Criminal Justice Statistical Analysis Center

January 2013

State of West Virginia
Department of Military Affairs & Public Safety
Division of Justice and Community Services
Office of Research and Strategic Planning

Improving State Capacity for Crime Reporting: An Exploratory Analysis of Data Quality and Imputation Methods Using NIBRS Data

Christina R. LaValle, M.S., ORSP Research Analyst
Stephen M. Haas, Ph.D., ORSP Director
Erica Turley, B.S., ORSP Research Analyst
James J. Nolan, Ph.D., West Virginia University

Crime reporting in the United States originates from two major sources of data, the Uniform Crime Reports (UCR) and the National Incident-Based Reporting System (NIBRS). The incident-based reporting (IBR) structure of NIBRS is an enhancement to the traditional summary reporting of UCR used to track crime in the U.S. While the law enforcement community initiated IBR to address the expanding complexity of crime, reporting crime using NIBRS, like UCR, is voluntary and susceptible to issues of data quality, missing data, and noncompliance.

Data collected using UCR and NIBRS are used for research and to document the status of crime at the national, state, and county levels. Data quality regarding accuracy and completeness are critical to reliable results and information. Further, reporting data "as is" without considering data quality and estimating for missing values may not be the most accurate depiction of the process and can result in criticism. As funding and resources lessen, coupled with the multitude of data fields involved with IBR, issues of data quality and missingness are areas of concern for analysts and researchers. In assessing data quality and handling missing data, appropriate and effective methods for resolving issues are necessary.

Currently, elaborate methods established by the Federal Bureau of Investigation (FBI) are used to evaluate UCR data quality in terms of outlier detection (see Akiyama & Propheter, 2005). The FBI also imputes, or estimates, missing UCR data using methods that were developed in 1958 (Maltz, 1999). These methods are not timely, accurate, or easy for state programs to administer since they often use data from regions involving multiple states rather than the individual state. Moreover, methods have not yet been applied to aggregate crime count totals using NIBRS data and often reports using NIBRS data are criticized for being incomplete or non-representative.

With the intended replacement of UCR with NIBRS, the need for data quality and imputation methods in the context of NIBRS aggregate totals grows. By inspecting data quality and applying imputation methods to IBR data, this research provides examples, recommendations, and easy-to-use tools that can be employed by state repository personnel, researchers, and data analysts who analyze state-level IBR data. This study explored

Office of
RESEARCH &
STRATEGIC PLANNING

data quality and imputation methods with the purpose of providing guidance to states reporting IBR data and improving the utility of NIBRS data.

Background

The UCR and NIBRS data are both considered primary sources of crime statistics (National Institute of Justice [NIJ], 2009). However, while the UCR tracks crime offense totals for eight crime categories (referred to as Index crimes), the NIBRS collects specific offense details of 22 crime categories (referred to as Group A crimes) (See Appendix A for list of Summary UCR and NIBRS offense and arrest crime categories). Further, crime rates produced using UCR data consist of incident tallies where only the most serious crime is reported (known as the hierarchy rule), while crime rates using NIBRS data do not follow the hierarchy rule[1]. Thus, the UCR is referred to as a summary reporting system where the NIBRS system involves incident-based reporting since it collects several details for each incident.

Data collected using an IBR system are particularly useful as they provide a wealth of information by gathering multiple attributes of each particular criminal incident in addition to aggregating the data to provide summary crime counts. Thus, IBR data are used to investigate complex criminological issues as well as provide crime statistics and trends[2]. However, while the use of NIBRS data for research has gained in reception, many studies are limited to specified populations since the data lack representative national coverage (Addington, 2008).

The conversion from UCR to NIBRS reporting has been slow (James, 2008). In 2010, the UCR covered 97.4 % of the total U.S. population and of that, NIBRS covered about 28% (Federal Bureau of Investigation [FBI], 2011a, 2011b). NIBRS data received by the FBI are converted to UCR data before any data processing occurs for the annual Crime in the U.S. report (L. Simmons, email, March 8, 2012).

As of June 2012, 43% of law enforcement agencies report NIBRS data to the FBI which covers 27% of crime reported nationally and 29% of the population (Justice Research and Statistics Association [JRSA], 2012). Currently, 32 states are certified to report crime data to the FBI using NIBRS; 13 of those states, including West Virginia[3] report 100% of their crime using NIBRS (see Table 1) (JRSA, 2012).

The NIBRS data, like the UCR, are not exempt from problems of data quality and missing values. The FBI checks UCR data monthly for anomalous reporting using several layers of control[4]. Data identified as outliers are flagged as missing and imputed.

Table 1: Status of NIBRS data collection in the United States

Percent Coverage	State
Less than 20%	Arizona, Louisiana, Missouri, Oklahoma, Texas
20% - 50%	Maine, Nebraska, Oregon, Washington, Wisconsin
50% - 90%	Colorado, Connecticut, Kansas, Kentucky, Massachusetts, Ohio, Utah
100%	Arkansas, Delaware, Idaho, Michigan, Montana, New Hampshire, Rhode Island, South Carolina, South Dakota, Tennessee, Vermont, Virginia, West Virginia
Testing/ Developing	Alabama*, California, Illinois*, Indiana, Maryland, Minnesota, Nevada, New Jersey, New Mexico, New York, North Carolina, Pennsylvania, District of Columbia*
No formal plans	Alaska, Florida, Georgia, Hawaii, Nevada, Wyoming

*NIBRS data from an individual agency is accepted by FBI

Source: *Status of NIBRS in States, 2012, JRSA*

Report Highlights...

Reporting crime data "as is" without considering data quality and estimating for missing values can lead to biased estimates of crime.

This research hopes to improve state methods for estimating crime statistics through the use of techniques that are accessible to state repository personnel, data analysts, and others who analyze state-level IBR data.

Applying data quality and imputation methods are pivotal to producing more accurate, stable, and reliable crime estimates.

This research provides evidence that alternative imputation methods improve the accuracy of crime reporting for WV and may hold promise for other states.

The FBI's method for imputing missing data depends on the amount of data missing. For agencies missing one to nine months of data, the annual crime total is estimated by multiplying the total number of crimes for the year by twelve then dividing by the number of months crime were reported. For agencies missing ten to twelve months of data, the annual crime total is estimated by multiplying the agencies' population by the crime rate for the agencies' population group divided by 100,000 (Maltz, 1999). For agencies that do not have an accompanying population "zero-population" agencies (e.g. State Police (SP), Park Police, college campus security, etc.), data are not imputed if not reported.

Though it may seem reasonable to apply FBI data quality and imputation methods to NIBRS data, it is worthwhile to explore other options for several reasons. First, the FBI imputation methods have been used since 1958 and are seemingly outdated given modern technology (Maltz, 1999). Second, the current imputation methods used on the UCR data were developed to produce annual national estimates of crime and are not appropriate for smaller units of analysis (Lynch & Jarvis, 2008).

The task of modernizing imputation methods has been investigated by Michael Maltz and colleagues to not only better national estimates of crime, but also provide estimates at smaller units of analysis.

In 2006, Maltz, Roberts, and Stasny compiled a report for the American Statistical Association Committee on Law and Justice Statistics proposing improved imputation methods for crime data. The report recommended different statistical distribution-based imputation methods for different situations[5]. The report also suggested possible improvements for future work, including using the average of months surrounding the missing month and Bayesian methods.

In 2011, Targonski suggested using longitudinal methods to impute for missing values in a final report to the Department of Justice. The proposed method suggested that for agencies missing ten to twelve months of data, a weight, calculated by dividing the current year's group crime rate by the previous year's group crime rate, was multiplied by the agency's data reported in the prior year; for an agency missing one to nine months of data, a weight, calculated using the agency's year-to-year increase based on matched months reported, was multiplied by the agency's previous year's monthly data (see Targonski, 2011).

Using suitable values to estimate missing data are vital to filling in known gaps and bettering analyses and reporting. However, establishing such methods is a task of balance. Methods for handling missing data have to be advanced enough to accurately estimate values yet accessible to be employed with reasonable guidance.

This research intended to develop and investigate methods that would improve the accuracy of state NIBRS data used for crime trend analyses. The WV IBRS data were used to develop techniques for assessing data quality and investigating the impact of imputation methods on state crime trends with agencies that report partial data or no data in a given year. Methods developed were required to be simple and accessible to state repository personnel, researchers, and data analysts who analyze or report NIBRS data.

Methodology

Assessing data quality and establishing a missing value pattern were critical initial analyses for studying imputation methods. There were two main issues of data quality that lurk in the WV IBRS data and had an effect on analysis at the local or state level: missing data and outliers. Neither NIBRS nor WV IBRS data have a designated variable or value to indicate a monthly crime report as missing. Rather, a monthly report of zero is the default value for missing data, but is also used to specify zero crimes reported for the month. Second, outlier detection, identifying a monthly crime count markedly different from others, is the responsibility of the analyst. Given that the WV IBRS data covers 100% of the population and crime, guidelines for classifying zeros as a true zero or missing and techniques for outlier detection were developed using the 2009 data.

Imputation methods were tested using simulation based on the missing value pattern identified by classifying zeros and detecting outliers. A second simulation study was performed to investigate the impact of estimating crime totals with complete agency data deleted at multiple missing data scenarios. The WV IBRS data was ideal to use for simulation because WV is a full reporting NIBRS state with 100% population and crime covered.

The 2009 WV IBRS data included aggregated monthly property and violent crime counts from 200 agencies and were used for analysis. All analyses were conducted using Microsoft Excel 2010, Visual Basic for Applications (VBA), Statistical Package for Social Science (SPSS v20), or R (v2.14.2).

Data Source

The WV Statistical Analysis Center (SAC) receives statewide IBR data compiled by UCR staff directly from the state repository. The SAC receives a data dump of all incidents reported for the requested calendar year typically in April to ensure the same amount of lag time in reporting for greater consistency and comparability over time.

The WV IBRS data used for this research was from Jan.1 - Dec. 31, 2009 and obtained in April of 2010. For validation and secondary analysis, data from 2007 (received in April 2008) and 2008 (received in April 2009) were also used. The WV IBRS offenses-known crime data were aggregated by crime type, property and violent, and were the focus of this research. Violent crimes consisted of murder, forcible rape, robbery, and aggravated assault; property crimes consisted of burglary/ breaking and entering, motor vehicle theft, all larceny, and arson.

Data Quality: Classifying Zeros

Developing guidelines for classifying zeros as a 'true zero' or missing data were based on a few

Report Highlights...

This study used West Virginia incident-base reporting data to assess data quality and investigate imputation methods for missing aggregate crime count data.

Two main issues impact data quality for WV and other states: missing data and outliers.

Guidelines were developed to classify reports of zero as 'true zero' or 'missing' based on agency type, crime total, and number of consecutive zeros reported across crime types.

Imputation methods were tested using simulation based on the missing value pattern identified by classifying zeros and detecting outliers.

Simulations were also conducted to determine the impact of estimating crime totals with complete agency data deleted at multiple missing data scenarios.

assumptions and observations in the 2007-2009 data.

1. If a nonzero crime count was reported in one crime type and a zero in the other crime type, the zero was considered a 'true zero'. For example, in the month of January, if 4 property crimes were reported and 0 violent crimes were reported, is was assumed that the report for violent crimes was a 'true zero'.

2. Missing data may occur when an agency reports zeros in both crime types in the same month when most other months are consistently nonzero.

3. Missing data may occur if there are consecutive months with observed zeros in both crime types.

4. Agencies covering smaller populations with sparse reporting may record zeros in both crime types for the same month or consecutive months because no crimes were reported (i.e., 'true zero').

Based on the aforementioned assumptions, two variables were created to assist in developing diagnostic rules for classifying zeros: a variable to indicate the total number of crime reported (Total) and a variable for counting the number of consecutive months where all crime counts were zero (NCZ).

Diagnostic cut points were determined using the R statistical package for k-means cluster analysis. Cluster analysis is often used to group data based on similarities among variables, understand data structure, and determine group characteristics; favorable for developing zero classification guidelines (Tan, Steinbach, & Kumar, 2006).

Data Quality: Outliers

Detecting outliers is often an initial step to any analysis as anomalous data have an effect on accuracy. The FBI applies a host of data quality measures to the UCR data which includes extensive outlier detection (see Akiyama & Propheter, 2005). Because NIBRS data are converted to UCR data, methods for detecting irregular reporting in IBR data are not as established.

There is no one method established for detecting outliers; rather, it is up to the analyst to select methods appropriate for the data. To the experienced analyst familiar with the data, spotting data inconsistencies may simply be done by inspection. Using automated outlier detection

Table 2: Description of automated outlier detection methods tested

Name	Algorithm	Details
Standard Deviation Method	$x_i > \bar{x} + 3*SD$ or $x_i < \bar{x} - 3*SD$	Data outside the calculated thresholds are potential outliers. This test assumes data come from a normal distribution.
Box Plot	$x_i > Q_3 + 3*IQR$ or $x_i < Q_1 - 3*IQR$	Data outside the calculated thresholds are potential outliers. This test does not assume data come from a normal distribution.
Dixon's Q test	$Q_{Upper} = \frac{x_n - x_{n-2}}{x_n - x_2}$ $\quad Q_{Lower} = \frac{x_3 - x_1}{x_{n-1} - x_1}$	Q is the ratio of 'gap' to 'range'. The Q test statistic is compared to a critical value at a specified level of significance. This test was developed for small sample sizes. The Dixon test assumes normality and known location of outlier.
Ratio to Median	$Yi = \frac{x_i}{\tilde{x}}$	Yi measures the number of times the monthly crime count is compared to the agency's median. The Yi test statistics is compared to a user defined critical value. No assumptions are made about the data distribution and the development of the test borrows concepts from FBI (see Akiyama & Propheter, 2005).
Ratio of Ranges	$Rr_{Top} = \frac{\lvert x_n - \tilde{x}\rvert}{\left(\frac{Total}{Range}\right)}$ $\quad Rr_{Bottom} = \frac{\lvert x_1 - \tilde{x}\rvert}{\left(\frac{Total}{Range}\right)}$	Rr measures the ratio of 'gap' to 'range' and was developed by the WV SAC. The Rr test statistic is compared to a user defined critical value and makes no assumption about the data distribution.

x_i = crime count at month i, $\bar{x}$ = mean, SD = standard deviation, Q_1 = first quartile, Q_3 = third quartile, IQR = inter quartile range, x_n = *ordered* monthly crime count where $x_1 < x_2 < ... < x_{12}$, $\tilde{x}$ = median, Total = annual crime total, Range = $x_n - x_1$

Table 3: Description of graphical techniques used to visualize data and detect outliers

Plot Name	Purpose	Method	Parameters	Outlier Diagnostics
Histogram	Assess the data distribution.	Frequency plot of data grouped by distinct intervals or 'bins'.	Number of bins = 5 according to k = 1 + $\log_2 12$, (Gentle, 2002). Bin width = data range / k, where k is the number of bins (Sturges, 1926).	Histograms that are skewed left may indicate potential outliers; skewed right or symmetric/bell shaped are supportive of the distributional characteristics we expect from count data (Poisson or Normal distributions).
Dot Plot	Assess the data spread and/or realize data clusters.	One-dimension plot of monthly crime counts on the horizontal axis.	Plot range set to 0 and 20 plus the maximum value rounded to the nearest 10 (allows for comparisons between plots).	Dot plots with a large spread and/or large gaps depicting data clusters may indicate outliers.
Line Chart	Assess the data reporting pattern and/or seasonality.	Bivariate plot of monthly crime count data and time.	Plot range set to 0 and 20 plus the maximum value rounded to the nearest 10 (allows for comparisons between plots).	Line charts with sharp peaks or valleys may indicate outliers.

methods are often preferred as they are regarded as being more objective. However, data detected by automated methods are merely "potential" outliers which require manual examination and rationale to classify the data as acceptable or unacceptable (Akiyama & Propheter, 2005).

Several well-known and novel methods for outlier detection were applied to the WV IBRS property and violent crime count data for each agency and are described in Table 2 (further explanation of Yi method in Appendix B and Rr method in Appendix C). All outlier detection methods were performed using Microsoft Excel 2010 and VBA.

Data visualization is often recommended for understanding and visualizing relationships among data (Gentle, 2002). Graphical analysis was used to supplement outlier detection as well as visualize data patterns and distributional characteristics. Three plots used for analysis were the histogram, dot plot, and line chart, described in Table 3.

To test and determine the efficacy of outlier detection methods described, Huntington PD was used as an indicator because it was known to have three months of outlying data. Methods that identified Huntington PD as having irregular data were deemed promising. All data flagged by outlier detection methods were manually investigated by the researchers.

Imputation Methods

Imputation methods systematically estimate missing data values. There were two types of missing data scenarios that required imputation methods in the NIBRS data: estimating for agencies that reported some but not all months of data and estimating for agencies that did not report any data.

While imputation methods applied to NIBRS data have been used to study crime clearance rates (see Roberts, 2007, 2008), little is known about imputation of NIBRS aggregate crime counts used for crime trend analyses. Though applying the FBI methods directly to NIBRS data seems reasonable, alternative imputation methods were sought to improve estimation accuracy.

