on year identified in W3CR5.• 1 = intercensal population estimate from the U.S. Census. There were 2 records for which this population data source was necessary.

W3CRIME: Wave III Crime Data

Wave III 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	
15187	100%	11316754-99886992	NOTE: Range of values omitted from display
1	0%	99886993	
1	0%	99886994	
1	0%	99886995	
1	0%	99886996	
1	0%	99886999	

W3CR1		Num 4	Total crimes per 100,000 NOTE: Smallest 5 and largest 5 values are displayed.
Frequency	Percent	Value	Label
2	0%	5	5
1	0%	149	149
3	0%	159	159
1	0%	202	202
2	0%	221	221
14668	97%	303-10123	NOTE: Range of values omitted from display

4	0%	10318	10318
5	0%	10623	10623
4	0%	15227	15227
199	1%	16688	16688
308	2%	99998	Geocode Missing

W3CR2		Num 2	Violent crimes per 100,000 NOTE: Smallest 5 and largest 5 values are displayed.
Frequency	Percent	Value	Label
3	0%	0	0
2	0%	2	2
3	0%	9	9
1	0%	18	18
1	0%	20	20
14540	96%	23-1647	NOTE: Range of values omitted from display
1	0%	1910	1910
4	0%	2188	2188
135	1%	2247	2247
199	1%	2745	2745
308	2%	9998	Geocode Missing

W3CR3		Num 4	Non-violent crimes per 100,000 NOTE: Smallest 5 and largest 5 values are displayed.
Frequency	Percent	Value	Label
2	0%	2	2
1	0%	32	32
1	0%	92	92
3	0%	150	150

1	0%	186	186
14670	97%	187-9258	NOTE: Range of values omitted from display
5	0%	9565	9565
3	0%	9664	9664
4	0%	13039	13039
199	1%	13943	13943
308	2%	99998	Geocode Missing

W3CR4		Num 2	Coverage indicator NOTE: Smallest 5 and largest 5 values are displayed.
Frequency	Percent	Value	Label
2	0%	31	31
1	0%	33	33
2	0%	38	38
2	0%	42	42
4	0%	43	43
1287	8%	48-96	NOTE: Range of values omitted from display
79	1%	97	97
902	6%	98	98
793	5%	99	99
11817	78%	100	100
308	2%	998	Geocode Missing

W3CR5		Num 2	Year of UCR data
Frequency	Percent	Value	Label
1	0%	1999	1999
199	1%	2000	2000
14689	97%	2001	2001

308	2%	9998	Geocode Missing
-----	----	------	-----------------

W3CR6		Num 1	Crime rate calculation method
Frequency	Percent	Value	Label
14242	94%	0	Single county and year
333	2%	1	Average of the closest adjacent counties
314	2%	2	Nearest county
308	2%	8	Geocode Missing

W3CR7		Num 1	Population data source
Frequency	Percent	Value	Label
14887	98%	0	Population based on year identified in W3CR5
2	0%	1	Intercensal population estimate
308	2%	8	Geocode Missing