JBS INTERNATIONAL

JBS International, Inc.
Child Welfare Reviews Project
Contract No. HHSP233201500038C

REVIEW AND VALIDATION OF THE ANALYTIC SYNTAX USED TO PRODUCE CFSR STATEWIDE DATA INDICATORS

January 2018

Submitted to:
Linda Mitchell, CFSR Unit Leader, Children's Bureau
James Gregory, Contracting Officer's Representative, Children's Bureau

Submitted by:
JBS INTERNATIONAL, Inc.
5515 Security Lane, Suite 800
North Bethesda, MD 20852

This information shall not be disclosed outside the Government or be duplicated in whole or in part for any purpose other than as determined by the Children's Bureau.

Contents

Executive Summary i
Introduction 1
Overview of CFSR Round 3 Data Indicators 1
Background and Scope 2
Methods 5
Timeline, Workplan, and Approach 5
Data Transfer, Security, and Infrastructure 6
Initial Review of Syntax and Data Files Received 7
Reliability Checks on SPSS Syntax Files 8
Verifying Observed Performance in State Data Profiles 8
Line-Level Verification of SPSS Syntax Files 9
Regular Discussions and Meetings With CB and ODARE 10
Findings 10
Syntax for Observed Performance Yields Valid and Reliable Calculations of Statewide Data Indicators 10
Conclusion and Next Steps 11
Table A1: Data Validity Processes in Syntax or Verified by JBS—Examples 12

Executive Summary

Overview

This report presents findings from a systematic review and verification by JBS International, Inc. (JBS) of the Observed Performance syntax used to produce Child and Family Services Reviews (CFSR) statewide data indicators for the Children's Bureau (CB). This review was conducted after a review of the syntax by internal Administration on Children, Youth and Families (ACYF) staff, completed earlier in 2017. JBS researchers determined that the SPSS syntax used to develop source data files, ensure data quality, and calculate state Observed Performance can validly and reliably calculate the seven statewide data indicators according to the definitions set forth in the Federal Register Notices (FRNs) and associated CB guidance. The JBS team verified that there are no errors according to the standards provided by the CB in using the syntax to report on the indicators. The review also demonstrated the relative feasibility for states to run statewide data indicator syntax files with only a few additional steps.

Introduction

The data indicators, as communicated in Data Profiles prepared by the CB, are important for informing states and the public of statewide performance related to child welfare Safety and Permanency outcomes measured in the CFSRs. The indicators are calculated using data submitted by states in the Adoption and Foster Care Analysis and Reporting System (AFCARS) and the National Child Abuse and Neglect Data System (NCANDS). JBS was tasked with validating the SPSS Observed Performance syntax against standards established in the relevant FRNs, technical bulletins, and operational definitions provided by CB.

Methods

JBS researchers used a multi-step process to systematically review and verify the Observed Performance indicator syntax:

1. Reviewed syntax for completeness, functionality on systems external to CB
2. Performed reliability checks against Observed Performance indicators and Data Profiles
3. Performed line-level validity checks of SPSS syntax files for accurate task execution, data quality verification
    a. Reviewed the syntax and assessed its function/purpose
    b. Selected and applied the appropriate validity check(s)
    c. Ran all syntax and inspected dataset(s) and output
    d. Debriefed on similar code segments across syntax files
4. Held regular meetings for reviewing syntax, discussing findings and questions

Findings

The JBS team confirmed that the syntax files incorporate all required data elements and accurately calculate Observed Performance of the seven statewide data indicators as designed. Line-level validity checks conducted by the JBS team did not identify any errors that would cause inaccuracies in the calculation of Observed Performance of statewide data indicators as designed. Errors in data submitted by states through AFCARS and NCANDS could still affect the accuracy of the statewide data indicators as calculated using the syntax.

Next Steps

JBS will expand this systematic review and verification process to include syntax files for calculating risk-standardized performance on the statewide data indicators.

Introduction

This report presents findings from a systematic review and verification conducted by JBS International, Inc. (JBS) of the analytic syntax used to produce Child and Family Services Reviews (CFSR) statewide data indicators for the Children's Bureau (CB). The following sections provide background about the syntax for Round 3 data indicators and the scope of this review and verification, an overview of the CFSR data indicators for Round 3, a detailed description of review and verification methods, a review of findings from the review and verification, and next steps.

Overview of CFSR Round 3 Data Indicators

This section briefly describes each of the seven data indicators for Round 3 of the CFSRs, the data used to calculate the data indicators, and the syntax used to calculate Observed Performance for each data indicator.