Agencies missing one to nine months of data are referred to as partial reporting agencies. Four different imputation methods for estimating missing monthly crime counts in property and violent crime for partial reporting agencies were investigated and compared to the FBI method. The alternative methods used averages calculated over a subdivision of months nearby the missing month to estimate

Table 4: Description of imputation methods for partial reporting agencies (missing one to nine months of data)

FBI Method **CT = PCT * 12 / number of reported months**
CT = Crime Total, PCT = Partial Crime Total
Quarter Method **CT = PCT + Q1*(N1) + Q2*(N2) + Q3*(N3) + Q4*(N4)**
N = number of missing values per period. Quarter 1: Q1 = average of Jan., Feb., Mar. crime counts; Q2 = average of Apr., May, Jun. crime counts; Q3 = average of Jul., Aug., Sept. crime counts; Q4 = average of Oct., Nov., Dec. crime counts. If N1or N4 = 3, then Q1⇔ Q4. If N2 or N3 = 3, then Q2⇔ Q3. If N1 and N4 = 3, then Q1 = Q2 and Q4 = Q3. If N2 and N3 = 3, then Q2 = Q1 and Q3 = Q4. If data for three entire quarters were missing, the average of the remaining values was used for the respective Qs. Quarter2: Q1 = average of Dec., Jan., Feb. crime counts; Q2 = average of Mar., Apr., May crime counts; Q3 = average of Jun., Jul., Aug. crime counts; Q4 = average of Sept., Oct., Nov. crime counts. If N1 = 3, then Q1 = min[Q2, Q3, Q4]. If N2 or N4 = 3, then Q2⇔Q4. If N3 = 3, then Q3 = max[Q1, Q2, Q4]. If N2 and N4 = 3, then Q2 = Q4 = average[Q1, Q3]. If data for three entire quarters were missing, the average of the remaining values was used for the respective Qs.
Tri-Annual Method **CT = PCT + T1*(N1) + T2*(N2) + T3*(N3)**
T1 = average of Jan., Feb., Mar., Apr. crime counts; T2 = average of May, Jun., Jul., Aug. crime counts; T3 = average of Sept., Oct., Nov., Dec. crime counts. If N1 = 4, then T1 = min[T2, T3]. If N2 = 4, then T2 = max[T1, T3]. If N3 = 4, then T3 = min[T1, T2]. If data for two entire tri-periods were missing, the average of the remaining values was used for the respective Ts.
Half Method **CT = PCT + B1*(N1) + B2*(N2)**
B1 = ave. of Jan., Feb., Mar., Apr., Nov., Dec. crime counts; B2 = ave. of May, Jun., Jul., Aug., Sept., Oct. crime counts. If data for B1 or B2 were missing, the average of the remaining values was used for the respective B.

Table 5: Description of FBI and three alternative imputation methods for non-reporting agencies (missing ten to twelve months of data)

$\textit{Crime Total} = \frac{\text{Population Group Crime Rate*Agency's Population}}{100{,}000}$
FBI Population Groups
Total of eight groups, developed by the FBI, are used to categorize jurisdictions. Six groups are based on population and are *roughly* half the size of the preceding group; colleges and universities are included in the smallest population group. Two groups use Metropolitan Statistical Area (MSA) designations for county and state detachments.
WV1 Population Groups
Total of eight groups. Six groups based on population are the FBI population groups divided by 10. The groups are scaled by a factor so that all population groups are occupied with at least five or more agencies and adequate data to compute crime rates. Colleges and universities are included in second smallest group. Two groups based on county and state MSA designations are the same as the FBI.
WV2 Population Groups
Total of seven groups. Five groups are established using k-means cluster analysis of all WV city population estimates for 2009 provided by the U.S. Census; cluster analysis was used to emulate the natural structure of the population data. The 'bin numeric variable' feature ("natural" method, k=5) of the 'Rcmdr' package in R was used to determine group membership. Colleges and universities are included in the smallest population group. Two groups based on county and state MSA designations are the same as the FBI.
WV3 Population Groups
Total of eight groups. Six groups are determined by splitting the data equally into 6 population groups using the 'bin numeric variable' feature ("proportions" method) of the 'Rcmdr' package in R. Since group membership is data specific, the population ranges for property and violent crimes will always be different. Two groups based on county and state MSA designations are the same as the FBI.

annual totals; this approach aligns with suggestions made by Maltz, Roberts, and Stasny regarding future work to improve imputation methods (2006) (see Table 4 for description of methods).

Imputation methods for agencies missing ten to twelve months of data (non-reporting agencies) were based on the assumption that similar agencies have similar crime volume and relate to population. Five alternative imputation methods were investigated for non-reporting agencies.

The FBI estimates non-reporting agency data by classifying the agency based on its population and metropolitan status and using the crime rate (total crime count divided by total population) for the population group (Lynch & Jarvis, 2008). Three of the alternative imputation methods used methods similar to the FBI but with modified population groups. The motive for adjusting the population group intervals stems from the fact that the FBI groups appear to be too broad for the WV population data (see Appendix D). The FBI and alternative imputation methods population groups are described in Table 5 and listed in Appendix D, Table d.

Two additional alternative imputation methods used the linear relationship between crime volume and population to estimate non-reporting agency data and follow the regression model in Table 6 where the regression coefficient, β, is calculated conventionally (labeled as SIZ) and according to a no intercept model (labeled as SNI).

Simulation

The 2009 WV IBRS property and violent crime count data were used to conduct a simulation study to investigate the imputation methods for partial reporting and non-reporting agencies using Microsoft Excel 2010 and Visual Basic for Applications (VBA). Two hundred agencies were used to create a full reporting simulation data set (reporting twelve months of data with no anomalies) and establish a missing value pattern to model. True and imputed values can be compared and accuracy assessed when using observed data. Simulation studies using observed data and associated missing value pattern have been applied to a variety of contexts such as surveys, UCR crime counts, and health studies (see Tremblay, 1994; Targonski, 2011; and Engels & Diehr, 2003).

To simulate partial reporting agencies, monthly data from selected agencies were selected at random to be removed. The simulation was repeated 1000 times to balance the chance of "good" or "bad" draws. Random seed[6] 96739 was used to generate 1000 random seeds used for selecting agencies to have data removed. Random seed 5540 was used to generate 1000 random seeds used for selecting the start month for removing data. The number of agencies and months of data to delete were determined by the missing value pattern.

A second simulation study investigated the impact of imputation methods for non-reporting agencies when entire agency data was deleted[7] at 10%, 20%, 30%, 40%, 50%, 60%, 70%, and 80%. The simulation was repeated 1000 times at each proportion and used random seed 63435 to generate 1000 random seeds for selecting agencies to remove. If data deletion were to result in no agencies available to compute a group crime rate, all estimates for the agencies belonging to the group were set to zero.

Assessing Estimation Accuracy

The simulation study estimated annual crime totals at the agency and state levels for each imputation method. The actual and imputed data were compared to test accuracy of the imputation

Table 6: Regression model and regression coefficient formulas for two alternative imputation methods for non-reporting agencies

Crime Total = β * Agency's Population where β is the regression coefficient.	
β, Traditional	β, No-intercept
$\beta_{SIZ} = \dfrac{\text{cov}(x,y)}{\text{var}(x)}$	$\beta_{SNI} = \dfrac{\sum(x*y)}{\sum(x^2)}$
where x is the independent variable (population) and y is the dependent variable (crime total).	

Table 7: Formulas and descriptions of accuracy measures for imputation methods

Mean Absolute Error $= \dfrac{\lvert y_1 - \hat{y}_1\rvert + ... + \lvert y_m - \hat{y}_m\rvert}{m}$	The MSE is the average of the absolute distance between imputed and original values. It describes how much, on average, imputed values differ from original values. Smaller values are better.
Root Mean Squared Error $= \sqrt{\dfrac{(y_1 - \hat{y}_1)^2 + ... + (y_m - \hat{y}_m)^2}{m}}$	The RMSE is the standard deviation of the prediction error. It is a measure of consistency and variation and sensitive to large over or under estimates. Smaller values are better.
Bias $= \dfrac{y_1 - \hat{y}_1 + ... + y_m - \hat{y}_m}{m}$	The Bias is the average distance between imputed and original values used to indicate the tendency for a method to over- or underestimate values. Bias of zero indicates no bias, negative bias indicates underestimation, and positive bias indicates overestimation.
where y is the original value, $\hat{y}$ is the imputed value, and m is the number of missing values.	

methods by calculating two statistics: the mean absolute error (MAE) and the root mean square error (RMSE) (Table 7). Bias was also measured to gauge whether estimates over- or underestimated crime totals (Table 7). Calculating MSE, RMSE, and Bias allows for the direct comparison of each imputation method's performance.

The MAE, RMSE, and Bias for each imputation methods were calculated for estimating totals at the agency level, (notated MAE_{ave}, $RMSE_{ave}$, and $Bias_{ave}$) and state totals (notated MAE_{tot}, $RMSE_{tot}$, and $Bias_{tot}$). When comparing MAE and RMSE, a smaller value indicates the more accurate method; Bias closest to zero indicates better performance.

Resulting MAE_{ave} and $RMSE_{ave}$ statistics were compared using a one-way analysis of variance (ANOVA) to determine whether there was a significant difference between methods. To identify specific differences, post hoc comparisons using Dunnett's test (with the FBI method as the control) were performed. To compare methods regarding state totals, ANOVA was performed on the absolute differences and labeled MAE_{tot}.

Two-way ANOVA was used to compare methods and different scenarios of missing data (e.g., 10% missing compared to 20% missing, 20% missing compared to 30% missing, etc.). Post hoc tests included sequential pairwise comparisons of missing data scenarios using Bonferroni correction. Due to the large number of simulations, statistical significance was set at 0.001.

The results of the best performing data quality and imputation methods were applied to 2007, 2008, and 2009 data to compute the state's crime trend for the three-year period and compared to the trend complied without data quality assessment or imputation methods.

Results

Data Quality: Zeros

Of the 200 agencies available for analysis, 63% (126 agencies) reported a nonzero crime count for either property or violent for all reporting months; the remaining 74 agencies required further investigation for classifying zeros. Although the third crime type, non-index crimes, was not the focus of this research, it was used to assist in classifying zeros. By including non-index crime counts, the percent of agencies with nonzero reporting in at least one of the three crime types increased to 76% (152 agencies) which left 48 agencies requiring further inspection for classifying zeros. For the 152 agencies that had nonzero crime counts in either violent, property, or non-index crimes, any zero observed was classified as a true zero.

The 48 agencies that had zeros reported in all three crime types in one or more months required further investigation. The 2008 and 2007 WV IBRS data was used to assist with classification. In comparing the 2009 data to the 2008 and 2007 WV IBRS data, if zeros were observed in all three crime types in the 2009 data, but no zeros reported in the 2008 and 2007 data, the zeros were flagged as missing. To generalize this pattern, an agency with consistent nonzero reporting with zeros simultaneously reported in all three crime types for one or two months was characteristic of classifying zeros as missing values (see Appendix E, Table a for an example).

Another irregular zero reporting pattern observed were agencies that had nonzero crime count months with several consecutive months of zeros in all crime categories. If zeros were observed in the 2009 data, but not in the 2008 and 2007 data, the zeros were flagged as missing (see Appendix E, Table b).

Within the reported data, there were instances where patterns of consecutive crime counts of zero were observed in the 2009, 2008, and/or 2007 data.

Report Highlights...

Crime trends resulting from imputed and non-imputed data showed explicit differences; thereby, showing that the accuracy of crime reporting is improved through imputation techniques.

The ratio of monthly to median crime count was robust to the presence of multiple outliers and identified the *month* with irregular reporting.

The two ratio-based outlier detection methods are complementary and were most effective at flagging anomalous data given 12 months of agency data.

Well-known outlier detection methods (i.e. Standard Deviation and Box Plot thresholds) were tested but failed to identify known irregularities in agency data.

For partial reporting agencies (missing 1 - 9 months of data), an imputation method using seasonal quarterly averages was more accurate than the FBI method.

For non-reporting agencies (missing 10 - 12 months of data), an imputation method using alternative population groups was more accurate than the FBI method.

The ratio of ranges (Rr) outlier detection method was developed during this research to identify an *agency* with irregular reporting.

Table 8: Summary of decision variables to assist with classifying zeros (total of 48 agencies)

# of Agencies	Agency Type	Zero Status	Total Range (Property)	Total Range (Violent)	NCZ Range
10	Population	True zero	[0,16]	[0,8]	[1,3]
14	Population	Missing	[1,226]	[0,22]	[1,9]
15	Population	Questionable	[1,30]	[0,5]	[1,9]
8	Zero-population	True zero	[0,36]	[0,2]	[1,6]
1	Zero-population	Missing	[5]	[0]	[7]

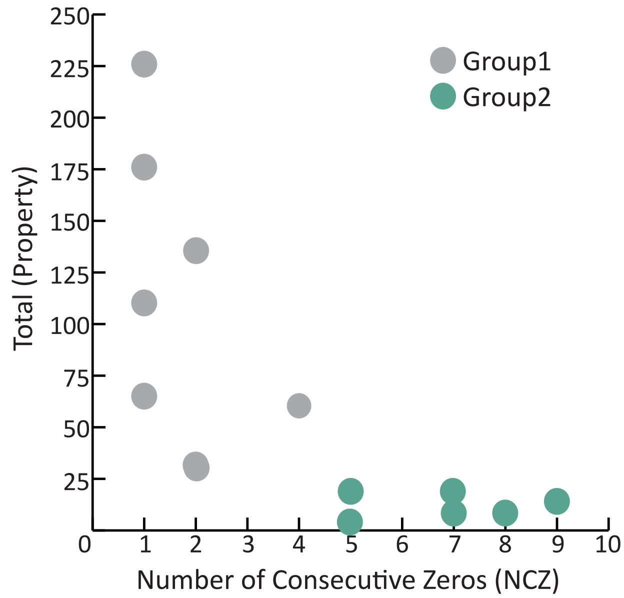
Since the pattern appeared similar from year to year, the zeros for these agencies were classified as true zeros (see Appendix E, Table c).

Researchers familiar with WV IBRS data were able to decisively classify 33 of the 48 agencies in the 2009 data as having true zeros or missing data using the 2008 and 2007. Of the 33 agencies, 15 agencies had data classified as missing and 18 agencies had data classified as true zeros. Fifteen agencies were not able to be classified with certainty due to lack of or unreliable historical data and were not used to establish classification guidelines.

After agencies were classified as having true zeros or missing data (zero status), quantitative variables for the annual crime total (Total) and number of consecutive zeros in all crime types (NCZ) were created. A qualitative variable for the agency's type (i.e. zero-population or population) was also created.

Commonalities and patterns among the Total and NCZ variables were summarized for zero-population and population agencies (Table 8). The reporting patterns for zero-population agencies (agencies that do not have an associated population such as colleges, universities, Task Forces, etc.) differed from population agencies (i.e., city and county) based on the created variables. Therefore, the classification guidelines for zero-population and agencies with associated populations were considered separately. Given the variety of functions zero-population agencies serve, the zeros observed in these types of agencies (colleges, universities, Department of Natural Resources (DNR), Task Forces, Turnpike, and State Police) needed to be

Figure 9: Plot of clusters using k-means and corresponding table of variables and data



AgencyID	Total(property)	NCZ	Cluster
2	30	2	1
9	32	2	1
3	60	4	1
10	65	1	1
11	110	1	1
12	136	2	1
13	176	1	1
14	226	1	1
1	1	5	2
5	8	8	2
4	8	7	2
8	14	9	2
7	19	7	2
6	19	5	2

classified on a case-by-case basis[8].

For population agencies (law enforcement agencies in cities and counties) that were classified as having true zeros, the Total was typically lower when compared to population agencies that had zeros classified as missing. Also, agencies that were classified as having true zeros had a smaller range of NCZ than agencies with zeros classified as missing. While the ranges for Total and NCZ overlap among the zero status categories, agencies with missing data seem to have larger Total crime counts with several consecutive months of missing data. However, applicable cut points for classifying the data were not obvious. To assist with guidelines and cut points for zero classification, additional analysis of Total and NCZ was needed.

While it seems acceptable that agencies reporting large crime counts were unlikely to report zeros, it

Table 10: Guidelines for classifying crime counts of zero as true zeros or missing data

Guideline1:	For any zero reported in a given month, if the violent, property, or non-index crime counts in the same month are non-zero, the reported zero is a true zero. If there are months when all crime counts simultaneously contain zeros, go to Guideline2.
Guideline2:	If zeros are observed in all crime types in the same month(s) AND the Total (property) is greater than 25, the zeros are flagged as missing (i.e., not reported) after checking Guideline 4. If Total (property) is less than or equal to 25, go to Guideline3.
Guideline3:	If zeros are observed in all crime types for more than 4 consecutive months, the zeros are considered missing after checking Guideline4. Extra consideration should be given to agencies where the number of consecutive months with zero reported is equal to 4; in this scenario, it is suggested to look at the number of crimes reported in other months to assist with classification.
Guideline4:	If the agency is a zero-population agency (colleges, universities, DNR, Task Force, Turnpike, or SP), separate examination is needed. For colleges/universities, it is not uncommon to observe crime counts of true zero for summer months (June-August) while having a total property crime count greater than 25. For DNR, Task Forces, Turnpike, and SP, it is not uncommon for NCZ to be greater than 4 and zeros classified as true zeros due to the nature of crime reporting for these agencies.

was useful to assign a value or cut point for decision making. For agencies classified as having missing values, there was a strong negative correlation (r = -0.858, p = 0.000) between Total (property) and NCZ, which suggested that lower Total (property) was associated with higher NCZ.

