Office of
RESEARCH &
STRATEGIC PLANNING

Improving State Reporting of Crime Trends: The Development and Testing of the NIBRS Data Quality and Imputation Package (NIBRS DQIP)

ASUCRP Conference
October 6, 2015

Stephen M. Haas, Ph.D., Director
Christina R. LaValle, Research Analyst
WV Office of Research and Strategic Planning

Division of
Justice &
Community
Services

Introduction

- Reporting data “as is” is often not accurate.
- IBR data is an essential data source for tracking crime trends in states without victimization surveys.
- Imputed data allows for more stable crime trends.
- WVIBR data are well suited to study data quality and imputation100% population and crime coverage with NIBRS.
- Established history of NIBRS reporting (since 2006).
- Relative stability and consistency in crime reporting.

Imputation Research in WV

- The WV Statistical Analysis Center, with funding from JRSA, conducted two research projects on data quality and imputation methods using WV incident based reporting (IBR) data
  - Resulting in the **NIBRS Data Quality and Imputation Package (DQIP)** for the states

- The first project used one year of state IBR data to develop data quality and several alternative imputation methods for partial and non-reporting agencies.

- The second project tested and validated the data quality techniques and imputation methods of NIBRS DQIP on longitudinal state IBR data.

WV NIBRS Imputation Reports

- Research reports for both projects are located on the ORSP/WVSAC web site
- http://www.djcs.wv.gov/ORSP/SAC/Documents/WV_ImputeReportJan2013_Final.pdf
- http://www.djcs.wv.gov/ORSP/SAC/Documents/WV_Impute2ReportJan2014_Final.pdf

Criminal Justice Statistical Analysis Center

June 2014

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

Testing the Validity of Demonstrated Imputation Methods on Longitudinal NIBRS Data

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

The Uniform Crime Reporting (UCR) Program and the National Incident-Based Reporting System (NIBRS) are the two major sources of crime data in the United States. The UCR is a summary reporting system while NIBRS is an incident-based reporting system which was established to modernize crime reporting. The data collected by NIBRS is much more detailed. Given that law enforcement agencies across the nation voluntarily submit data to the Federal Bureau of Investigation (FBI) using either UCR or NIBRS, the presence of irregular reporting, missing data, and noncompliance are likely to compromise data quality.[1]

For many states, crime data collected using UCR or NIBRS are used to generate state and local crime reports and statistics. These data are most often reported "as is" and are thereby assumed correct. Since victimization data are typically not collected at the state or local levels to corroborate crime reports, there is an increased need for crime data to be as reliable as possible. Given the voluntary nature and inherent limitations of crime data collection systems, however, these data come with the caveat of being incomplete, or dubbed non-representative.

Previous research on state incident-based reporting (IBR) data revealed issues with completeness, resulting from partial and non-reporting agencies, and accuracy, due to irregular reporting (LaValle, Haas, Turley, & Nolan, 2013). The previous work found that imputation, particularly alternative imputation methods developed by the West Virginia Statistical Analysis Center (WVSAC), can be used to reliably estimate for missing data. In conclusion of applying imputation methods to IBR data, the study also revealed that reporting data "as is" may not be the most accurate representation of IBR data. Additional studies have been conducted on national UCR data which found similar concerns with data quality, particularly issues related to missing data (see Maltz, Roberts, & Stasny, 2006; Targonski, 2011).

Tools to detect and adjust for issues that are known to exist in crime data can improve data that are used as a basis for information and research. This research seeks to test and validate data quality techniques and imputation methods which will provide evidence that reliable and stable estimates of crime data can be attained with consistency over time. The study examines the performance of alternative imputation methods in comparison to FBI methods and provides a framework for the use of techniques on state-level IBR data. We apply

Office of RESEARCH & STRATEGIC PLANNING

JRSA NIBRS Imputation Webinar

- Step-by-step procedures using the data and tool are described on JRSA's website:
  - http://www.jrsa.org/webinars/index.html#wv

- 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.*

- LaValle, C. R., Haas, S. M, and Nolan, J. J. (June, 2014). *Testing the Validity of Demonstrated Imputation Methods on Longitudinal NIBRS Data.*

Imputation Background

- FBI methods have been used since 1958.

- Recently, researchers have taken on the task of improving imputation methods
  - Use more advanced computational techniques
  - Account for seasonality and other data issues

- WV methods incorporate these recommendations.
  - Advanced enough to be accurate yet accessible with reasonable guidance
  - Use seasonality and variation in reporting patterns between agencies

Imputation Background (cont.)

