# Fragile Families Uniform Crime Report Restricted Use Data Appendage

## Baseline, 1, 3, 5, 9, and 15 year Waves

May 2019

**Bendheim-Thoman
Center for Research on Child Wellbeing**
Wallace Hall
Princeton University
Princeton, NJ 08544
http://crcw.princeton.edu

**Columbia Population Research Center**
1255 Amsterdam Avenue, Room 715
Columbia University
New York, NY 10027
http://cupop.columbia.edu/

## DATA APPENDAGE OVERVIEW

The Fragile Families Uniform Crime Report Restricted Use Data Appendage (ff_UCR_pub1.dta) county-level crime rate data that correspond to the residence of respondents at Baseline and during the 1, 3, 5, 9, and 15-year follow-up waves of the Fragile Families and Child Wellbeing Study (FFCWS). Variables consist of county level crime rates, per 100,000 people. For waves 1-5, these rates correspond to the current residence of the biological mother and biological father at the wave(s) in which they were interviewed. Crime rate data for wave 6 correspond to the current residence of the focal child's primary caregiver (PCG) -- biological mother, biological father, other family, or other non-family caregivers.

## FILE LAYOUT

The file contains 4,898 observations (one per family), sorted by *idnum.*

## VARIABLE NAMING CONVENTION

Contextual variable names are 7-14 characters long. The first 3 characters contain the variable prefix. The next 4-5 characters contain the variable suffix, which includes an abbreviated description of the contextual data. The last 4 characters for the 15-year data, indicate if the variable was constructed from 2015 ACS 5-year estimates, the 2010 Census or the 2000 Census. The variable names are constructed as follows:

| Position | Character | Indicates |
|---|---|---|
| 1 | r | Restricted data |
| 2 | g | Geographic Measure |
| 3 | 1-6 | Wave of data collection |
| 4-7 | ucr_ | Data source, Uniform Crime Report (UCR) |
| 8 | [m/f/p] | Respondent: m:Mother, f:Father, p:PCG |
| 9-end | allrt | Total crime rate per 100,000 people |
| 9-end | viort | Violent crime rate per 100,000 people |
| 9-end | prprt | Property crime rate per 100,000 people |
| 9-end | covind | County coverage indicator |

For example, variable rg5ucr_fprprt: the prefix (position 1-3) rg5 refers to a geographical (county) level measure (g) at the fifth follow-up interview (9-year) (5); the suffix (ucr_fprprt) refers to the source of the data (UCR) for the father's county (f) property crime rate (prprt).

# ABOUT UNIFORM CRIME REPORT

Crime rate data in waves 1-6 included in the current data appendage was obtained from the National Archive of Criminal Justice Data (NACJD). Data was downloaded from ICPSR here: https://www.icpsr.umich.edu/icpsrweb/NACJD/series/57?q=

The following section comes from the ICPSR website here: https://www.icpsr.umich.edu/icpsrweb/content/NACJD/about.html

## About NACJD

Established in 1978, the National Archive of Criminal Justice Data (NACJD) archives and disseminates data on crime and justice for secondary analysis. The archive contains data from over 2,700 curated studies or statistical data series. NACJD is home to several large-scale and well known datasets, including the National Crime Victimization Survey (NCVS), the FBI's Uniform Crime Reports (UCR), the FBI's National Incident-Based Reporting System (NIBRS), and the Project on Human Development in Chicago Neighborhoods (PHDCN).

The following sections come from the ICPSR website here: https://www.icpsr.umich.edu/icpsrweb/content/NACJD/guides/ucr.html#desc_cl

## About the FBI's Uniform Crime Reporting Program

Uniform Crime Reports (UCR) serve as periodic nationwide assessments of reported crimes not available elsewhere in the criminal justice system. By 1985, there were approximately 17,000 law enforcement agencies contributing reports either directly or through their state reporting programs. By 1998, the number of agencies was over 18,500. More information about the development of the UCR Program can be found on the FBI website.

## County-Level UCR Data

The **county-level** Uniform Crime Reporting files contain only arrests and crimes reported data and are distributed annually.