To assist with developing classification rules, k-means clustering (with k=2 clusters) was applied to the standardized variables Total (property) and NCZ for data classified as missing. The Total (property) variable was selected over Total (violent) because there was more variation to yield meaningful results.

The two resulting clusters seemed to have distinct features regarding Total (property) and NCZ values. The agencies in the Group 1 (G1) cluster had larger total property crime counts, Total (property) greater than 30 (based on the range: [30, 226]), and NCZ less than 4 (based on the range: [1,4]). The Group 2 (G2) cluster had smaller total property crime counts, Total (property) less than 19 (based on the range: [1,19] and NCZ grater than 5 (based on the range: [5,9]) (Figure 9). Explicit direction for classification guidelines were established using the clusters' characteristics. The cut point used for Total (property) was chosen to be 25 since it is roughly halfway between 16 and 30, the upper and lower limits for the two clusters. There seemed to be a straightforward value of 4 for NCZ to used for classification.

In summary, zero-population agencies must be examined individually to classify zeros. Further, attention should be given to historical reporting (if available) and seasonality to determine classification. Agencies with larger total property crime counts are unlikely to have zero reports. Agencies with small total property crime counts that display several months of consecutive zero reports in all crime types are suspected as missing values. Agencies with small total property crime counts and a few months and/or consecutive months of zero reports are likely to be classified as having true zeros (i.e., agencies with sparse reporting). The characteristics of the decision variables for the two missing data clusters and true zero categories used for establishing guidelines for classifying zeros are summarized in Table 10.

Using the developed classification guidelines for zero reports, the 15 agencies with questionable data were classified. In conclusion, of the 48 agencies that required reports of zero to be classified as true zeros or missing values, 23 agencies had reports of zeros classified missing values and 25 agencies were classified as true zeros.

Data Quality: Outliers

All agency data were checked for outliers using five automated methods and graphical analysis. Outlier methods that identified Huntington PD were considered successful since the agency was known to have anomalous data for February, March, and April. Data, descriptive statistics, and outlier statistics for Huntington PD are located in Appendix F.

Data (or agencies) identified by the outlier detection methods were manually inspected. The identified data were classified as outliers or valid data based on historical data, known data collection

Figure 11: Outlier plots for Huntington PD property crime counts

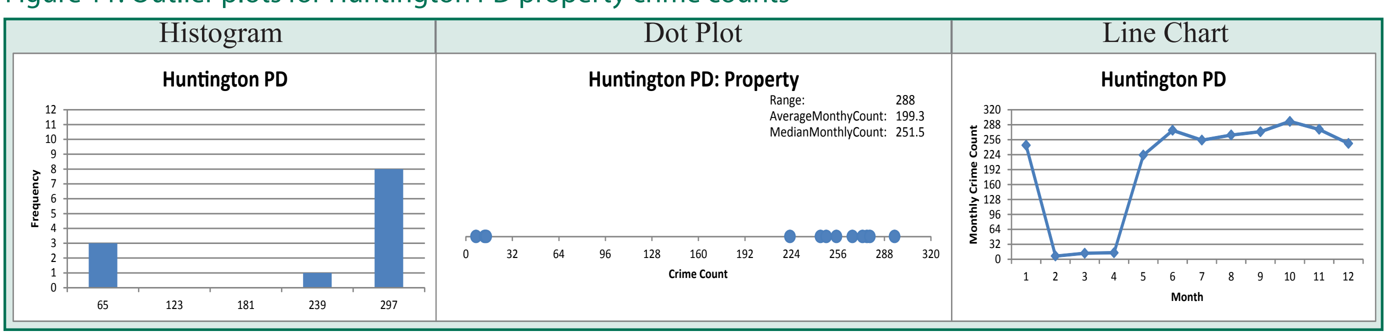


Table 12: Data for agencies identified as having irregular data (irregular data are circled)

Violent												
Agency	Jan.	Feb.	Mar.	Apr.	May	Jun.	Jul.	Aug.	Sept.	Oct.	Nov.	Dec.
HuntingtonPD	24	3	5	3	21	29	28	24	21	31	27	31
Property												
Agency	Jan.	Feb.	Mar.	Apr.	May	Jun.	Jul.	Aug.	Sept.	Oct.	Nov.	Dec.
HuntingtonPD	244	7	13	14	223	276	255	266	273	295	278	248
MarionCo	1	4	19	9	34	missing	1	missing	4	19	18	1
ParkersburgPD	11	67	102	107	138	114	101	119	121	139	105	115
ViennaPD	21	18	21	29	23	22	17	20	3	7	8	32

issues, and graphical methods.

The first method investigated was the Standard Deviation Method which flags potential outliers that are more or less than three standard deviations above or below the mean. While this method is fairly well-known, it did not identify Huntington PD data as having outliers in violent or property crimes. The Standard Deviation method identified 11 agencies in violent crime and 7 agencies in property crime as having potential outliers. After manual inspection, none of the data were judged as irregular.

The Box Plot method identifies potential outliers that fall outside a threshold range determined using the median and quartiles. Despite being based on robust statistics, the Box Plot method failed to identify Huntington PD data in property and violent crimes. Forty-seven agencies were identified in violent crimes and 23 agencies were identified in property crimes as having potential outliers. After manual inspection, one agency (Parkersburg PD) was flagged as having one outlier in property crimes.

The Dixon's Q test was developed to handle data with small sample size; however, it did not identify Huntington PD data in violent or property crimes. This method identified 19 agencies with potential outliers in violent crimes and 27 agencies with potential outliers in property crimes. After manual inspection, one agency (Parkersburg PD) was flagged as having one outlier in property crimes.

The ratio of monthly count to median (Yi) measured how large or small a monthly crime count was compared to its median. This outlier detection method seemed promising since it correctly identified Huntington PD in property and violent crimes when critical values Yi > 4 and Yi < 0.25 were used. However, there were a few computational issues with this method. Agencies with a median equal to zero had an undefined Yi, these agencies were inspected case-by-case. Agencies with a monthly report of zero would result in Yi equal to zero, which would be identified as an outlier (since zero is always less than any threshold). However, any Yi equal to zero was accepted as valid data since these reports were already screened while classifying zeros. The Yi method identified 15 agencies in violent crimes and 30 agencies in property crimes as having potential outliers. After manual inspection, Huntington PD was flagged as having irregular reporting in violent crimes and 4 agencies were flagged as having outliers in property crimes (Huntington PD, Marion Co, Parkersburg PD, and Vienna PD).

The Ratio of Ranges (Rr) method was developed by the WV SAC to identify agencies with irregular reporting using 12 months of data. The method was simple and specific to 12 months of aggregated count data. Since critical values were not established, the Rr values were ranked to determine a comparison value (Rr values for all agencies are located in Appendix G). As a result, agencies with Rr greater than 2 were suspected as having potential outliers. This method seemed promising and identified Huntington PD in property and violent crimes as having outliers. Four agencies were identified in violent and 27 agencies were identified in property crimes as having potential outliers. After manual inspection, Huntington PD was flagged as having irregular reporting in violent crimes and 4 agencies were flagged as having outliers in property crimes (Huntington PD, Marion Co, Parkersburg PD, and Vienna PD).

Graphical analysis supplemented the automated outlier detection methods. Three different plots were used for visualizing the data for each agency: a histogram, dot plot, and line chart (see Figure 11 for an agency example with outliers and Appendix H for agency examples with and without outliers). In general, viewing plots for all agencies simultaneously was helpful for differentiating agency plots that looked 'different'.

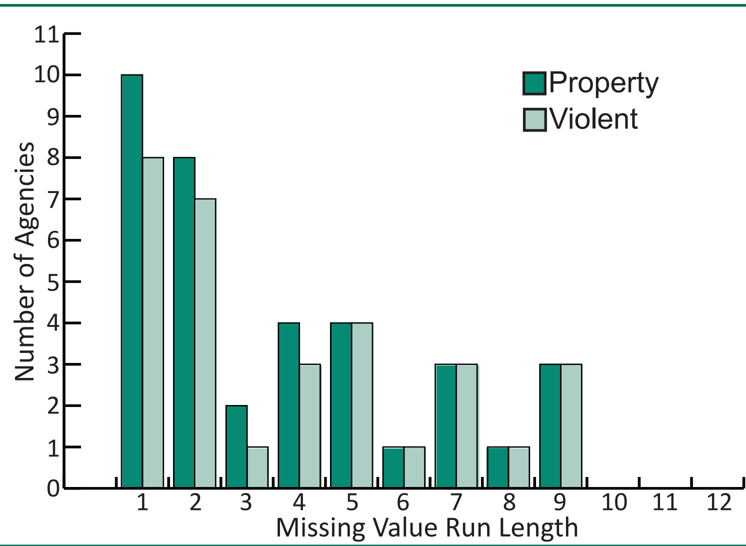
The histogram was a powerful graphic used to show the data distribution. When the histogram appeared skewed to the left, it was suspected to have anomalies. This is based on the expectation that the histogram will be either skewed to the right (i.e. follow a Poisson distribution) or symmetric (i.e. bell shaped curve or follow a Gaussian distribution). Huntington PD (violent and property data) and Parkersburg PD (property data) showed skewed left histograms (see Figure 11 and Appendix I).

The dot plot illustrated the spread and clustering pattern of data. Outliers were generally spotted as an isolated point or small cluster separated from the main cluster of data; agencies with irregular data had dot plots with these characteristics (see Figure 11 and Appendix I). This plot was also useful for classifying potential outliers as valid data for agencies with highly variable data (i.e., data with a wide spread).

The line chart displayed an agency's data trend over time; this plot allowed visualization of seasonality as well as a sharp increase (peak) or decrease (valley) in reporting. The line chart was useful in providing justification for higher or lower crime counts attributed to seasonal factors and a relatively flat pattern was suggestive of consistent reporting. All agencies with anomalous data showed a noticeable valley indicating decreased reporting (see Figure 11 and Appendix I). One agency (Marion Co) showed inconsistent reporting (peaks and valleys).

Overall, one agency was identified as having irregular reporting in violent crimes (Huntington PD) and four agencies were identified as having irregular reporting in property crime (Huntington PD, Marion Co, Parkersburg PD, and Vienna PD) (Table 12). The Yi and Rr automated outlier detection methods combined with graphical analysis

Figure 13: Missing value pattern of WV IBRS data (2009) and corresponding proportion of data by run length used in simulation



Missing Value Run Length	Proportion Missing(Property)	Proportion Missing(Violent)
1	0.05	0.04
2	0.04	0.035
3	0.01	0.005
4	0.02	0.015
5	0.02	0.02
6	0.005	0.005
7	0.015	0.015
8	0.005	0.005
9	0.015	0.015
10	0	0
11	0	0
12	0	0

Table 14: Accuracy and Bias results for estimating agency crime totals

Simulation 1: Agency						
Property				Violent		
Method	MAE_{ave}	$RMSE_{ave}$	$Bias_{ave}$	MAE_{ave}	$RMSE_{ave}$	$Bias_{ave}$
FBI	12.81	25.10	0.41	2.72	4.28	0.02
Q1	13.31	25.28	0.43	3.09*	4.92*	0.03
Q2	**11.90***	**22.23***	0.50	2.97*	4.73	0.12
Tri	**12.67**	**23.84**	-1.17*	2.87*	4.59	-0.23*
Bi	**12.51**	**24.32**	0.69	2.88*	4.50	0.09
Deleted	64.14*	135.87*	-64.14*	7.19*	16.36*	-7.19*

Note: Results in bold indicate better performance than FBI methods. Results with an asterisk show a significant difference from FBI methods

*Significant at 0.001

were the most capable at flagging irregular data (see Appendix I for outlier statistics). Three of the four agencies containing anomalous data had more than one data value identified as an outlier; the presences of multiple irregularities seemed to have an impact on effectiveness of outlier detection methods.

Imputation & Simulation

Two simulation studies were conducted. The first simulation investigated imputation methods for partial reporting agencies and the second looked at non-reporting agencies. Both simulation studies used data from full reporting agencies with no anomalies (174 agencies for property crimes and 177 agencies for violent crimes). The MAE, RMSE, and Bias were calculated for estimating agency totals and the state total to determine the most accurate method.

The missing value pattern was established so it could be modeled in the simulation (see Figure 13). Among the 200 agencies used for initial analysis, 26 agencies had missing data identified by classifying zeros and/or outlier detection. The most common missing value pattern was one missing month followed by two consecutive missing months of data. Proportions of missing run lengths were used to model the number of agencies and months deleted in the simulation study. Missing more than nine months of data was not observed in the 2009 data.

Three out of the four alternative imputation methods used to estimate agency and state property crimes for partial reporting agencies were more accurate than the FBI method according to MAE_{ave},

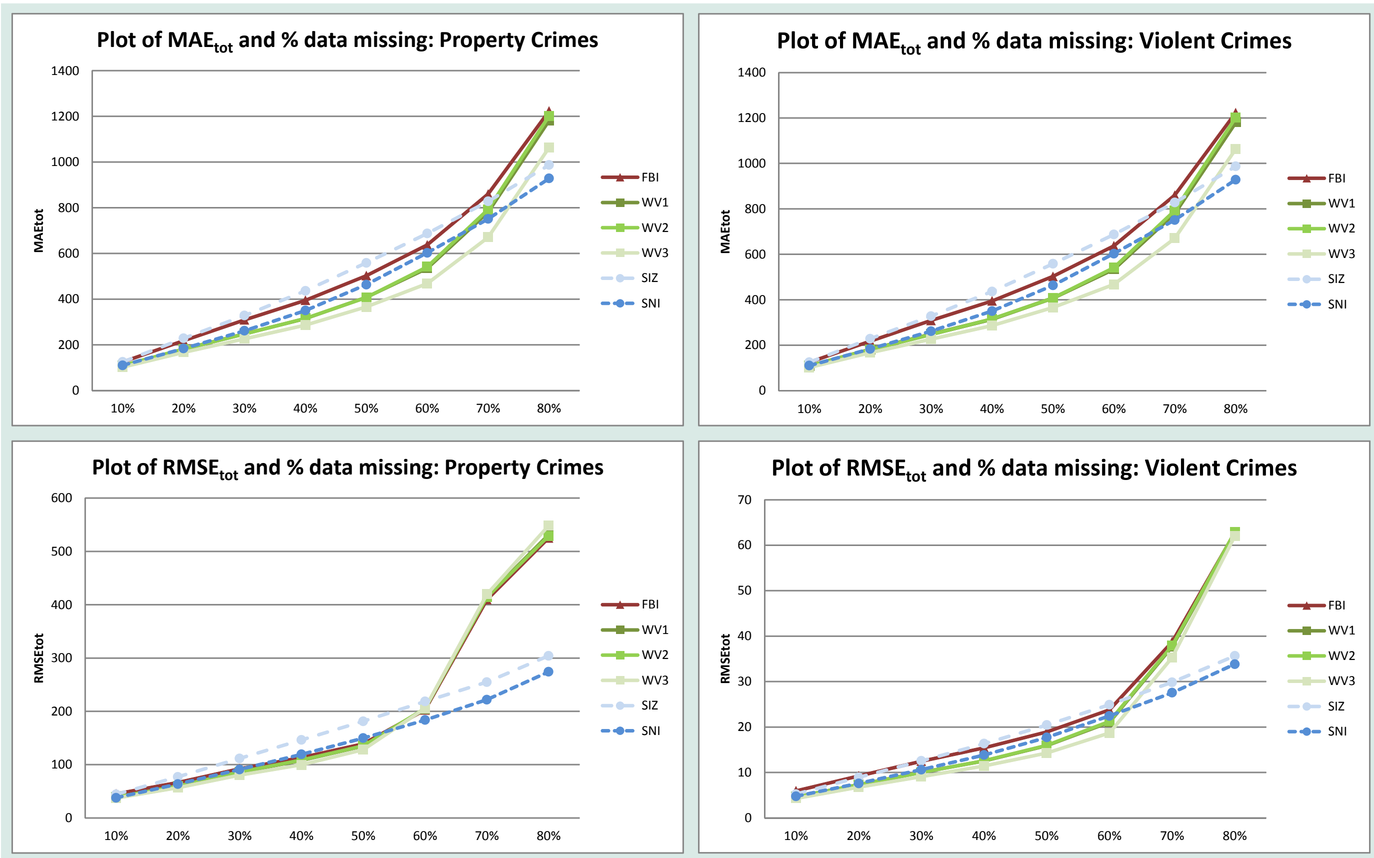
Table 15: Accuracy and Bias results for estimating state crime totals

Simulation 1: State						
Property				Violent		
Method	MAE_{tot}	$RMSE_{tot}$	$Bias_{tot}$	MAE_{tot}	$RMSE_{tot}$	$Bias_{tot}$
FBI	122.54	160.05	13.91	24.21	0.32	0.76
Q1	123.12	160.59	14.72	27.86*	0.68	0.97
Q2	**109.90***	**143.05**	17.01	26.74	**0.19**	3.85
Tri	**119.36**	**153.37**	-39.92*	25.96	0.39	-7.40*
Bi	**118.83**	**154.38**	23.51	25.47	**0.14**	2.90
Deleted	2180.72*	2288.78	-2180.73*	230.14*	7.21	-230.14*

Note: Results in bold indicate better performance than FBI methods. Results with an asterisk show a significant difference from FBI methods

*Significant at 0.001

Figure 16: Plots of property and violent crime MAE_{tot} and $RMSE_{tot}$ over all missing data scenarios for each non-reporting agency imputation method



$RMSE_{ave}$, MAE_{tot}, and $RMSE_{tot}$ (see Tables 14 and 15). The best performing imputation method for property crimes was Quarter 2 (Q2) which used seasonal quarterly averages. The Q2 MAE_{ave}, $RMSE_{ave}$, and MAE_{tot} were significantly less than the FBI method. The low $RMSE_{ave}$ and $RMSE_{tot}$ suggest that the Q2 method was more stable when estimating property crime totals compared to the FBI.