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

Our Approach

- Designed to assist states in being more timely and accurate in reporting their state crime trends
- Accessibility is important – want the approach to be easy for state programs to administer and report
- NIBRS data is often criticized for being incomplete or non-representative
- We sought to develop an "easy-to-use" tool that can be employed by state repository personnel and data analysts who work with state-level IRB data

Our Approach (cont.)

- 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.
- Different imputation methods were tested using simulation based on the missing value patterns 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.

Process of Applying Imputation Methods

1. Obtain crime data (ORI, agency name, crime counts).
2. Identify missing data (zero reports), manually inspect, and classify as missing data or true zeros.
3. Identify irregular data, manually inspect, and classify as outlier or acceptable.
4. Identify non-reporting agencies.
5. Obtain population estimates for municipal police departments and county sheriff departments.
6. Obtain MSA status for all non-municipal agencies.
7. Apply imputation methods.
8. Calculate statistics.

Imputation Considerations

- Imputation methods systematically estimate missing data values.

- There were two types of missing data situations:
  - Estimating for agencies that reported some but not all months of data (i.e., partial reporting agencies) and,
  - Estimating for agencies that did not report any data (i.e., non-reporting agencies).

- Several different imputation methods for missing data for both partial reporting (4) and non-reporting (5) situations were investigated and compared to the FBI method.

Focus of Imputation Studies in WV

- Two studies were conducted:
- **Study #1:** The goal was to explore alternative approaches to imputation developed by the WV SAC and compare the various methods to those traditionally used by the FBI.
- **Study #2:** Take the promising alternative approaches found in study #1and test and validate them with longitudinal data
  - Provide evidence that reliable and stable estimates of crime data can be attained with consistency over time.

Study #1: Design and Methods

- Two simulations were conducted: One for partial reporting agencies and another for 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).
- WV IBRS data is ideal for studying imputation methods for non-reporting agencies because of its 100% population and crime coverage.
- Observed and imputed values can be compared and accuracy assessed, and have been used in a variety of context (see Tremblay, 1994; Targonski, 2011; and Engels & Diehr, 2003)

Study #1: Simulation and Assessing Estimation Accuracy

- 2009 WV IBRS property and violent crime count data were used

- <u>To simulate partial reporting agencies</u>: Monthly data from specific agencies were selected at random to be removed. Using specific random seeds, the simulation was repeated 1000 times to balance the chance of "good" or "bad" draws.

- <u>To simulate non-reporting agencies</u>: investigated impact of imputation when entire agency data was deleted at10%, 20%, 30%, 40%, 50%, 60%, 70%, and 80%

Accuracy Measures for Implementation

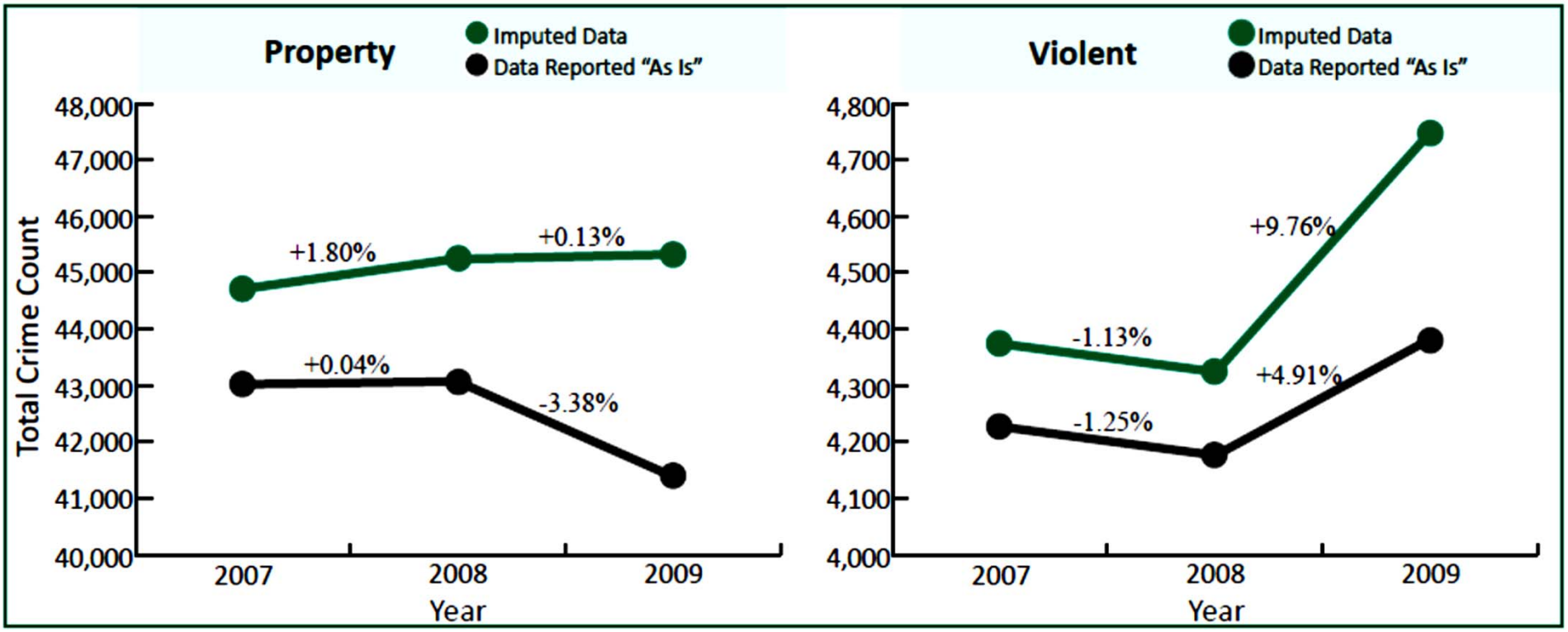
- **Mean Absolute Error (MAE):** the average of the absolute distance between imputed and original values. It describes how much, on average, imputed values differ from original values. Smaller values are better.