## Study Design

County-level UCR files are *created by NACJD* based on agency records in a file obtained from the FBI that also provides aggregated county totals. NACJD imputes missing data and then aggregates the data to the county-level. Specific information about the imputation algorithm can be found in the codebooks for each year of the county-level files.

IMPUTATION PROCEDURES USED FOR 1994 UCR COUNTY-LEVEL FILES AND ONWARD: The data for any ORI reporting 12 months were used for county aggregation as submitted. Data for an ORI reporting 3 to 11 months were increased by a weight of [12/months reported]. For ORIs reporting 0 to 2 months, data for these ORIs were set to zero and then data were estimated using rates calculated from ORIs reporting 12 months of data located in the ORIs

geographic stratum based on UCR Population Groups within their state. UCR Population Groups are defined as follows:

Population Group
1. Cities 250,000 and over
2. Cities 100,000 - 249,999
3. Cities 50,000 - 99,999
4. Cities 25,000 - 49,999
5. Cities 10,000 - 24,999
6. Cities 2,500 - 9,999
7. Cities under 2,500
8. Non-MSA counties and non-MSA State Police
9. MSA counties and MSA State Police

COVERAGE INDICATOR: For releases of UCR county-level files before 1994, data from jurisdictions reporting less than 6 months of data were not included in county totals in an effort to ensure cross-sectional data comparability and quality. With the new procedures to adjust for incomplete reporting, data will be provided for each active ORI that reports less than 12 months of data, whether through weighting of partial year data or substitution of a value based on population group and state. Instead of exercising an a priori judgment that 6 months of data is the minimum threshold for acceptable data quality, a new Coverage Indicator variable has been created that will allow users to set their own threshold for acceptable data quality and to include or exclude data based on the standards they set themselves. The Coverage Indicator variable represents the proportion of county data that is not imputed for a given year. The indicator ranges from 100, indicating that all ORIs in the county reported for 12 months in the year, to 0, indicating that all data in the county are based on estimates (as described above), not reported data. The Coverage Indicator is calculated as follows:
$$CI_x = (1-(sum((ORI_i \text{ pop/county pop})((12\text{-months reported}/12)))))*100$$

where

CI = Coverage Indicator
x = county
i = ORI within county

Some ORIs do not have a population associated with their jurisdiction. These ORIs report for jurisdictions such as national parks, colleges and universities, toll bridges and tunnels, and most state police departments. As the coverage indicator is based on months of reporting and the population of each agency, this variable will not show estimation that did occur for statewide ORIs and for ORIs as listed above that do not have a population but reported 3 to 11 months of data. Conversely, the coverage indicator will indicate that estimation has occurred for ORIs with a population that reported 3 to 11 months of data even if the ORIs actually reported no crimes or arrests. Similarly, the coverage indicator will indicate that estimation had occurred for ORIs with a population that reported 0 to 2 months of data, even though no rate was calculated to estimate data because of the lack of ORIs in the agency's geographic stratum reporting 12 months of data. Finally, since data for ORIs that reported 0 to 2 months of information are set to zero, users

should be aware that no estimation of data was possible for ORIs without a population that reported 0 to 2 months of data.

It should be emphasized that, while UCR staff were consulted in developing the new adjustment procedures, these UCR county-level files are not official FBI UCR releases and are being provided for research purposes. Users with questions regarding these UCR county-level data files can contact the National Archive of Criminal Justice Data at the ICPSR.

ADDITIONAL NOTES: In the UCR county-level arrest files, the population and data for jurisdictions located in multiple counties are provided only in the county containing the largest population component of the jurisdiction. Counties containing smaller population components of multiple-county jurisdictions will contain no population or arrest data for these jurisdictions. Data in counties affected by one or more multiple-county jurisdictions are indicated by a multi-county jurisdiction flag variable. In the county-level crimes reported files, the population and crime data for jurisdictions located in multiple counties are provided by the UCR porportioned to each county (maximum of three) in which the jurisdiction is located.
Cities designated by the Census Bureau as independent cities are reported separately and have their own "county" codes. Some jurisdictions, such as state parks and some state police, provide data only on a statewide basis. In these cases, data are allocated to counties proportionate to their share of the total state population. State Police data for Vermont that were not reported within a county and the State Police data for Alaska are not allocated to the counties. These two State Police records are identified by the county code 999.