For estimating agency violent crime, the MAE_{ave} and $RMSE_{ave}$ for the FBI method were less than the alternatives; however, the $RMSE_{ave}$ for Q2, Tri, Bi methods did not significantly differ from the FBI (Table 14). When estimating the state total violent crime, the FBI method did not significantly differ from the MAE_{tot} for Q2, Tri, or Bi methods (Table 15). The $RMSE_{tot}$ for Q2 and Bi methods were less than the FBI method for violent state totals suggesting that the estimates resulting from Q2 and Bi methods were more stable than the FBI method.

The Bias at the agency and state levels suggested that all methods except Tri had a tendency to overestimate counts. However, in most cases, the Bias was relatively close to zero (Table 14). Further, the Bias for estimating the state total in property and violent crimes were not significantly different from the FBI method with the exception of the Tri method (Table 15).

The second simulation study focused on the impact of imputing missing data for non-reporting agencies. The WV IBRS data was ideal for studying imputation methods for non-reporting agencies because of its 100% population and crime coverage. The optimal imputation method, alternative methods compared to the FBI, depended on how much data were missing and what type of

Table 17: Summary of significance for comparisons between adjacent non-reporting missing data scenarios (pairwise comparisons) accuracy measures by method

Simulation 2								
	Statistic	10% - 20%	20% - 30%	30% - 40%	40% - 50%	50% - 60%	60% - 70%	70% - 80%
Property	MAE_{ave}	$NS_{1,2,3,4,5}$	$NS_{1,2,3,4,5}$	$NS_{1,2,3,4,5}$	$NS_{1,2,3,4,5}$	$NS_{1,2,3,4,5}$	$NS_{4,5}$	NS_{4}
	$RMSE_{ave}$	$NS_{1,2,3,4,5}$	$NS_{1,2,3,4,5}$	$NS_{1,2,3,4,5}$	$NS_{1,2,3,4,5}$	$NS_{1,2,3,4,5}$		
	MAE_{tot}	$NS_{1,2,3}$	$NS_{1,2,3}$	$NS_{1,2,3}$	$NS_{1,2,3}$			
Violent	MAE_{ave}	$NS_{1,2,3,4,5}$	$NS_{1,2,3,4,5}$	$NS_{1,2,3,4,5}$	$NS_{1,2,3,4,5}$	$NS_{1,2,3,4,5}$	$NS_{3,4,5}$	$NS_{4,5}$
	$RMSE_{ave}$	$NS_{1,2,3,4,5}$	$NS_{1,2,3}$	$NS_{1,2,3,4,5}$	$NS_{1,2,3,4,5}$	$NS_{1,2,3,4,5}$	$NS_{4,5}$	$NS_{4,5}$
	MAE_{tot}	$NS_{1,2,3}$	$NS_{1,2,3}$	$NS_{1,2,3}$	$NS_{1,2,3}$	NS_{3}		
	1=WV1, 2=WV2, 3=WV3, 4=SNI, 5=SIZ							

Note: NS = nonsignificant

crime total was desired (i.e., agency or state).

The MAE_{ave} and $RMSE_{ave}$ for all proposed alternative methods were smaller than those of the FBI method when 10% - 70% of agencies were simulated missing or non-reporting for estimating agency property crime totals. For violent crimes, all but SNI had smaller MAE_{ave} compared to the FBI method when 10% - 70% agency data were missing. The $RMSE_{ave}$ for all methods at all test scenarios (10% - 80%) were smaller than the FBI method. For all test scenarios (10% - 80%) the SIZ and WV3 had smaller MAE_{ave} and $RMSE_{ave}$ than the FBI (Appendix J).

When estimating the state total, the MAE_{tot} for WV1, WV2, and WV3 were less than the FBI method when 10% - 70% of the data were missing for property crime (Figure 16 and Appendix K). At 70% - 80% missing, the SNI and WV3 MAE_{tot} were less than the FBI. The WV1, WV2, and WV3 $RMSE_{tot}$ were smaller than the FBI method when 10% - 50% of the data were simulated missing. At 50% - 80% missing, the SNI had a lesser $RMSE_{tot}$ than the FBI method. For violent crimes, WV1, WV2, WV3, and SNI had smaller MAE_{tot} and $RMSE_{tot}$ than the FBI at 10% - 70% missing. At 80% missing, all methods had lower MAE_{tot} than the FBI. All methods except WV2 had smaller $RMSE_{tot}$ than the FBI at 80% missing.

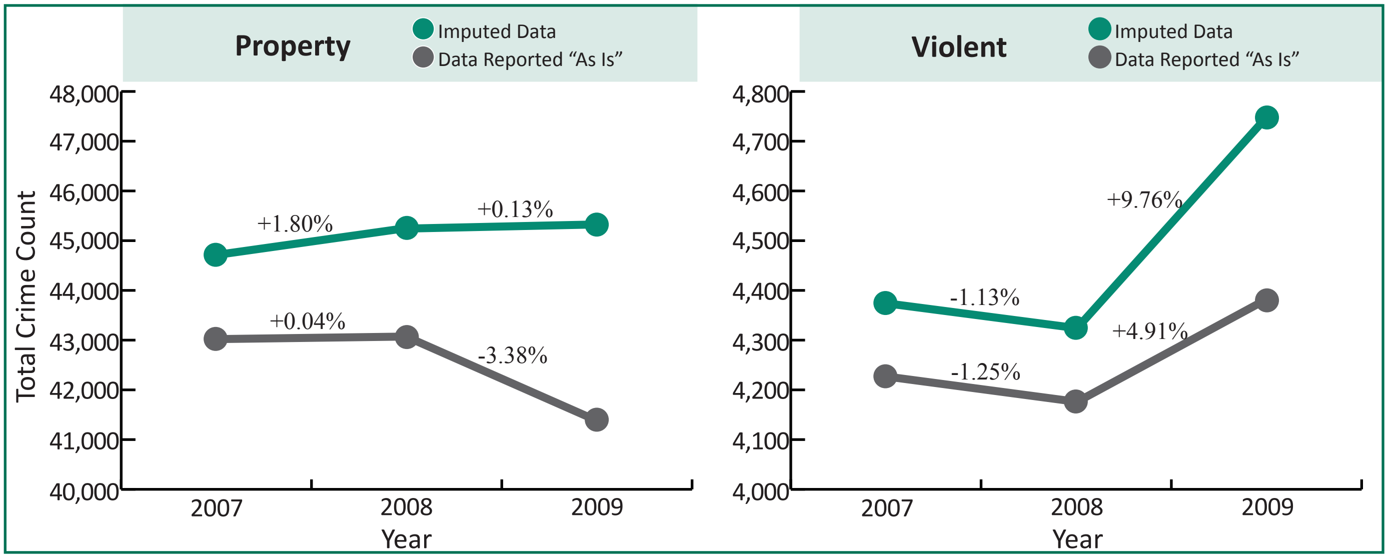
The investigation of non-reporting imputation methods' performance at different levels of missing agency data was conducted to give insight for determining how much reported data were needed to provide accurate estimates. When estimating the state total, at 80% of the data missing, the mean absolute difference was roughly 33% and 27% of the total for property and violent crimes, respectively (MAE_{tot}/Total Crime Count*100%). At 10% missing, the estimates are about 4% and 3% of the total for property and violent crime, respectively. Though some error is expected when estimating data, too much error yields unreliable results.

To determine how much data was needed to provide reliable crime total estimates, the MAE_{ave}, $RMSE_{ave}$, and MAE_{tot} were compared at consecutive missing data scenarios (e.g. comparing measures at 10% missing to 20%, 20% missing to 30%, etc.).

Agency total estimates using all methods seemed to be reliable with up to 60% of data missing for most methods based on MAE_{ave} and $RMSE_{ave}$ pairwise comparisons of property and violent data (see Table 17 and Appendix L). For the two regression based imputation methods, there were no significant differences between the MAE_{ave} at nearly all pairwise comparisions[9] for property and violent crimes. This suggested relatively consistent estimation for agency totals using regression even when up to 70% of the data are missing.

When estimating the state property and violent crime totals, the accuracy measures seemed stable

Figure 18: Three-year crime trend for state property and violent imputed and unimputed data between 2007 and 2009



using WV1, WV2, WV3, or FBI methods as long as 50% was available for imputation based on the MAE_{tot} (see Table 17 and Appendix L). The regression based methods were significantly different at all pairwise comparisons suggesting unreliable state total estimation.

The negative Bias observed for estimating agency and state violent crime totals suggest that on average, imputation methods, including the FBI, underestimated crime counts. For property crime, the Bias for WV1 and WV2 tended to fluctuate between over- and underestimating across missing data scenarios. On average, the WV3 method consistently overestimated and the two regression methods (SIZ and SNI) consistently underestimated property crime totals (Appendix M).

To illustrate the impact of imputation methods, the Q2 and WV1 imputation methods were applied to 2007, 2008, and 2009 data to compute the state's property and violent crime trends. The alternative imputation methods were selected based on MAE, RMSE, and Bias performance. As expected, the number of crimes for the imputed data were greater than the unimputed data (Figure 18).

The three-year trend for property crime showed similar trajectory from 2007 to 2008 in the imputed and "As Is" (or unimputed) trend lines (rates of change was 1.18% and 0.04%, respectively). However, the trend from 2008 to 2009 showed a slight increase in the imputed data (0.13%) and relatively steep decrease in the unimputed data (-3.38%) (Figure 18).

The three-year violent crime trend showed a similar decrease then increase pattern among the imputed and "As Is" data (Figure 18). While the rates of change between 2007 and 2008 were comparable for the imputed and unimputed trend (-1.13% and -1.25%, respectively), the rate of change between 2008 and 2009 was greater in the imputed data compared to the unimputed data (9.76% and 4.91%, respectively).

In both the property and violent three-year crime trend, there seemed to be a marked difference between the trends from 2008 to 2009. It is possible that the results could be connected to the number of agencies reporting missing data. In 2007, there were 67 agencies with missing property data and 66 with missing violent data. In 2008, there were 68 and 65 agencies with missing property and violent data, respectively. However, in 2009, there were 81 and 79 agencies that had missing data in property and violent crimes. The increase in the number of agencies with missing data in 2009 may account for the discrepancies in trend between 2008 and 2009.

In effect, by reporting data "as is", trends may show underreported crime counts or skewed rates of change when compared to trends using imputed data.

Conclusion

Data quality is a concern for researchers, analysts, and stakeholders; especially in terms of accuracy and completeness. Although several states have laws that mandate agencies report crime, it still remains voluntary and vulnerable to data quality issues. Assessing data quality should always be an initial step to any analysis; however, challenges arise when data attributes are unique, have ambivalent values, and conventional methods fail.

Zeros reported in UCR and NIBRS data have one of two meanings, no crime to report (true zero) or no report was submitted (data missing). The duel meaning of zero reports requires inspection for the sake of data quality. While it seems reasonable to assume that agencies with consistent reporting are not likely to report zeros, determining 'consistent reporting' is challenging. By inspecting and analyzing reporting patterns in the WV IBRS data, guidelines and decision variables were created to assist with classifying zero reports[10]. Although specific cut-points were listed in the guidelines, it should be noted that agencies with values near or on the cut-point values may require additional consideration or the use of historical data (if available) to assist with classification.

Despite the acceptability of well-known Standard Deviation, Box Plot, and Dixon's Q test outlier detection methods, they were not suited for detecting outliers in the WV IBRS data. The Standard Deviation and Box Plot methods use measures of central tendency and dispersion; in the presence of outliers, these statistics become inflated and affect the performance of the methods. In the WV IBRS data, it was common for an agency to have more than one outlier; an agency with three outliers equates to 25% of the data being irregular. Since outliers cause a shift in the location of the mean and an inflated variance, the conventional thresholds become so large that they failed to detect outliers, especially in the case of multiple outliers. Even with the more robust measures used for the Box Plot and Dixon's Q test, the presence of multiple outliers seemed to have an effect on their reliability.

Both the Standard Deviation and Dixon's Q test assume data come from normal distribution. Crime count data, which can be considered a process of observing a discrete number of events, is more characteristic of the Poisson distribution (Osgood, 2000). The distribution of the data was illustrated by the histogram plot. Many of the histograms of agencies with regular data appeared skewed right (although it should be acknowledged that some agencies did display symmetric plots).

As a result, the ratio of monthly crime counts to the median (Yi), ratio of ranges (Rr), and graphical analysis were found to be most effective methods for detecting outliers in 12 months of crime count data[10]. The two ratio methods were complementary; while the Rr identified an agency with irregular reporting, the Yi identified the month of irregular reporting. These tests also seemed to be successful in flagging data when multiple outliers were present.

Crime trend statistics are dependent on completeness. Since classifying zeros and identifying outliers result in missing values, the need for imputing, or estimating for missing data, is essential. Although the FBI has been using the same imputation methods for decades, its appropriateness is questioned in this time of computational capacity and interest in smaller units of analysis. Despite efforts to improve the FBI methods, alternatives are often complex or involve years of reported data. This research looked to improve methods while using techniques that were accessible, particularly for states wanting to increase the utility of IBR data.

Overall, when comparing the true reported crime totals to imputed crime totals, the results reinforced that imputing for missing data was statistically better than doing nothing.

Given the characteristics of the WV IBRS data (i.e. 100% crime and population covered), two analyses investigated imputation methods for partial and non-reporting agencies. For partial reporting agencies, the imputation method using seasonal quarterly averages (Q2) gave more accurate crime count estimates than the FBI method for property crime. The development of imputation methods based on observed data nearby missing data supported suggestions made in previous research investigating improvements to imputation methods (see Maltz et al., 2006). While the accuracy measures for estimating violent crime counts were less for the FBI method when compared to the alternatives, they were not statistically different from the Q2 method. While statistical significance between the Q2 and FBI methods was observed in property crime, the lack of significance in violent crime may be attributed to the sparseness of the violent crime data and/or lack of reporting pattern fitting the seasonal quarterly model. Additional research using data with higher levels of reporting, particularly in violent crimes, may assist in determining whether imputation methods using smaller ranged averages (e.g., seasonal) give more accurate results.

The second simulation focused on estimating data for non-reporting agencies at several missing data scenarios (10% - 80% data missing). For each missing data scenario, many of the alternative imputation methods were more accurate than the FBI's method. One plausible reason for improvement was that alternative methods were specific to the state's population. Selecting the best performing imputation method depended on the missing data scenario, crime type, and whether the estimate was at the agency or state level.

The imputation methods using regression were more accurate when estimating agency totals in all missing data scenarios when compared to the FBI method. Further, the accuracy did not significantly change as the amount of data deleted increased, which suggests that imputation using regression is relatively stable when estimating agency totals and moderate amounts of data are missing. The performance of the regression-based imputation method used for estimating agency totals supports the concept that crime volume and crime rates are related to population (Nolan, 2004). Using regression to impute data for agency totals seems promising; however, accuracy hinges on the availability of an associated population.

Regarding the parameter used for regression-based imputation, it appeared that the regression model using the slope from the slope-intercept model and setting the intercept to zero (SIZ) performed better than the slope with no intercept model (SNI) in the majority of the missing data scenarios for property and violent crime estimates. When calculating the different slopes in the simulation, the slope for the SIZ method tended to be larger than in the SNI method.

Conversely, imputation using regression to estimate the state total were not as accurate as the FBI in a majority of scenarios. The weak performance was likely due to the fact that regression methods do not have a mechanism for incorporating zero population crime counts. Imputation methods using population groups, on the other hand, are able to capture zero population crime counts by including counts when calculating crime rates, which inflates estimates, but residually account for non-reported zero population crimes when imputing (Barnett-Ryan, 2007). Further work detailing the characteristics of non-reporting agencies would assist in developing more precise imputation methods for non-reporting agencies, specifically, methods for estimating missing data for zero-population agencies separately from agencies associated with a population.