- **Root Mean Squared Error (RMSE):** 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:** 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.

Study #1: Impact of Imputation Methods

- Two alternative imputation methods were selected based on performance (MAE, RMSE, and Bias) and applied to 2007-2009 property and violent crime trends.



Study #1: Impact of Imputation Methods (cont.)

- As expected, the number of crimes for the imputed data were greater than the unimputed or "as is" data

- For both property and violent crime trends, there is a marked difference between 2008 and 2009 (likely related to the number of agencies reporting missing data).
  - 2007 (agency missing data) = 67 property; 66 violent
  - 2008 = 68 property; 65 violent
  - 2009 = 81 property; 79 violent

- 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.

Study #2: Testing the Validity of Demonstrated Imputation Methods

- Goals:
  - To examine the number and proportion of agencies with missing data (classifying zeros) and outliers for the purpose of validating the previously demonstrated methods.
  - To present the results of a simulation study which is used to test imputation methods on longitudinal data.

- WV IBR data 2007 to 2011; each data file contains incident data from Jan. 1 to Dec. 31 of the respective year.

- Offenses-known crime data are aggregated into monthly totals by violent and property crime and types for each reporting agency

Study #2: Validation Results

- About 1 out of 5 agencies are identified using the zero classification guidelines. After manual inspection, 1out of 10 agencies are actually classified as having missing data.

- Thus, about half of the data identified by zero classification guidelines were classified as missing.

- For property crimes, on average, about 1 out of 50 agencies were classified as having irregular data.

- For violent crimes, very few agencies were classified as having irregular reporting in violent crimes.

Study #2: Testing Imputation Methods Using Simulation

- Overall, the WV imputation methods provide strong evidence that simple modifications can be made to improve state total estimates.
- Similar results were found when estimating agency totals, allowing for improved reporting at smaller units of analysis.