The original data from the FBI contains one record for New York City. Data from New York City are allocated into New York City's five counties on the basis of the proportion of the population in each county. For example, the population for Queens county is divided by the total population of New York City and the resulting proportion is multiplied with data from each of New York City's arrest and offense categories to apportion data to Queens county.

## DATA CLEANING

### Reshaping the UCR Data
The source data files were organized by FIPS state and county codes, containing counts for each county of the following types of crimes: murders, rapes, robberies, aggravated assaults, burglaries, larcenies, motor vehicle thefts, and arsons. The file also contained a coverage indicator for the each county. The raw data files for the years 1998-2014 and 2016 were merged together and crime rates were generated, for each year, as follows:

**Total crime rate**: the sum of the counts for each crime listed above, divided by the county population.
**Violent crime rate**: the sum of the counts of murder, rape, robbery, and aggravated assault, divided by the county population.
**Property crime rate**: the sum of the counts of burglary, larceny, motor vehicle theft, and arson, divided by the county population.

This resulted in an intermediate file with 73 variables: a county code along with the county crime rates and coverage indicators for each year listed above.

**Merging with Fragile Families**

To merge the above intermediate file with our family identifiers (idnum), a second intermediate file containing the idnum, respondents' interview year and county at each wave was created. Then, for each wave respondent combination, the crime rates file was merged to this second intermediate file, using respondent's county at that wave. Next, the crime rate variables were generated for each idnum by taking the crime rate from the year that matched the year of the respondent's interview. Due to missing UCR data for 2015 and 2017, **respondents with interviews in 2015 were assigned crime rates from 2014, and respondents with interviews in 2017 were assigned crime rates from 2016.** Missing values were assigned as described below.

## MISSING VALUES

All missing values for contextual variables are coded using similar conventions found in other FFCWS data files. More specifically, the following three missing codes are used:

| | |
|---|---|
| *-9 Not in wave* | The respondent did not participate in interview/assessment |
| *-7 N/A* | The respondent participated in interview/assessment; however, the respondent's county could not be determined. |
| *-3 Missing* | The respondent participated in interview/assessment and the residence was determined; however, crime data is unavailable. |

## DATA DICTIONARY

| Variable | Description |
|---|---|
| rg{w}ucr_mallrt | Mother's county total crime rate (per 100,000 people) in year of Wave {w} interview |
| rg{w}ucr_mviort | Mother's county violent crime rate (per 100,000 people) in year of Wave {w} interview |
| rg{w}ucr_mprprt | Mother's county property crime rate (per 100,000 people) in year of Wave {w} interview |
| rg{w}ucr_mcovind | Mother's county coverage indicator in year of Wave {w} interview |
| rg{w}ucr_fallrt | Father's county total crime rate (per 100,000 people) in year of Wave {w} interview |
| rg{w}ucr_fviort | Father's county violent crime rate (per 100,000 people) in year of Wave {w} interview |
| rg{w}ucr_fprprt | Father's county property crime rate (per 100,000 people) in year of Wave {w} interview |
| rg{w}ucr_fcovind | Father's county coverage indicator in year of Wave {w} interview |
| rg6ucr_pallrt | PCG's county total crime rate (per 100,000 people) in year of Wave 6 interview |
| rg6ucr_pviort | PCG's county violent crime rate (per 100,000 people) in year of Wave 6 interview |
| rg6ucr_pprprt | PCG's county property crime rate (per 100,000 people) in year of Wave 6 interview |
| rg6ucr_pcovind | PCG's county coverage indicator in year of Wave 6 interview |

*{w} indicates this variable was repeated for waves 1-5. At wave 6, only the PCG was interviewed