Alternative imputation methods that used different population groups were more accurate than the FBI method for the majority of the missing data scenarios when estimating state crime totals.

While the imputation method using equal

proportion population groups (WV3) gave more accurate estimates when compared to the FBI method, the method consistently overestimated crime totals. In addition, the improved performance of the method may be attributed to mathematical properties rather than criminological theory[11].

The imputation methods using scaled population groups (WV1) and k-means clustering (WV2) were also more accurate than the FBI method. Throughout the missing data scenarios, WV1 and WV2 performed similarly. With the objective of using simple techniques, the imputation method that used scaled population groups was favored. For the WV IBRS data, the WV1 population groups were determined by scaling the national FBI population groups by ten. The specific scale used to establish population groups for other states should be state specific. The primary factor for determining the scale for the WV IBRS data was the number of agencies available in group 1; a sufficient number of agencies were needed to calculate a reliable crime rate.

Highlights...

The results of this study illustrate that the use of imputation methods are statistically better than not adjusting for missing crime data.

Alternative imputation methods explored in this study performed better than the FBI's method for a the majority of the missing data scenarios when estimating state crime totals for WV.

This research resulted in the development of guidelines for classifying zero reporting as missing data or zero counts and techniques for detecting outliers given one year of crime count data.

The results of this study provide the groundwork for refining methods using different state/jurisdictional or additional longitudinal data in an effort to improve the accuracy of state crime reporting and encourage greater use of incident-based reporting data.

The performance of imputation methods for non-reporting agencies decreased as the amount of missing data increased. Deterioration in accuracy was expected as more data were missing. The observed statistical differences in accuracy between missing data scenarios assisted in determining a cut point for how much data are needed to give reliable results. When estimating state crime totals, this study showed that there was a significant increase in error (MAE_{tot}) when 50% and 60% missing data were compared for property and violent crimes (see Appendix L and Table 12). This research suggests that reliable state crime totals can be estimated when no more than 50% of the data are missing.

Applying data quality and imputation methods to three years of property and violent crime count data, there were observed differences in crime trends from imputed and unimputed data. First, as a result of applying imputation methods, the overall volume of property and violent crime was greater than crime counts reported without imputation. This was expected as the imputation methods accounted for missing data. Second, there were distinct differences in rates between the imputed and unimputed trends between 2008 and 2009 in both property and violent crime. The 2008 to 2009 change in property crime for the imputed data showed a slight increase while the unimputed data showed a marked decrease. During the same time period, both the imputed and unimputed trends displayed an increase in violent crime; however the imputed trend had a sharper rate. It is plausible that the difference in rates between the imputed and unimputed trends may be associated with the number of full reporting agencies. During the 2007 and 2008 period, the number of full reporting agencies changed by one and the trends between imputed and unimputed property and violent data were comparable. However, during the 2008 and 2009 period, the number of full reporting agencies lessened by 13 for agencies reporting property crimes and 14 for agencies reporting violent crimes potentially connecting the decrease in property crime and damped increase in violent crime to the decrease in reporting. Therefore, this research suggests that reporting data without assessing data quality and estimating missing values, or leaving data "as is", may not be sufficiently accurate or reliable when depicting crime trends or calculating rates of change.

The findings of this research are not without limitations. Methods were explored and developed using one year of data; therefore, the investigation of partial and non-reporting imputation methods were done separately. The primary advantage for isolating the two contexts was that exploring imputation methods for non-reporting agencies allowed the close examination of estimating crime counts at various levels of non-reporting. The information could assist states that are interested in estimating crime totals but are not yet reporting NIBRS at 100%. Future work would include longitudinal data for validation and modeling purposes. In addition, although WV is a 100% reporting NIBRS state, it is a moderately low crime state; therefore, applying and testing the developed techniques and tools to additional data sets would have the potential to increase versatility and utility of methods.

This research resulted in the development of tools to assist with assessing data quality in NIBRS crime count data by providing guidelines for classifying zero reporting as missing data or zero counts and techniques for detecting outliers given one year of crime count data. Imputation methods seemed to reliably estimate missing data for producing stable crime trends as a means to count crime not reported. The research of alternative imputation methods for estimating missing crime data seemed promising and results provide the groundwork for refining methods using different state/jurisdictional or additional longitudinal data and extending methods to specific crime categories and smaller, regional areas of interest.

Endnotes

1. While the conversion to IBR among states continues, research directly comparing crime rates produced from the UCR and NIBRS data have been conducted to investigate consistency. Addington (2008) found that the variation between UCR and NIBRS crime rates depended on the crime and population group; Rantala (2000) concluded that differences between UCR and NIBRS crime rates were slight. Overall, the FBI NIBRS General FAQs concluded that there are not substantial differences in overall crime statistics between UCR and NIBRS (FBI, 2009).

2. West Virginia has used IBR data to study the validity of hate crime reporting, gun availability and crime in WV, selected populations groups and victim-offender relationships, domestic violence victimization in WV, patterns of violent crime and weapon use, juvenile arrests, and school violence (West Virginia Statistical Analysis Center, 2012). These studies were possible because of the additional information available in the IBR data, a clear advantage of using IBR data. Annual crime statistics are also compiled using the WV IBRS data; however, data are reported "as is" and the fidelity of the report is, at times, questioned.

3. West Virginia became the sixteenth state certified to submit data using NIBRS in September 1998. As of January 1, 1999, the WV repository began only accepting data in the IBR format. All WV IBRS data are currently submitted to, compiled, and maintained by the WVSP UCR Section of the state's repository. By 2006, all policing agencies were reporting IBR data, with only a small number of county and local agencies reporting no incidents.

4. FBI data quality measures include: cross-sectional comparisons made between the agency's crime rate and the median crime rate for the agency's strata, the lowest and highest ranked agencies are manually inspected for outliers; longitudinal comparisons (measuring consistency) made between the agency's current to the previous year's data as well as the agency's current data to the stratum's previous year's data, large increases or decreases are manually inspected for outliers; and proportional comparisons between the agency's crime distribution to the stratum's distribution which is taken as the norm, and highest and lowest deviations from the norm

distribution are manually inspected for outliers (see Akiyama & Propheter, 2005).

5. Maltz, Roberts, and Stasny recommended that for an agency with an average crime count of less than 1 per month, the Poisson distribution was assumed and the missing months were replaced with the mean monthly count; for an average crime count between 1 and 35, Poisson regression was used to estimate missing data; and for an average crime count over 35 per month, a seasonal autoregressive-moving-average (SARIMA) model was used (see Maltz, Roberts, & Stasny, 2006).

6. A random seed is a fixed starting point for the random number sequence; it is useful for replicating simulated results.

7. City, county, and state agencies had an equal chance of being deleted in the second simulation study.

8. Crime counts reported by some colleges or universities often had zeros reported in the summer months. Given the change in operations on campus during the summer months, it is reasonable that the zeros reported were true zeros. Zero-population agencies that serve very specific functions such as the Department of Natural Resources, Task Forces, and Turnpike, showed a consistent pattern of many consecutive zeros in the 2009, 2008, and 2007 data leading to the conclusion that sparse reporting is common and zeros reported are true zeros. Finally, the WVSP detachments are zero-population agencies, but historically have consistent nonzero crime counts and unreported data are extremely rare. Therefore, should zeros be observed in WVSP data in all crime types for a given month (or months) they are most likely indicative of missing data.

9. The SIZ method had significantly larger measures when comparing 70% to 80% missing in property crimes (see Appendix L for pairwise comparison p-values).

10. The methods established for classifying zeros, outlier detection, and graphical analysis can be applied to similar IBR datasets using a Microsoft Excel workbook equipped with macros (12 months of data are required). This tool was developed by the WV SAC to assist with data quality in NIBRS data and will result in a cleaned dataset for crime trend analysis (see attachment 'NIBRS Zero Classifier and Outlier Detection and Data Plots DEMO.xlsm' and 'NIBRS Zero Classifier and Outlier Detection and Data Plots.xlsm'). The workbook allows the user to input critical values for outlier detection to run desired 'what if' scenarios.

11. The procedure for determining WV3 population groups results in an equal number of agencies in each group, which meant that crime rates would be available for all simulation iterations and various missing data scenarios. This is unlike the other imputation methods that may have empty population groups due to random chance and the resulting crime rate would be set to zero as a penalty.

References

Addington, L. A. (2008). Assessing the Extent of Nonresponse Bias on NIBRS Estimates of Violent Crime. Journal of Contemporary Criminal Justice, 24(1), 32 – 49.

Akiyama, Y., & Propheter, S.K. (2005). Methods of Data Quality Control: For Uniform Crime Reporting Programs. Clarksburg, WV, Criminal Justice Information Services Division, Federal Bureau of Investigation. http://www.fbi.gov/about-us/cjis/ucr/additional-ucr-publications/ucr_methodsqc.pdf

Barnett-Ryan, C. (2007). Missing UCR Data and Divergence of the NCVS and UCR Trends. In . J. P. Lynch & L. A. Addington (Eds.), Introduction to the Uniform Crime Reporting Program (55 - 89). New York, NY: Cambridge University Press.

Dean, R. B., & Dixon, W. J. (1951). Simplified Statistics for Small Numbers of Observations. Analytical Chemistry, 23(4), 636-638.

Engels, J. M., & Diehr, P. (2003). Imputation of Missing Longitudinal Data: A Comparison of Methods. Journal of Clinical Epidemiology, 56(10), 968-976.

Federal Bureau of Investigation. (2000). Uniform Crime Reporting: National Incident-Based Reporting System: Volume 1: Data Collection Guidelines. Washington, DC: U.S. Government Printing Office.

Federal Bureau of Investigation. (2009). NIBRS General FAQs. Retrieved from: http://www.fbi.gov/about-us/cjis/ucr/frequently-asked-questions/nibrs_faqs/

Federal Bureau of Investigation. (2011a). Crime in the United States, 2010. Retrieved from: http://www.fbi.gov/about-us/cjis/ucr/crime-in-the-u.s/2010/crime-in-the-u.s.-2010/aboutucrmain

Federal Bureau of Investigation. (2011b). Crime in the United States, 2010. Retrieved from: http://www.fbi.gov/about-us/cjis/ucr/crime-in-the-u.s/2010/crime-in-the-u.s.-2010/about-cius

Gentle, J. E. (2002). Elements of Computational Statistics. New York, NY: Springer.

James, N. (2008). How crime in the United States is measured. Library of Congress Washington DC Congressional Research Service.

Justice Research and Statistics Association. (2012). Status of NIBRS in the States. Retrieved from: http://www.jrsa.org/ibrrc/background-status/nibrs_states.shtml

Little, R. J. A., & Rubin, D. B. (2002). Statistical Analysis of Missing Data (2nd ed). New York, NY: John Wiley & Sons.

Lynch, J. P., & Jarvis, J. P. (2008). Missing Data and Imputation in the Uniform Crime Reports and the Effects on National Estimates. Journal of Contemporary Criminal Justice, 24(1), 69 – 85.

Maltz, M. D. (1999). Bridging the Gaps in Police Crime Data. Report No. NCJ-1176365, Bureau of Justice Statistics, Office of Justice Programs, U.S. Department of Justice, Washington, D.C., September, 1999. http://www.ojp.usdoj.gov/bjs/pub/pdf/bgpcd.pdf

Maltz, M. D. (2007). Missing UCR Data and Divergence of the NCVS and UCR Trends. In . J. P. Lynch & L. A. Addington (Eds.), Understanding Crime Statistics: Revisiting the Divergence of the NCVS and UCR (269 – 294). New York, NY: Cambridge University Press.

Maltz, M. D., Roberts, C., & Stasny, E. A. (2006). Developing Imputation Methods for Crime Data. Report to the American Statistical Association Committee on Law and Justice Statistics. Retrieved from Ohio State University Academia.edu web site: http://osu.academia.edu/MichaelMaltz/Papers/1162511/Developing_Imputation_Methods_for_Crime_Data

National Institute of Justice. (2009). Sources of Crime Data: Uniform Crime Reports and the National Incident-Based Reporting System. Retrieved from: http://www.nij.gov/topics/crime/ucr-nibrs.htm

Nolan, J. J. (2004). Establishing the statistical relationship between population size and UCR crime rate: Its impact and implications. Journal of Criminal Justice, 32, 547-555.

Osgood, W. (2000). Poisson-based Regression Analysis of Aggregate Crime Rates. Journal of Quantitative Criminology, 16(1), 21-43.

Rantala, R. (2000). Effects of NIBRS on Crime Statistics. Bureau of Justice Statistics: Special report (NCJ 178890).

Roberts, A. (2007). Predictors of Homicide

Clearance by Arrest: An Event History Analysis of NIBRS Incidents. Homicide Studies, 11(2), 82 – 93.

Roberts, A. (2008). The Influences of Incident and Contextual Characteristics on Crime Clearance of Nonlethal Violence: A Multilevel Event History Analysis. Journal of Criminal Justice, 36(1), 61 – 71.

Tan, P. N., Steinbach, M., & Kumar, V. (2006). Introduction to Data Mining. Boston, MA: Addison-Wesley.

Targonski, J. R. (2011). A Comparison of Imputation Methodologies in the Offenses-Known Uniform Crime Reports. Final Report (U.S. Department of Justice). Retrieved from National Criminal Justice Reference Service web site: https://www.ncjrs.gov/pdffiles1/nij/grants/235152.pdf

Tremblay, A. (1994). Longitudinal Imputation of SIPP Food Stamp Benefits. Washington, DC, U.S. Bureau of the Census.

Sturges, H. A. (1926). The Choice of a Class Interval. Journal of the American Statistical Association, 21(153), 65-66.

West Virginia Statistical Analysis Center. (2012). Publications (2004 – present). Retrieved from http://www.djcs.wv.gov/SAC/Pages/publications-2004-present.aspx

Witten, I. H., & Frank, E. (2005). Data Mining: Practical Machine Learning Tools and Techniques (2nd ed.). San Francisco, CA: Morgan Kaufmann Publishers.

Funding Source

This project was supported by subaward agreement #WV11-2012-001, awarded by the Justice Research and Statistics Association (JRSA). The views expressed in this report are those of the authors and do not necessarily reflect the opinions of the U.S. Department of Justice, the Bureau of Justice Statistics, or the West Virginia Division of Justice and Community Services.

Recommended Citation

LaValle, C. R., Haas, S. M, Turley, E., and Nolan, J. J. (January, 2013). *Improving State Capacity for Crime Reporting: An Exploratory Analysis of Data Quality and Imputation Methods Using NIBRS Data.* Charleston, WV: Criminal Justice Statistical Analysis Center, Office of Research and Strategic Planning, Division of Justice and Community Services.

DJCS Administration

J. Norbert Federspiel, *DJCS Director*
Jeff Estep, *DJCS Chief Deputy Director*
Stephen M. Haas, *Ph.D., DJCS Deputy Director and ORSP Director*
Leslie Boggess, *Deputy Director*

Division of Justice & Community Services

1204 Kanawha Boulevard, East
Charleston, WV 25301
Phone: 304-558-8814
Fax: 304-558-0391
www.djcs.wv.gov

Appendix A: Summary UCR and NIBRS offences and arrest categories

Summary UCR	NIBRS
Offenses and arrests are reported for the following, listed in hierarchical order:	Offenses and arrests are reported for the following, for which a hierarchy does not apply:
Part I (Index) offenses Arson (not subject to hierarchy rule) Aggravated assault Burglary/Breaking or entering Forcible rape Larceny Motor vehicle theft Murder **Part II offenses** (only arrests are reported) Curfew and loitering law violations Disorderly conduct Driving under the influence Drug abuse violations Drunkenness Embezzlement Forgery and counterfeiting Fraud Gambling Liquor laws Offenses against family and children Other assaults Prostitution and commercial vice Runaways Sex offences (except forcible rape and prostitution) Stolen property: buying, receiving, possessing Suspicion Vagrancy Vandalism Weapons: carrying, possessing, other All other offenses (except traffic)	**Group A offenses** Arson Assault offenses Bribery Burglary/Breaking and entering Counterfeiting/Forgery Destruction/Damage/Vandalism of property Drug/Narcotic offenses Embezzlement Extortion/Blackmail Fraud offenses Gambling offenses Homicide offenses Kidnapping/Abduction Larceny/Theft offenses Motor vehicle theft Pornography/Obscene material Prostitution offenses Robbery Sex offenses, Forcible Sex offenses, Nonforcible Stolen property offenses Weapon law violations **Group B offenses** (only arrests are reported) Bad Checks Curfew/Loitering/Vagrancy violations Disorderly conduct Driving under the influence Drunkenness Family offenses, Nonviolent Liquor law violations Peeping Tom Runaway Trespass of real property All other offenses
Adapted from "Effects of NIBRS on Crime Statistics," by Rantala, R., 2000, Bureau of Justice Statistics: Special report (NCJ 178890).	

Appendix B: Additional details on the Yi outlier detection method

The Yi method compares an agency's monthly data to its annual median; thus the comparison is relative to the specific agency.

Specifically, Yi is calculated by dividing each the monthly crime count by the agency's median crime count where i = 1 to 12 corresponding to each month of the year. Each agency will have a Yi statistic for each month it reported data. This statistic is easily interpretable as it measures the number of times the monthly crime count is compared to the median.