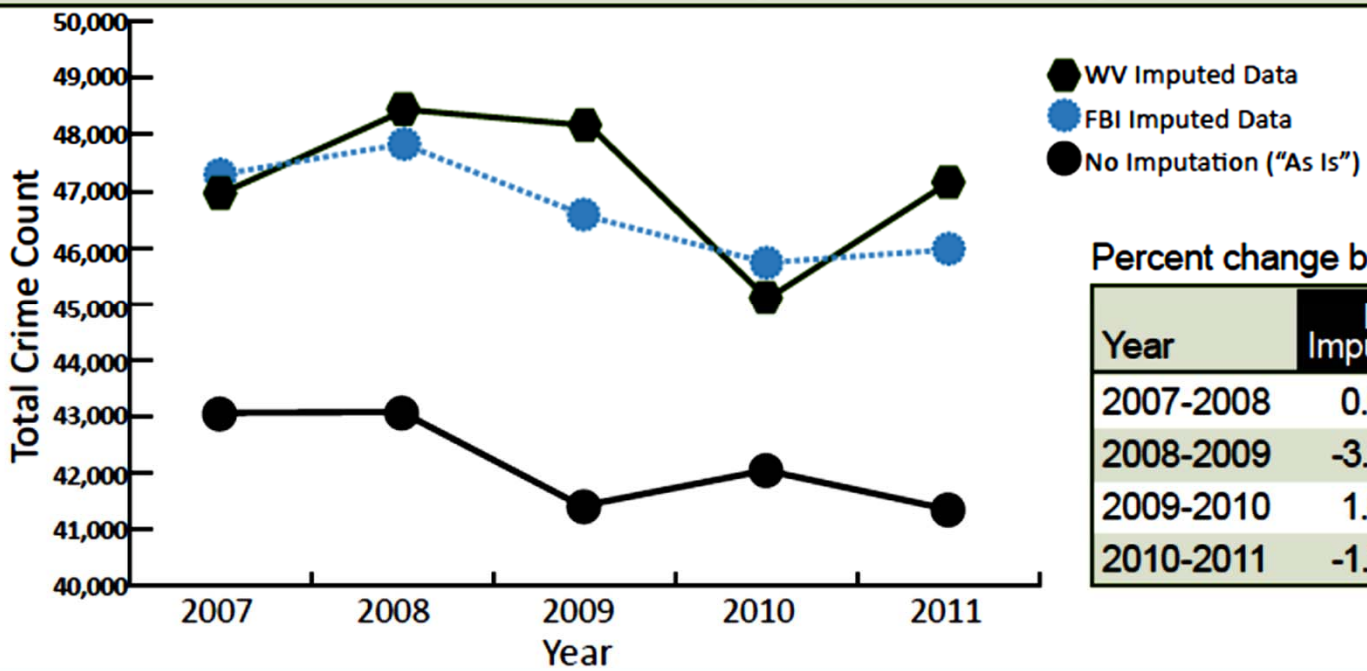
	Violent						Property					
	WV Methods			FBI Methods			WV Methods			FBI Methods		
Year	MAE_{tot}	$RMSE_{tot}$	$Bias_{tot}$	MAE_{tot}	$RMSE_{tot}$	$Bias_{tot}$	MAE_{tot}	$RMSE_{tot}$	$Bias_{tot}$	MAE_{tot}	$RMSE_{tot}$	$Bias_{tot}$
2011	**367***	**485**	**18****	416	523	-137	**2,789**	**4,080**	**34***	2,986	4,146	-624
2010	**361***	**476**	**7****	413	519	-146	**2,753**	**4,030**	**55***	2,933	4,076	-558
2009	**330**	**431**	**9****	357	449	-129	**2,634**	**3,871**	**56***	2,842	3,922	-574
2008	**370**	**485**	**9****	377	476	-139	**2,783**	**4,064**	**22***	2,997	4,119	-617
2007	**365**	**477**	**11****	378	475	-131	**2,820**	**4,124**	**26***	3,060	4,191	-646
5 yr	**1,351**	**1,788**	**47****	1,479	1,865	-536	**10,928**	**16,051**	**138***	11,826	16,295	-2,461

Results in bold indicate better performance. * Significant at 0.05 ** Significant at 0.001

Study #2: The Utility and Impact of Imputation Methods on State-wide IRB Data

- For 2007-2011 data, the imputed <u>property</u> crime trends are compared to non-imputed data (i.e., "as-is")
- # reporting agencies, stable all across 5 years, except 2009

