

## Procedures for Estimating in the Uniform Crime Reporting (UCR) Program

James H. Noonan, PhD

The UCR Program will begin offense estimation in the 2013 *Crime in the United States (CIUS)* publication. The following information details the intended method for achieving a valid and reliable estimate and uses an example of Police Employee (PE) estimation to demonstrate the employed statistical model.

Currently, the UCR Program publishes PE data by agency population group, shown in PE Tables 70-75, by state in PE Tables 76-77, and by individual law enforcement agency (*CIUS* Tables 78-81). PE totals are represented as a rate per 1,000 persons. Based on aggregated population group rates in PE Table 70 and individual agency data, the UCR Program is capable of reporting agency imputations and national estimations for PE counts in its publications.

Excerpt from Table 70<sup>1</sup>

### Full-time Law Enforcement Employees

by Population Group

Rate per 1,000 Inhabitants, 2012

Population group	Agency Count	Population	Rate
TOTAL CITIES	10,826	193,021,74	2.9
Group I (250,000 and over)	75	56,040,960	3.5
Group II (100,000 to 249,999)	196	29,282,253	2.3
Group III (50,000 to 99,999)	426	29,406,808	2.1
Group IV (25,000 to 49,999)	798	27,574,550	2.2
Group V (10,000 to 24,999)	1,743	27,664,124	2.3
Group VI (under 10,000)	7,588	23,053,050	4.6
Total Counties	3,180	91,998,321	4.4
Suburban areas <sup>2</sup>	7,503	124,099,81	3.7

<sup>1</sup> US Federal Bureau of Investigation (2012). This is not Table 70 in its entirety but an excerpt from Table 70, *CIUS*. Retrieved from [http://www.fbi.gov/about-us/cjis/ucr/crime-in-the-u.s/2012/crime-in-the-u.s.-2012/police\\_employee\\_data/police\\_employee\\_data](http://www.fbi.gov/about-us/cjis/ucr/crime-in-the-u.s/2012/crime-in-the-u.s.-2012/police_employee_data/police_employee_data)

<sup>2</sup> Suburban areas include law enforcement agencies in cities with less than 50,000 inhabitants and county law enforcement agencies that are within a Metropolitan Statistical Area. Suburban areas exclude all metropolitan agencies associated with a principal city. The agencies associated with suburban areas also appear in other groups within this table.

## **National Estimation**

For 2012, 17,010 agencies reported data to the FBI and of those agencies 14,006 reported a once a year police employee count which means the missing 3,004 agencies missing police employee information must be imputed. National estimates for police employee data will be computed by summing agencies reporting a complete set of data with agencies who are estimated using the imputation method based on if the agency reported for the calendar year.

## **Imputation for Non-reporting Agencies for Police Employee Counts**

For agencies not reporting a police employee count, the FBI UCR Program will impute the police employee count based on the rate of police employees per 1,000 inhabitants within the agency's population group multiplied by the agency's population.

$$\text{Population Group PE} \frac{\text{Rate} * \text{Agency Population}}{1,000} = \text{Estimated PE Count}$$

For example, Philadelphia police employee counts were not reported in 2012. The PE rate for agencies over one million in population is 3.5 per 1,000 persons and Philadelphia has a population of 1,538,957 persons. Therefore its estimated PE count for 2012 is 5,386.

$$\frac{3.5 * 1,538,957 \text{ persons}}{1,000} = 5,386.35 \approx 5,386$$

## **Combining Estimates with Reported Totals**

Once the agencies are imputed, their PE totals will be summed with the agencies that reported PE counts to derive the national estimate. For 2012, there were 194,995 police employees reported and 32,397 estimated police employees. Therefore, the estimated number of police employees in the United States in 2012 is 227,392 police employees.

$$194,995 + 32,397 = 227,392$$

## **Limitations and Future Improvements**

The imputation method demonstrated in this paper's estimation model is based on the assumption that agencies within population groups are most similar to one another. Further

study needs to be done to determine if regional characteristics of crime would help better stratify agencies into more representative groups. Until other characteristics, such as poverty rates, employment rates, or other socioeconomic variables become attributable to specific agencies and statistically tested for validity, the best stratification variable the FBI UCR Program has is agency population group. The FBI UCR Program is also assuming that the best representation of an agency's missing data is the agency's data reported in prior years, but there is no evidence which definitively shows that an agency's own data qualifies as a representative sample.

Once new variables are tested and validated, other techniques for imputation may be used, such as regression analysis. Other types of imputation models, such as bootstrapping, are not currently available with the FBI UCR Program's current technological or staffing resources; however, the Program is working toward a partnership with the Bureau of Justice Statistics (BJS) which will allow for shared resources in the future which could handle more sophisticated and statistically valid and reliable methods. Police employee data may vary from one law enforcement department to another because of administrative constraints, such as departmental budgets, that will need to be studied and identified for use in future estimations.

Agencies with zero-population are not represented in the imputation model. The FBI UCR Program will examine the value of longitudinal and ratio-based models for future estimation methods. For example, there may be a reliable ratio of offense data to PE counts which can best represent missing data, including those with zero-population. Using an agency's previous reporting patterns over the last five years may also yield valuable estimation results, but these models are not yet tested. Discussions concerning these methods with the BJS showed there is value in examining new cross-sectional, longitudinal, and ratio models, but they will take some time to develop and validate.

A final potential future improvement for nationally estimating crime data is testing UCR data for representativeness. If the data are tested and found to support the idea that UCR data is a statistically representative sample of reported crime data, then sampling estimation models can be used to estimate national reported crime totals. Of course, should the tests show the data is not representative; the other aforementioned imputation models will still be available to the FBI UCR Program to meet the need for determining national crime estimates.

Given the limitations, this estimation proposal is proposed as the next step for the FBI UCR Program to enhance its value for estimating crime data. It is not intended as a once-and-for-all solution for estimation. It is also expected that a current National Academy of Sciences and future partnering opportunities with the BJS and other UCR stakeholders will provide recommendations for estimation that are not currently considered or anticipated in this document.

This proposed method also has other uses within the UCR Program, including providing estimations for other reported data collections, including Part I crimes, arson data, arrest data,

hate crime, cargo theft, human trafficking, and the Law Enforcement Officers Killed and Assaulted program. Specific methodological considerations need to be addressed for unique data concerns within each collection, but a method similar to the PE estimation example can be used as the base estimation model for those programs.