For example, Yi = 1 indicates that the monthly crime count and median are the same. This is what we would expect in data that is consistent.

As Yi grows larger (or smaller), anomalous data are suspected. A large Yi value indicates a larger than expected observation while a small Yi indicates a lower than expected crime count. Given that the terms 'larger' (and 'smaller') are relatively vague, Yi values become suspect when it is compared to some user-defined critical value.

Let the user-defined critical value for Yi be x. If Yi is greater than x (or less than 1/x), the data for month i is flagged as a potential outlier.

For example, setting x = 5 translates to any monthly count 5 times larger or 5 times smaller than the median is identified as a potential outlier.

One disadvantage to this method is that critical values are not established, rather it is user-defined. When selecting the critical value, if x is too liberal, too many agencies will be identified and many falsely flagged. When x is set too conservative, there is the potential for irregular data to go undetected. Therefore, it is helpful to test several scenarios for x to identify agency data. In analyzing the WV IBRS data, it was found that selecting a critical value for x such that 25% of agencies were identified as having potential irregular reporting was ideal. In the WV IBRS data, this was achieved when x = 4 (and 1/4 = 0.25).

An additional issue with this method is when the median or any monthly report is zero. When the median is zero, the Yi statistic is undefined. When a monthly report is zero, the Yi statistic is zero and will always be flagged as a potential outlier (since zero is always less than 1/x). However, since zero reports were addressed as the first step to assessing data quality, the zero reports that remain are deemed valid.

As a result, agencies with a median equal to zero are omitted from this outlier test. Likewise, when Yi = 0, the month is excluded from outlier detection.

This method makes no assumption about the data distribution and borrows concepts from the cross-sectional outlier detection used by the FBI on the UCR data. The FBI employs this type of comparison by calculating the ratio between agency data and its stratum's median (see Akiyama & Propheter, 2005, p. 12-13).

Appendix C: Additional details on the Rr outlier detection method

The Rr statistic assumes that agencies with outliers have distinctive features in relation to calculated distances. The Rr statistic compares a 'gap' to a 'range' much like Dixon's Q test. A 'large' gap or range is may indicate an outlier, where a 'small' gap or range would indicate no outlier since agencies with consistent data would have a 'small' spread.

However, terms such as 'large' and 'small' are relative. In fact, agencies with larger volumes of crime tend to have larger data spread (the correlation between annual crime total and range was $r = 0.87$ for property crimes and $r = 0.88$ for violent crimes in the 2009 WV IBRS data). In an attempt to account for the relationship between total and range, the denominator of the Rr statistic was adjusted and seemed to emphasize the effect of outlying values relative to the total of agency's crime count. As a result, a large Rr statistic is indicative of abnormal data.

To examine the mechanics of the Rr statistic, let's consider the statistic's numerator, or 'gap', the distance between the maximum or minimum value from the median. This distance can account for multiple outliers. A 'large' gap would most likely indicate an outlier, where a 'small' gap would indicate no outlier.

The Rr statistic's denominator involves an agency's total and range. A 'large' range is suggestive of an outlier since agencies with consistent data would have a 'small' range.

As previously mentioned, the relationship between crime volume and range was adjusted by taking the total and dividing it by the range. To illustrate, let

G = 'large' gap,
T = 'large' total,
R = 'large' range,
g = 'small' gap,
t = 'small' total,
r = 'small' range,
D = 'large' denominator, and
d = 'small' denominator
where $D > d$.

In calculating the *denominator* of Rr, we know that the total ≥ range; $R > r$; and $T > t$, therefore,

$T / R < T / r$ so
$T/R = d$ (but ≥ 1) and
$T/r = D$.
$t / R < t / r$ so
$t/R = d$ (but ≥ 1) and
$t/r = D$.

The two ratios above that are suspected to contain outliers have a 'large' range (ratios with 'R') and resulted in a 'small' denominator or 'd'.

In calculating the final Rr statistic, we assume that $G > g$; then,

$G / D < G / d$ and
$g / D < g / d$.
Thus, $g / d < G / d$.

Given the illustration above, the ratio G / d would yield the largest value of Rr. The final Rr statistic resulting in G / d suggests that the data has at least one irregularity since both the gap is 'large' (or 'G') *and* the denominator is 'small' (or 'd').

To summarize,

G / d would indicate a 'large' gap and range suggestive of potential outliers;

G / D would indicate a 'large' gap but the ratio between total and range do not suggest anomalies;

g / D would indicate a 'small' gap and a relationship between the total and range that do not suggest outliers;

g / d indicates that although 'd' was said to potentially contain anomalous data, the fact that the gap is 'small' is contradictory and therefore, does not suggest outlying data.

In conclusion, an agency with a 'large' Rr is suggestive of having irregular data. Since this outlier test is novel, critical values for Rr have not been developed. Rather, it is suggested that the resulting RrTop and RrBottom statistics for each agency are sorted in descending order; the agencies with the largest RrTop and RrBottom values are to be checked for outliers. See next page (Appendix G) for the top 50 ranked Rr statistics using the 2009 WV IBRS property and violent data.

Appendix D: Frequency charts and corresponding portions of city populations according to various population groups a.) U.S. cities according to FBI population groupings 1-6, b.) WV cities according to FBI population groupings 1-6, c.) WV cities according to scaled FBI population groupings, d.) Population group limits for FBI and alternative non=reporting agency imputation methods

Table a.)

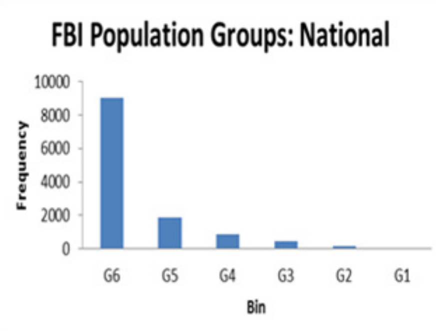


Table b.)

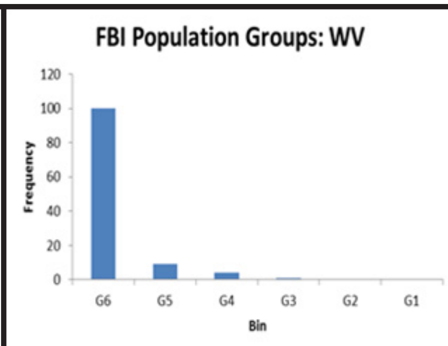
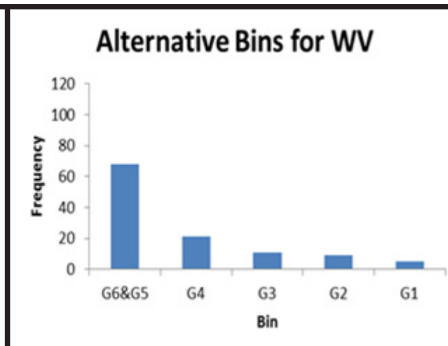


Table c.)



Population Group (City)	National Agencies, FBI Population Bins	WV Agencies, FBI Population Bins	WV Agencies, WV1 Population Bins
G1	0.0058	0.0000	0.0439
G2	0.0151	0.0000	0.0789
G3	0.0377	0.0088	0.0965
G4	0.0678	0.0351	0.1842
G5	0.1524	0.0789	0.3860
G6	0.7211	0.8772	0.2105

Table d.) Population groups for FBI and alternative non-reporting agency imputation methods.

Group	FBI	WV1	WV2	WV3 (Property)	WV3 (Violent)
1	250,000+	25,000+	23,917+	11,026+	11,821+
2	100,000 - 249,999	10,000 - 24,999	11,822 - 23,917	5,399 - 11,026	6,266 - 11,821
3	50,000 - 99,999	5,000 - 9,999	5,066 - 11,821	3,280 - 5,398	3,396 - 6,265
4	25,000 - 49,999	2,500 - 4,999	1,754 - 5,065	2,563 - 3,279	2,361 - 3,395
5	10,000 - 24,999	1,000 - 2,499 + colleges	Less than 1,754 + colleges	1,464 - 2,562	1,464 - 2,360
6	Less than 10,000 + colleges	Less than 1,000	N/A	Less than 1,464 + colleges	Less than 1,464 + colleges
7	Non-MSA counties & SP	Non-MSA counties & SP	Non-MSA counties & SP	Non-MSA counties & SP	Non-MSA counties & SP
8	MSA counties & SP	MSA counties & SP	MSA counties & SP	MSA counties & SP	MSA counties & SP

Appendix E: Examples of zero reporting patterns in agency data (V=violent, P=property, Z=non-index)

Table a: Example of consistent crime reporting with one month of zeros reported in all crime types (labeled V, P, and Z); irregularities are highlighted in grey

Agency	Jan.	Feb.	Mar.	Apr.	May.	Jun.	Jul.	Aug.	Sept.	Oct	Nov.	Dec.	Total
Mason Co SO_V07	1	0	0	1	1	3	0	1	0	1	1	0	9
Mason Co SO_P07	17	8	19	20	19	13	15	18	10	20	19	13	191
Mason Co SO_Z07	17	13	22	15	20	13	15	21	21	18	11	17	203
Mason Co SO_V08	0	2	1	1	0	3	0	0	2	0	1	0	10
Mason Co SO_P08	12	11	14	7	31	17	26	16	13	11	14	13	185
Mason Co SO_Z08	26	18	17	17	17	9	21	11	8	11	11	5	171
Mason Co SO_V09	1	1	1	2	0	0	1	0	0	1	1	0	8
Mason Co SO_P09	11	11	15	16	17	0	22	25	13	14	19	13	176
Mason Co SO_Z09	11	15	12	18	7	0	10	11	9	8	6	3	110

Table b: Example of nonzero reporting months with several months of consecutive zeros reported in all crime types (labeled V, P, and Z); irregularities are highlighted in grey

Agency	Jan.	Feb.	Mar.	Apr.	May.	Jun.	Jul.	Aug.	Sept.	Oct	Nov.	Dec.	Total
Grafton PD_V07	0	0	0	0	1	1	0	0	0	0	0	0	2
Grafton PD_P07	5	3	2	2	1	0	6	6	2	1	1	1	30
Grafton PD_Z07	3	3	2	2	2	3	7	6	2	2	2	0	34
Grafton PD_V08	0	0	0	0	0	0	0	0	0	0	0	0	0
Grafton PD_P08	1	2	1	4	0	1	4	2	2	2	1	0	20
Grafton PD_Z08	5	0	2	2	0	2	1	2	3	4	1	2	24
Grafton PD_V09	1	0	0	0	0	0	0	0	0	0	0	0	1
Grafton PD_P09	1	0	0	0	0	0	0	0	0	0	0	0	1
Grafton PD_Z09	1	2	1	1	0	0	1	0	0	0	0	0	6

Table c: Examples of consecutive zeros reported in all crime types in multiple years of data; patterned zero reporting is highlighted in grey

Agency	Jan.	Feb.	Mar.	Apr.	May.	Jun.	Jul.	Aug.	Sept.	Oct	Nov.	Dec.	Total
Concord College _V07	0	0	0	0	0	0	0	0	0	0	0	0	0
Concord College _P07	1	0	1	4	0	0	0	0	0	0	0	0	6
Concord College _Z07	1	0	4	0	0	0	0	2	3	0	0	0	10
Concord College _V08	0	1	0	1	0	0	0	0	0	0	0	0	2
Concord College _P08	0	2	3	4	2	0	0	3	1	1	1	0	17
Concord College _Z08	0	2	2	3	0	0	0	3	1	8	2	0	21
Concord College _V09	0	1	0	0	0	0	0	0	0	0	1	0	2
Concord College _P09	1	3	0	5	1	2	0	0	6	4	5	9	36
Concord College _Z09	0	15	4	2	1	0	0	0	5	9	0	0	36

Agency	Jan.	Feb.	Mar.	Apr.	May.	Jun.	Jul.	Aug.	Sept.	Oct	Nov.	Dec.	Total
Paden City PD_V07	0	0	0	0	0	0	1	1	0	0	0	0	2
Paden City PD_P07	3	0	0	0	0	0	0	0	0	0	0	0	3
Paden City PD_Z07	0	3	1	2	0	0	0	1	1	1	1	0	10
Paden City PD_V08	1	1	0	0	0	0	0	0	1	0	2	0	5
Paden City PD_P08	0	1	0	0	0	0	0	0	0	0	1	0	2
Paden City PD_Z08	0	0	0	1	0	2	0	0	0	1	0	0	4
Paden City PD_P09	0	0	1	0	0	0	0	0	0	1	0	0	2
Paden City PD_V09	0	0	0	0	0	0	0	0	0	0	0	0	0
Paden City PD_Z09	2	0	2	2	0	0	1	0	0	1	2	2	12

Appendix F: Data, descriptive statistics, and outlier detection statistics for Huntington PD property (P) and violent (V) crime counts (2009)

Table 1: Data for monthly aggregate property and violent crime counts.

Agency	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sept	Oct	Nov	Dec
Huntington P	244	7	13	14	223	276	255	266	273	295	278	248
Huntington V	24	3	5	3	21	29	28	24	21	31	27	31

Table 2: Descriptive Statistics for monthly aggregate property and violent crime count data.

Agency	Total	Mean	Median	Min.	Max.	Range	SD	IQR
Huntington P	2392	199.33	251.5	7	295	288	114.88	103
Huntington V	247	20.58	24	3	29	28	10.74	11.25

SD = standard deviation, IQR = interquartile range

Table 3: Outlier Detection Statistics for property and violent crime count data using the Standard Deviation method (SD), Box Plot method (BP), Dixon's Q test (Q), and the Ratio of Ranges test (Rr).

Agency	SD_{Upper}	SD_{Lower}	BP_{Upper}	BP_{Lower}	Q_{Top}	Q_{Bottom}	Rr_{Top}	Rr_{Bottom}
Huntington P	543.97	-145.31	582.75	-138.25	0.067	0.026	5.237	29.438
Huntington V	52.80	-11.64	62.00	-16.75	0.071	0.071	0.794	2.381

Upper = upper threshold value, Lower = lower threshold value, Top = suspected value too large, Bottom = suspect value too small. Data greater than the 'Upper' and less than the 'Lower' thresholds are flagged as outliers and require manual inspection.

Critical Values for Dixon's Q test are $QCriticalValue_{\alpha=0.05} = 0.546$ and $QCriticalValue_{\alpha=0.01} = 0.642$; if QCriticalValue < Q_{Top} or Q_{Bottom}, then a potential outlier exists and the data requires manual inspection.

Critical Values for Rr do not exist, refer to Appendix C for advice on flagging potential outliers.

Table 4: Outlier Detection Statistics for property and violent crime count data using the Ratio to Median test (Yi where i = month).

Agency	Y_{Jan}	Y_{Feb}	Y_{Mar}	Y_{Apr}	Y_{May}	Y_{Jun}	Y_{Jul}	Y_{Aug}	Y_{Sept}	Y_{Oct}	Y_{Nov}	Y_{Dec}
Huntington P	0.97	0.03	0.05	0.06	0.89	1.10	1.01	1.06	1.09	1.17	1.11	0.99
Huntington V	1.00	0.13	0.21	0.13	0.88	1.21	1.17	1.00	0.88	1.29	1.13	1.29

Critical Values for Yi do not exist, refer to Appendix B for advice on flagging potential outliers.

Huntington PD's data was selected as the litmus test for effectiveness of outlier detection methods for several reasons. First, the aggregate crime counts for Feb., Mar., and Apr. are undeniably too low. While the violent crime counts for Feb., Mar., and Apr. appear irregular, the severity of irregularities observed in property crime counts are obvious. A second important feature of Huntington's data was that more than one month of irregular data exists. Thus, the most effective outlier detection method must be able to function when the sample size is small and more than one outlier exists.

Using Huntington's data, the well-known Standard Deviation (SD) and Box Plot (BP) methods fail to identify Huntington PD as having outliers in violent or property crimes. Both SD and BP upper and lower thresholds are inflated (see Table 3), therefore all the agency's data are not beyond the thresholds. Further, Dixon's Q test also failed even though it was specifically developed to handle data with a small sample size (see Table 3, the values of QTop and QBottom are not greater than critical values).

The alternative ratio tests, Ratio of Ranges (Rr), developed to account for multiple outliers and Yi successfully identified Huntington's data as being irregular. The Rr statistic in property is quite large; when the Rr value in violent crimes is compared to other agencies, it is considered large (see Appendix C). The ratio of monthly count to median (Yi) identified Feb., Mar., and Apr. in Huntington's property and violent data when Yi > 4 or Yi < 0.25. Therefore, the Rr and Yi statistics seem most suited for twelve months of aggregate agency data where more than one outlier may exist.