State Property Crimes: 5 year trend



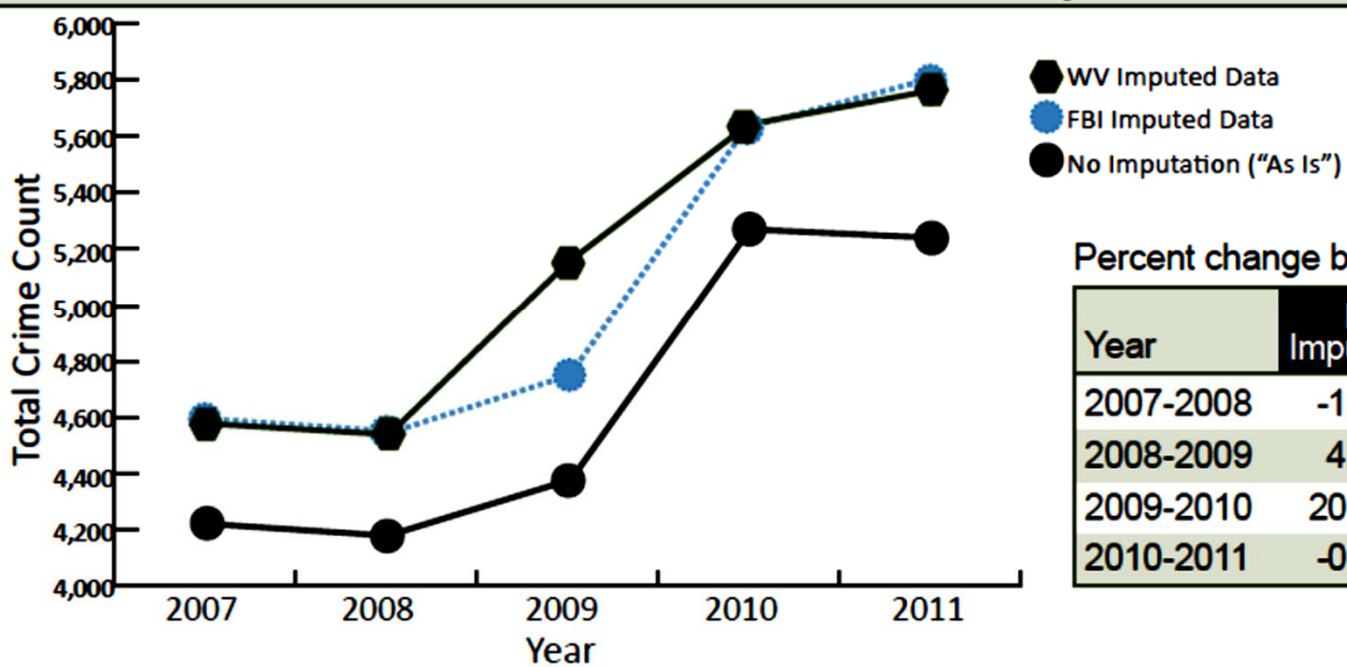
Percent change by year and imputation method.

Year	No Imputation	WV Imputation	FBI Imputation
2007-2008	0.04%	3.05%	1.25%
2008-2009	-3.83%	-0.49%	-2.64%
2009-2010	1.61%	-6.37%	-1.82%
2010-2011	-1.75%	1.90%	3.04%

Study #2: The Utility and Impact of Imputation Methods on State-wide IRB Data (Cont.)

- For 2007-2011 data, the imputed violent crime trends are compared to non-imputed data (i.e., "as-is")
- # reporting agencies, stable all across 5 years, except 2009

State Violent Crimes: 5 year trend



Percent change by year and imputation method.

Year	No Imputation	WV Imputation	FBI Imputation
2007-2008	-1.25%	-1.02%	-1.49%
2008-2009	4.91%	13.68%	4.74%
2009-2010	20.31%	9.30%	18.40%
2010-2011	-0.51%	2.06%	3.04%

Comparison of Approaches: Key Differences

- WV vs. FBI Approach:
  - For partial reporting states (i.e., 3-11 months of data), WV uses a moving average (i.e., average of a quarter) while FBI uses an average.
    - Based on assumption that crime is not evenly distributed over the year (i.e., month-to-month relationship and seasonality)
  - For non-reporting agencies (i.e., reporting 0-2 months of data), WV BINS pop. uses a scaled range fitting better WV's population characteristics
    - WV agencies fall in only 4 of the 6 FBI groups, resulting in nearly 88% of the agencies falling into one group (i.e., group 6, population less than 10,000)

Discussion and Conclusions (Cont.)

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

- Alternative imputation methods for partial and nonreporting agencies were more accurate than the methods used by the FBI.

Discussion and Conclusions (Cont.)

- Imputation can reasonably estimate for missing data, that would otherwise go undetected and uncounted, and offers a way to strengthen data quality.

- About 20% of agencies had data quality issues and remained consistent across all years.

- Based on MAE and RMSE, the research suggests that reliable state crime totals can be estimated when up to 40% of the data are missing.

Future Directions

- Our research provides evidence that alternative imputation methods improve the accuracy of crime reporting for WV, and may hold promise for other states
- Testing the imputation methods using other state or jurisdictional data.
- Investigating imputation methods that do not depend solely on population data…other community characteristics?
- Continued development of imputation methods for zero-population agencies.

Contact Information

Stephen M. Haas, Ph.D.

- Stephen.M.Haas@wv.gov
- 304.558.8814 ext. 53338

Christina LaValle, M.S.

- Chrissylavalle@yahoo.com
- 612.360.0090

Important links:

http://www.djcs.wv.gov/SAC/

http://www.facebook.com/wvorsp

http://www.twitter.com/wvorsp

Office of RESEARCH & STRATEGIC PLANNING