Appendix G: Top 50 ranked Rr statistics for 2009 WV IBRS property and violent data

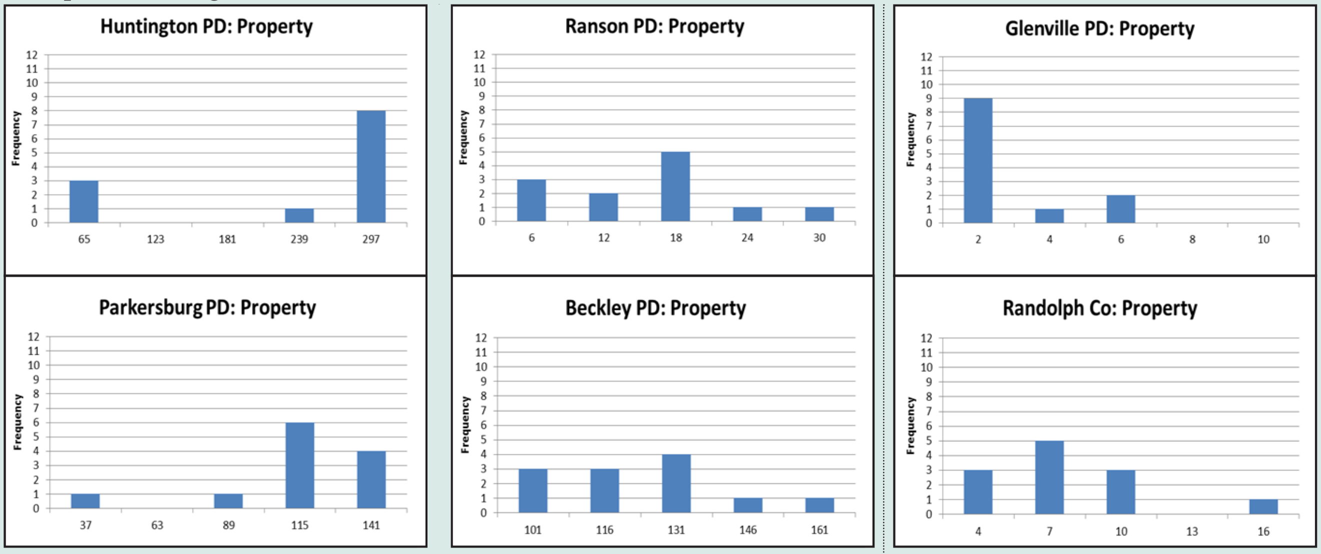
Ranked Rr Statistics for Property Crime				
Rank	Agency	RrTop	RrBottom	LargestRr
1	Huntington PD	5.24	29.44	29.44
2	Parkersburg PD	2.94	10.28	10.28
3	Marion Co SO	9.27	1.24	9.27
4	Oak Hill PD	6.05	4.14	6.05
5	Wayne SP	3.47	2.59	3.47
6	Stonewood PD	3.20	0.00	3.20
7	Winfield PD	3.13	0.00	3.13
8	Ranson PD	3.09	2.51	3.09
9	Berkeley Co SO	1.91	2.81	2.81
10	Nicholas Co SO	1.37	2.79	2.79
11	WVUniversityPolice	2.77	1.35	2.77
12	Monongalia Co SO	2.51	2.72	2.72
13	Monongah PD	2.67	0.00	2.67
14	Buckhannon PD	2.66	0.82	2.66
15	Wayne Co SO	2.56	1.68	2.56
16	Moundsville PD	2.53	1.59	2.53
17	Tyler Co SO	2.50	0.00	2.50
18	Weston SP	2.40	0.74	2.40
19	Kanawha Co SO	1.48	2.40	2.40
20	South Charleston PD	1.77	2.40	2.40
21	Wyoming Co SO	2.39	2.02	2.39
22	Wheeling PD	2.37	1.29	2.37
23	Harrison Co SO	2.34	1.55	2.34
24	Vienna PD	1.51	2.30	2.30
25	Charleston PD	2.23	1.32	2.23
26	Keyser SP	2.19	1.22	2.19
27	Point Pleasant PD	2.12	1.03	2.12
28	Beckley PD	2.07	1.56	2.07
29	Mason Co SO	1.49	2.06	2.06
30	Williamson SP	2.02	0.81	2.02
31	Doddridge Co SO	2.00	0.00	2.00
32	Richwood SP	2.00	0.25	2.00
33	Morgan Co SO	1.39	1.92	1.92
34	Charles Town PD	1.90	0.68	1.90
35	Fort Gay PD	1.89	0.00	1.89
36	Wayne PD	1.89	0.00	1.89
37	Elkins SP	1.88	0.69	1.88
38	Philippi SP	1.88	1.25	1.88
39	Roane Co SO	1.79	0.00	1.79
40	Ripley PD	1.78	1.03	1.78
41	Bridgeport PD	1.55	1.75	1.75
42	Morgantown PD	1.75	0.90	1.75
43	So Charleston SP	1.41	1.73	1.73
44	Harpers Ferry PD	1.73	0.19	1.73
45	Martinsburg SP	1.38	1.72	1.72
46	Greenbrier Co SO	1.20	1.72	1.72
47	St Albans PD	1.71	0.76	1.71
48	Spencer PD	1.67	0.89	1.67
49	Logan SP	1.67	0.99	1.67
50	Braxton Co SO	1.67	0.49	1.67

Ranked Rr Statistics for Violent Crime				
Rank	Agency	RrTop	RrBottom	LargestRr
1	Marion Co SO	2.55	0.36	2.55
2	WVUniversityPolice	2.50	0.44	2.50
3	Huntington PD	0.79	2.38	2.38
4	Ripley PD	2.25	0.00	2.25
5	Petersburg SP	2.00	0.00	2.00
6	Westover PD	2.00	0.00	2.00
7	Princeton PD	1.92	0.64	1.92
8	Oceana PD	1.78	0.00	1.78
9	Morgantown PD	1.54	0.70	1.54
10	South Charleston PD	1.48	0.89	1.48
11	Grant Co SO	1.45	0.00	1.45
12	Wyoming Co SO	1.40	0.56	1.40
13	Parkersburg PD	0.76	1.37	1.37
14	Gilmer Co SO	1.33	0.00	1.33
15	Jackson Co SO	1.33	0.00	1.33
16	Ranson PD	1.33	0.00	1.33
17	Summersville SP	1.33	0.00	1.33
18	Braxton Co SO	1.29	0.00	1.29
19	Oak Hill SP	1.29	0.00	1.29
20	Ravenswood PD	1.29	0.00	1.29
21	Mc Dowell Co SO	1.27	0.18	1.27
22	Monongalia Co SO	1.23	0.85	1.23
23	Boone Co SO	1.13	0.38	1.13
24	Charles Town PD	1.13	0.00	1.13
25	Nutter Fort PD	1.09	0.36	1.09
26	Sutton SP	1.08	0.15	1.08
27	Raleigh Co SO	0.69	1.04	1.04
28	Ohio Co SO	1.02	0.85	1.02
29	Barbour Co SO	1.00	0.00	1.00
30	BCI - Beckley	1.00	0.00	1.00
31	BCI - Bluefield	1.00	0.00	1.00
32	DANVILLE SP	1.00	0.25	1.00
33	Grafton PD	1.00	0.00	1.00
34	Harrisville PD	1.00	0.00	1.00
35	Huntington SP	1.00	0.50	1.00
36	Kingwood PD	1.00	0.00	1.00
37	Lewisburg PD	1.00	0.00	1.00
38	Moundsville SP	1.00	0.00	1.00
39	Mullens PD	1.00	0.00	1.00
40	New Haven PD	1.00	0.00	1.00
41	Parsons SP	1.00	0.00	1.00
42	Pleasants Co SO	1.00	0.00	1.00
43	Rainelle SP	1.00	0.00	1.00
44	Ronceverte PD	1.00	0.00	1.00
45	tf Parkersburg VioCrNarc	1.00	0.00	1.00
46	Wayne Co SO	1.00	0.00	1.00
47	Wellsburg SP	1.00	0.00	1.00
48	WVU Tech-Security	1.00	0.00	1.00
49	Kanawha Co SO	0.90	0.99	0.99
50	Fayette Co SO	0.28	0.98	0.98

Appendix H: Outlier Plots for agency data with and without outliers

Histogram

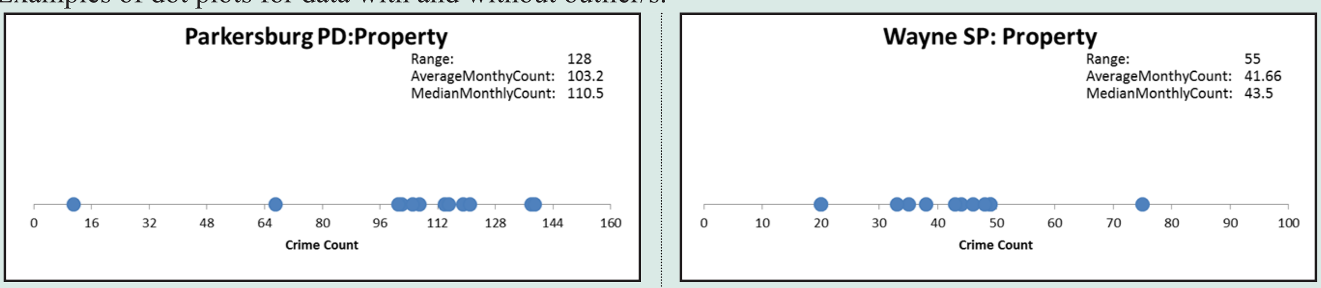
Examples of histograms for data with and without outlier/s.



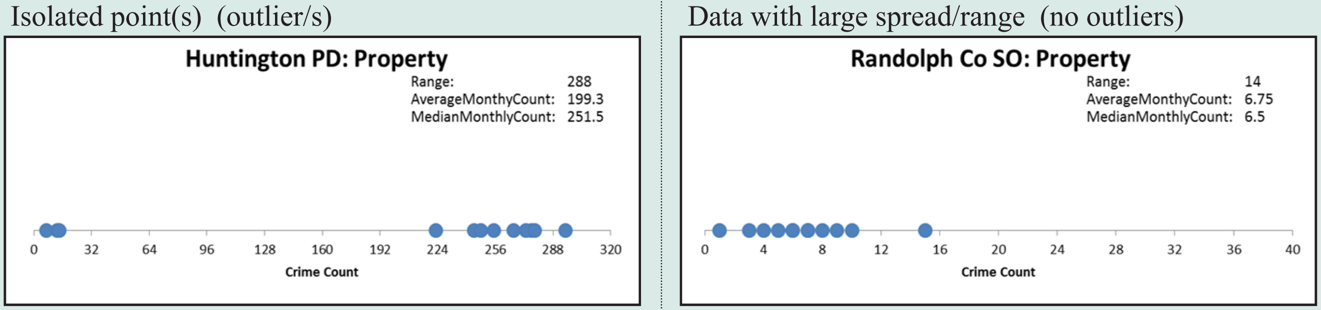
Left-skewed (outlier/s) | Symmetric (no outliers) | Right-skewed (no outliers)

Dot Plot

Examples of dot plots for data with and without outlier/s.



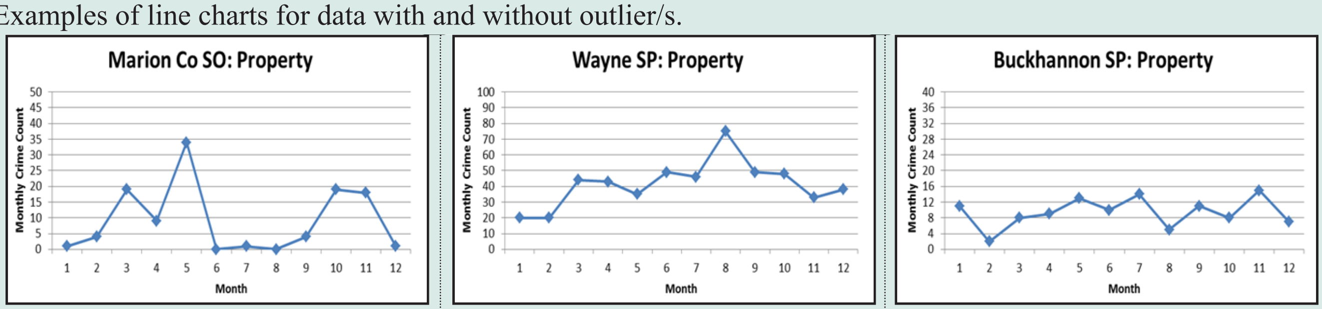
Isolated point(s) (outlier/s) | Data with large spread/range (no outliers)



Small isolated cluster (outliers) | Data with moderate spread/range (no outliers)

Line Chart

Examples of line charts for data with and without outlier/s.



Sharp peak(s)/valley(s) (outlier/s) | Seasonality (no outliers) | Consistent reporting (no outliers)

Appendix I: Descriptive statistics, outlier statistics, and outlier plots for agencies identified with irregular data. (1 agency in violent crimes and 4 agencies in property crimes)

Table 1: Descriptive Statistics for agencies identified with outliers (V = violent crimes, P = property crimes).

Agency	Total	Mean	Median	Min.	Max.	Range	SD	IQR
Huntington V	247	20.58	24	3	29	28	10.74	11.25
Huntington P	2392	199.33	251.5	7	295	288	114.88	103
Marion Co P	110	9.17	4	0	34	34	10.91	17.25
Parkersburg P	1239	103.25	110.5	11	139	128	34.58	17.75
Vienna P	221	18.42	20.5	3	32	29	8.66	7.5

Table 2: Outlier Detection Statistics for agencies identified with outliers using the Standard Deviation method (SD), Box Plot method (BP), Dixon's Q test (Q), and the Ratio of Ranges test (Rr).

Agency	SD_{Upper}	SD_{Lower}	BP_{Upper}	BP_{Lower}	Q_{Top}	Q_{Bottom}	Rr_{Top}	Rr_{Bottom}
Huntington V	52.80	-11.64	62.00	-16.75	0.071	0.071	0.79	2.38
Huntington P	543.97	-145.31	582.75	-138.25	0.067	0.026	5.24	29.44
Marion Co P	43.11	-23.11	71.00	-51.50	0.441	0.053	9.27	1.24
Parkersburg P	206.98	-0.48	172.75	48.50	0.250	0.709	2.94	10.28
Vienna P	44.40	-7.56	44.75	-7.75	0.360	0.192	1.51	2.30

Table 3: Outlier Detection Statistics for agencies identified with outliers using the Ratio to Median test (Yi where i = month).

Agency	Y_{Jan}	Y_{Feb}	Y_{Mar}	Y_{Apr}	Y_{May}	Y_{Jun}	Y_{Jul}	Y_{Aug}	Y_{Sept}	Y_{Oct}	Y_{Nov}	Y_{Dec}
Huntington V	1.00	0.13	0.21	0.13	0.88	1.21	1.17	1.00	0.88	1.29	1.13	1.29
Huntington P	0.97	0.03	0.05	0.06	0.89	1.10	1.01	1.06	1.09	1.17	1.11	0.99
Marion Co P	0.25	1.00	4.75	2.25	8.50	0.00	0.25	0.00	1.00	4.75	4.50	0.25
Parkersburg P	0.10	0.61	0.92	0.97	1.25	1.03	0.91	1.08	1.10	1.26	0.95	1.04
Vienna P	1.02	0.88	1.02	1.41	1.12	1.07	0.83	0.98	0.15	0.34	0.39	1.56

Appendix J: Agency MAEave and RMSEave for property & violent data

Agency: Property

10%		20%		30%		40%		50%		60%		70%		80%	
Method	MAEave	Method	MAEave	Method	MAEave	Method	MAEave	Method	MAEave	Method	MAEave	Method	MAEave	Method	MAEave
SIZ	128.03	SIZ	127.81	SIZ	128.56	SIZ	129.18	SIZ	129.95	SIZ	131.52	SIZ	134.03	SIZ	138.59
SNI	135.55	SNI	135.05	SNI	135.58	SNI	136.02	SNI	136.20	SNI	137.11	SNI	138.58	SNI	141.22
WV3	166.53	WV3	166.86	WV3	169.43	WV3	171.01	WV3	173.59	WV3	181.72	WV3	194.64	WV3	215.90
WV2	170.29	WV2	171.22	WV2	175.48	WV2	178.05	WV2	181.66	WV1	191.71	WV1	206.06	FBI	227.71
WV1	170.29	WV1	171.35	WV1	175.64	WV1	178.24	WV1	181.72	WV2	191.84	WV2	206.39	WV2	228.59
FBI	175.71	FBI	177.23	FBI	182.17	FBI	184.13	FBI	186.63	FBI	194.81	FBI	207.53	WV1	229.14
Delete	210.07	Delete	209.51	Delete	210.80	Delete	210.84	Delete	210.52	Delete	210.18	Delete	210.02	Delete	209.63

Method	RMSEave	Method	RMSEave	Method	RMSEave	Method	RMSEave	Method	RMSEave	Method	RMSEave	Method	RMSEave	Method	RMSEave
SIZ	201.08	SIZ	205.54	SIZ	208.86	SNI	210.78	SNI	212.52	SNI	215.07	SNI	219.18	SNI	225.81
SNI	202.01	SNI	206.00	SNI	208.91	SIZ	210.82	SIZ	213.34	SIZ	217.20	SIZ	223.16	SIZ	233.07
WV3	248.23	WV3	255.56	WV3	264.03	WV3	269.16	WV3	276.27	WV3	298.51	WV3	337.29	WV3	401.54
WV1	258.23	WV1	268.12	WV1	282.04	WV1	290.45	WV1	301.70	WV1	330.06	WV1	373.44	WV1	439.47
WV2	258.67	WV2	268.83	WV2	283.01	WV2	291.78	WV2	303.42	WV2	332.56	WV2	376.76	WV2	442.12
FBI	285.73	FBI	304.16	FBI	324.92	FBI	332.51	FBI	341.18	FBI	361.90	FBI	399.42	FBI	455.22
Delete	363.68	Delete	378.34	Delete	389.79	Delete	392.51	Delete	393.96	Delete	394.29	Delete	395.83	Delete	395.82

Agency: Violent

Method	MAEave	Method	MAEave	Method	MAEave	Method	MAEave	Method	MAEave	Method	MAEave	Method	MAEave	Method	MAEave
SIZ	15.79	SIZ	15.70	SIZ	15.79	SIZ	15.73	SIZ	15.83	SIZ	15.86	SIZ	15.89	SIZ	16.22
WV3	15.81	WV3	15.94	WV3	16.15	WV3	16.27	WV3	16.65	WV3	17.07	SNI	17.06	SNI	16.98
WV2	16.36	WV2	16.55	WV2	16.84	WV2	17.03	SNI	17.29	SNI	17.21	WV3	18.23	WV3	20.65
WV1	16.44	WV1	16.64	WV1	16.88	WV1	17.04	WV2	17.47	WV1	18.05	WV1	19.49	FBI	22.10
FBI	16.49	FBI	16.72	FBI	17.01	FBI	17.24	WV1	17.47	WV2	18.09	WV2	19.57	WV1	22.17
SNI	17.47	SNI	17.34	SNI	17.38	SNI	17.27	FBI	17.74	FBI	18.29	FBI	19.63	WV2	22.22
Delete	22.89	Delete	-22.91	Delete	23.05	Delete	22.98	Delete	23.14	Delete	23.18	Delete	23.10	Delete	23.07

Method	RMSEave	Method	RMSEave	Method	RMSEave	Method	RMSEave	Method	RMSEave	Method	RMSEave	Method	RMSEave	Method	RMSEave
WV3	26.31	WV3	27.89	WV3	29.02	WV3	29.61	WV3	30.72	SIZ	31.22	SIZ	31.50	SIZ	32.42
SIZ	27.27	SIZ	28.81	SIZ	29.84	SIZ	30.21	SIZ	30.89	WV3	31.90	SNI	32.46	SNI	32.86
SNI	28.01	SNI	29.56	SNI	30.68	SNI	31.10	SNI	31.84	SNI	32.24	WV3	35.15	WV3	42.36
WV1	28.05	WV1	30.09	WV1	31.56	WV1	32.46	WV1	34.01	WV1	35.94	WV1	40.45	WV1	48.46
WV2	28.11	WV2	30.13	WV2	31.67	WV2	32.64	WV2	34.19	WV2	36.24	WV2	40.85	WV2	48.87
FBI	30.33	FBI	33.88	FBI	36.42	FBI	38.12	FBI	40.19	FBI	41.95	FBI	45.56	FBI	51.99
Delete	43.26	Delete	46.68	Delete	48.51	Delete	49.32	Delete	50.46	Delete	50.94	Delete	51.18	Delete	51.35

Appendix K: State MAEtot and RMSEtot for property & violent data

State: Property, total state property crime from simulation: 36,556

10%		20%		30%		40%		50%		60%		70%		80%	
Method	MAETot	Method	MAETot	Method	MAETot	Method	MAETot	Method	MAETot	Method	MAETot	Method	MAETot	Method	MAETot
WV3	925.58	WV3	1402.93	WV3	1914.92	WV3	2396.97	WV3	2962.22	WV3	4319.18	SNI	6038.30	SNI	7485.56
WV1	960.64	WV1	1472.72	WV1	2050.85	WV1	2604.22	WV1	3196.01	WV1	4576.87	WV3	6302.39	SIZ	8599.92
SNI	962.40	WV2	1475.18	WV2	2057.92	WV2	2627.95	WV2	3217.83	WV2	4607.84	WV1	6573.86	WV3	9724.53
WV2	969.14	FBI	1635.41	FBI	2296.64	FBI	2821.30	FBI	3405.61	FBI	4706.25	WV2	6611.47	FBI	9809.83
FBI	1092.08	SNI	1659.25	SNI	2448.88	SNI	3263.09	SNI	4139.84	SNI	5047.64	FBI	6656.51	WV1	9856.23
SIZ	1178.15	SIZ	2137.76	SIZ	3191.32	SIZ	4258.69	SIZ	5306.72	SIZ	6346.47	SIZ	7311.00	WV2	9898.08
Delete	3781.29	Delete	7333.01	Delete	11172.26	Delete	14758.95	Delete	18315.63	Delete	22068.90	Delete	25622.26	Delete	29348.67

Method	RMSETot	Method	RMSETot	Method	RMSETot	Method	RMSETot	Method	RMSETot	Method	RMSETot	Method	RMSETot	Method	RMSETot
WV3	37.25	WV3	56.59	WV3	80.32	WV3	99.68	WV3	128.11	SNI	183.64	SNI	221.85	SNI	274.16
SNI	37.88	WV1	59.42	WV1	85.76	WV1	108.09	WV1	134.86	FBI	202.77	SIZ	254.83	SIZ	304.16
WV1	38.93	WV2	59.67	WV2	86.11	WV2	108.46	WV2	135.41	WV1	204.46	FBI	409.51	FBI	525.31
WV2	39.24	SNI	63.58	SNI	90.85	FBI	113.12	FBI	139.10	WV2	205.30	WV2	414.13	WV2	529.04
SIZ	44.24	FBI	66.44	FBI	92.33	SNI	119.26	SNI	149.81	WV3	205.56	WV1	414.98	WV1	531.29
FBI	45.19	SIZ	77.09	SIZ	111.66	SIZ	146.31	SIZ	181.28	SIZ	218.52	WV3	419.56	WV3	548.45
Delete	127.26	Delete	239.47	Delete	359.79	Delete	472.30	Delete	583.72	Delete	701.26	Delete	812.77	Delete	929.77

State: Violent, total state violent crime from simulation: 4,058

Method	MAETot	Method	MAETot	Method	MAETot	Method	MAETot	Method	MAETot	Method	MAETot	Method	MAETot	Method	MAETot
WV3	102.31	WV3	166.94	WV3	226.39	WV3	286.03	WV3	366.08	WV3	467.80	WV3	670.51	SNI	928.62
WV1	109.28	WV1	180.93	WV2	247.66	WV1	314.79	WV2	407.29	WV1	534.88	SNI	751.16	SIZ	987.30
WV2	109.80	WV2	181.22	WV1	248.41	WV2	315.66	WV1	407.61	WV2	541.52	WV1	781.18	WV3	1064.01
SNI	110.31	SNI	183.92	SNI	261.68	SNI	350.18	SNI	463.48	SNI	602.84	WV2	792.53	WV1	1182.11
FBI	125.15	FBI	217.89	FBI	309.16	FBI	394.81	FBI	502.68	FBI	637.19	SIZ	828.24	WV2	1202.25
SIZ	125.20	SIZ	228.57	SIZ	328.32	SIZ	436.36	SIZ	558.59	SIZ	687.29	FBI	860.13	FBI	1223.51
Delete	412.05	Delete	824.64	Delete	1221.42	Delete	1631.93	Delete	2036.35	Delete	2456.83	Delete	2864.71	Delete	3252.49

Method	RMSETot	Method	RMSETot	Method	RMSETot	Method	RMSETot	Method	RMSETot	Method	RMSETot	Method	RMSETot	Method	RMSETot
WV3	4.31	WV3	6.77	WV3	9.09	WV3	11.45	WV3	14.29	WV3	18.69	SNI	27.55	SNI	33.84
WV2	4.63	WV2	7.31	WV2	10.01	WV1	12.56	WV1	15.91	WV1	21.04	SIZ	29.85	SIZ	35.66
WV1	4.63	WV1	7.31	WV1	10.04	WV2	12.58	WV2	15.96	WV2	21.26	WV3	35.30	WV3	62.00
SNI	4.77	SNI	7.60	SNI	10.62	SNI	13.85	SNI	17.71	SNI	22.45	WV1	37.72	WV1	62.63
SIZ	5.25	SIZ	8.92	FBI	12.46	FBI	15.41	FBI	18.92	FBI	23.77	WV2	37.94	FBI	62.76
FBI	5.94	FBI	9.22	SIZ	12.57	SIZ	16.35	SIZ	20.44	SIZ	24.92	FBI	38.74	WV2	62.86
Delete	14.24	Delete	27.40	Delete	39.63	Delete	52.46	Delete	65.10	Delete	78.28	Delete	90.02	Delete	103.14

Appendix L: P-values for pairwise comparisons

MAEtot	Method	10to20	20to30	30to40	40to50	50to60	60to70	70to80
Property	FBI	1	0.659	1	1	0	0	0
	WV1	1	1	1	1	0	0	0
	WV2	1	1	1	1	0	0	0
	WV3	1	1	1	1	0	0	0
	SNI	0	0	0	0	0	0	0
	SIZ	0	0	0	0	0	0	0
Violent	FBI	0.062	0.072	0.131	0.01	0	0	0
	WV1	0.524	0.751	0.823	0.065	0.001	0	0
	WV2	0.527	0.808	0.71	0.072	0	0	0
	WV3	1	1	1	0.269	0.028	0	0
	SNI	0	0	0	0	0	0	0
	SIZ	0	0	0	0	0	0	0

MAEave	Method	10to20	20to30	30to40	40to50	50to60	60to70	70to80
Property	FBI	1	1	1	1	0.144	0	0
	WV1	1	1	1	1	0.01	0	0
	WV2	1	1	1	1	0.008	0	0
	WV3	1	1	1	1	0.086	0	0
	SNI	1	1	1	1	1	1	0.058
	SIZ	1	1	1	1	1	0.19	0
Violent	FBI	1	1	1	1	1	0.0011	0
	WV1	1	1	1	1	1	0	0
	WV2	1	1	1	1	0.802	0	0
	WV3	1	1	1	1	1	0.001	0
	SNI	1	1	1	1	1	1	1
	SIZ	1	1	1	1	1	1	1

RMSEave	Method	10to20	20to30	30to40	40to50	50to60	60to70	70to80
Property	FBI	0.852	0.412	1	1	0.417	0	0
	WV1	1	1	1	1	0.012	0	0
	WV2	1	1	1	1	0.008	0	0
	WV3	1	1	1	1	0.16	0	0
	SNI	0.178	1	1	1	0.14	0	0
	SIZ	0.154	1	1	1	0.461	0	0
Violent	FBI	1	0	1	1	1	0.026	0
	WV1	1	0.0011	1	1	0.705	0	0
	WV2	1	0.0011	1	1	0.489	0	0
	WV3	1	0.032	1	1	1	0.003	0
	SNI	1	0	1	1	1	1	1
	SIZ	1	0	1	1	1	1	0.632

Appendix M: Agency BiasAve and state BiasTot for property & violent data

Agency: Property

10%		20%		30%		40%		50%		60%		70%		80%	
Method	BiasAve	Method	BiasAve	Method	BiasAve	Method	BiasAve	Method	BiasAve	Method	BiasAve	Method	BiasAve	Method	BiasAve
Delete	-210.07	Delete	-209.51	Delete	-210.8	Delete	-210.84	Delete	-210.52	Delete	-210.18	Delete	-210.02	Delete	-209.63
SIZ	-61.39	SIZ	-59.84	SIZ	-59.72	SIZ	-60.68	SIZ	-60.74	SIZ	-60.11	SIZ	-59.26	SIZ	-59.33
SNI	-44.36	SNI	-43.57	SNI	-44.15	SNI	-45.67	SNI	-46.76	SNI	-47.07	SNI	-47.85	SNI	-50.24
FBI	-14.08	FBI	-12.77	FBI	-13.69	FBI	-14.26	FBI	-13.94	FBI	-10.27	FBI	-3.98	FBI	5.27
WV1	-0.33	WV2	1.63	WV1	0.93	WV2	-0.61	WV1	-1.59	WV2	0.17	WV2	3.67	WV2	8.17
WV2	-0.12	WV1	1.70	WV2	0.95	WV1	-0.53	WV2	-1.47	WV1	0.35	WV1	4.67	WV1	9.25
WV3	1.72	WV3	4.32	WV3	5.65	WV3	5.42	WV3	6.88	WV3	12.75	WV3	20.14	WV3	27.30

Agency: Violent

10%		20%		30%		40%		50%		60%		70%		80%	
Method	BiasAve	Method	BiasAve	Method	BiasAve	Method	BiasAve	Method	BiasAve	Method	BiasAve	Method	BiasAve	Method	BiasAve
Delete	-22.89	Delete	-22.91	Delete	-23.05	Delete	-22.98	Delete	-23.14	Delete	-23.18	Delete	-23.10	Delete	-23.07
SIZ	-5.50	SIZ	-5.61	SIZ	-5.74	SIZ	-5.83	SIZ	-6.00	SIZ	-6.12	SIZ	-6.28	SIZ	-6.37
FBI	-4.63	FBI	-4.72	FBI	-4.85	FBI	-4.83	FBI	-4.87	FBI	-5.05	FBI	-4.57	SNI	-4.72
WV2	-2.48	WV1	-2.52	WV1	-2.62	WV1	-2.71	SNI	-3.07	SNI	-3.52	SNI	-4.06	FBI	-3.64
WV1	-2.47	WV2	-2.52	WV2	-2.59	WV2	-2.69	WV2	-2.80	WV2	-3.19	WV2	-3.19	WV2	-2.88
WV3	-2.22	WV3	-2.25	SNI	-2.37	SNI	-2.62	WV1	-2.79	WV1	-3.10	WV1	-3.01	WV1	-2.61
SNI	-1.83	SNI	-2.06	WV3	-2.27	WV3	-2.26	WV3	-2.11	WV3	-2.06	WV3	-1.41	WV3	-0.44

State: Property

10%		20%		30%		40%		50%		60%		70%		80%	
Method	BiasTot	Method	BiasTot	Method	BiasTot	Method	BiasTot	Method	BiasTot	Method	BiasTot	Method	BiasTot	Method	BiasTot
Delete	-3781.29	Delete	-7333.01	Delete	-11172.26	Delete	-14758.95	Delete	-18315.63	Delete	-22068.90	Delete	-25622.26	Delete	-29348.67
SIZ	-1104.99	SIZ	-2094.25	SIZ	-3164.93	SIZ	-4247.69	SIZ	-5284.39	SIZ	-6311.95	SIZ	-7229.20	SIZ	-8305.63
SNI	-798.39	SNI	-1524.78	SNI	-2340.02	SNI	-3196.77	SNI	-4068.55	SNI	-4941.87	SNI	-5837.62	SNI	-7034.21
FBI	-253.45	FBI	-446.80	FBI	-725.54	FBI	-998.39	FBI	-1213.01	FBI	-1078.05	FBI	-485.87	FBI	737.11
WV1	-5.96	WV2	57.09	WV1	49.28	WV2	-42.89	WV1	-138.22	WV2	17.60	WV2	447.24	WV2	1143.80
WV2	-2.24	WV1	59.61	WV2	50.21	WV1	-36.90	WV2	-127.91	WV1	36.51	WV1	570.04	WV1	1295.17
WV3	30.94	WV3	151.05	WV3	299.55	WV3	379.20	WV3	598.15	WV3	1338.84	WV3	2457.09	WV3	3822.13

State: Violent

10%		20%		30%		40%		50%		60%		70%		80%	
Method	BiasTot	Method	BiasTot	Method	BiasTot	Method	BiasTot	Method	BiasTot	Method	BiasTot	Method	BiasTot	Method	BiasTot
Delete	-412.05	Delete	-824.64	Delete	-1221.42	Delete	-1631.93	Delete	-2036.35	Delete	-2456.83	Delete	-2864.72	Delete	-3252.49
SIZ	-98.92	SIZ	-202.08	SIZ	-304.18	SIZ	-414.27	SIZ	-527.91	SIZ	-648.92	SIZ	-778.73	SIZ	-898.07
FBI	-83.29	FBI	-169.98	FBI	-257.29	FBI	-343.11	FBI	-428.85	FBI	-535.33	FBI	-566.62	SNI	-665.35
WV2	-44.66	WV1	-90.77	WV1	-138.93	WV1	-192.51	SNI	-270.33	SNI	-372.92	SNI	-503.69	FBI	-513.70
WV1	-44.39	WV2	-90.70	WV2	-137.35	WV2	-191.09	WV2	-246.27	WV2	-338.55	WV2	-396.15	WV2	-405.89
WV3	-40.03	WV3	-81.06	SNI	-125.37	SNI	-185.97	WV1	-245.71	WV1	-328.94	WV1	-373.50	WV1	-368.70
SNI	-33.03	SNI	-74.32	WV3	-120.45	WV3	-160.50	WV3	-185.66	WV3	-218.51	WV3	-174.55	WV3	-62.22