Table 1 below lists each data indicator, the source data used to calculate each indicator, the definition according to the relevant Federal Register Notices (FRNs),[1] and any additional elements of the operational definitions provided by CB and documented in a table of decisions about each indicator.

Table 1: CFSR Indicators Overview: Statewide Data Indicators for Foster Care (FC)

Category	Title	Data Source(s)	Description
Safety 1	Maltreatment in FC	AFCARS NCANDS U.S. Census	Of all children in FC during a 12-month period, what is the rate of victimization per day of FC?
Safety 2	Recurrence of Maltreatment	NCANDS U.S. Census	Of all children who were victims of a substantiated or indicated report of maltreatment during a 12-month period, what percent were victims of another substantiated or indicated report of maltreatment within 12 months of their initial report?
Permanency 1	Permanency in 12 Months for Children Entering FC	AFCARS U.S. Census	Of all children who entered FC in a 12-month period, what percent discharged to permanency within 12 months of entering care?
Permanency 1	Permanency in 12 Months for Children in FC 12-23 Months	AFCARS	Of all children in FC on the first day of a 12-month period who had been in FC (in that episode) between 12 and 23 months, what percent discharged to permanency within 12 months of the first day of the 12-month period?
Permanency 1	Permanency in 12 Months for Children in FC 24+ Months	AFCARS	Of all children in FC on the first day of a 12-month period who had been in FC (in that episode) for 24 months or more, what percent discharged to permanency within 12 months of the first day of the 12-month period?

[1] (1) Notice of Statewide Data Indicators and National Standards for Child and Family Services Reviews. 79 Fed. Reg. 22604 (April 23, 2014). (2) Final Notice of Statewide Data Indicators and National Standards for Child and Family Services Reviews. 79 Fed. Reg. 61241 (Oct. 10, 2014). (3) Final Notice of Statewide Data Indicators and National Standards for Child and Family Services Reviews; correction. 80 Fed. Reg. 27263 (May 13, 2015).

Category	Title	Data Source(s)	Description
Permanency 2	Re-Entry to Foster Care in 12 Months	AFCARS U.S. Census	Of all children who entered FC in a 12-month period, who discharged within 12 months to reunification, live with relative, or guardianship, what percent re-entered FC within 12 months of their discharge?
Permanency 3	Placement Stability	AFCARS	Of all children who entered FC in a 12-month period, what is the rate of placement moves per day of FC?

Background and Scope

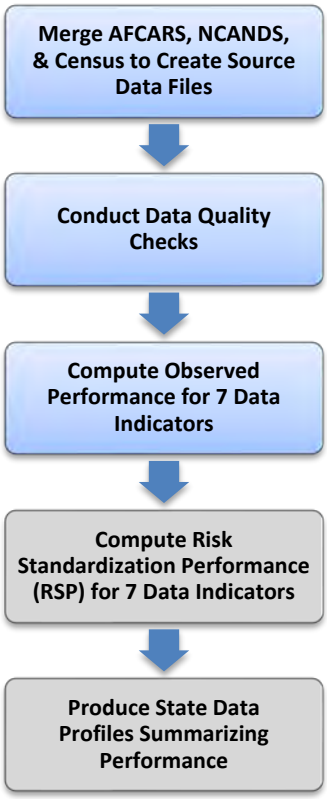
At the outset of Round 3 of the CFSRs in 2014, the CB used a state's performance on national standards developed for statewide data indicators as part of the determination of substantial conformity with child welfare outcomes measured in the CFSRs. State performance was calculated using data on the population of children served by each state's child welfare agency as submitted by states in the Adoption and Foster Care Analysis and Reporting System (AFCARS) and the National Child Abuse and Neglect Data System (NCANDS). In May 2015 (the following year), CB notified states of corrections to the calculation methodology and national standards. Since the May 2015 notice, the CB discovered additional technical errors in the syntax (i.e., statistical programming/code) and formulation of the statewide data indicators. As a result, in October 2016, CB notified states that CB would suspend the use of statewide data indicators in conformity decisions in the CFSRs during Round 3.

Nevertheless, statewide data indicators, as communicated in Data Profiles prepared by the CB and disseminated to states, are important for informing states and the public of state performance on these indicators.

With this report, CB has now completed two reviews of the syntax to identify and address syntax errors. During the past year, internal ACYF staff reviewed the syntax used to calculate Observed Performance on the statewide data indicators, identified potential changes that could screen for data quality problems in AFCARS and NCANDS submissions and preempt calculation errors, agreed on solutions to refine operational definitions of the indicators as needed, corrected the syntax, and began to issue new Data Profiles using the corrected syntax. Subsequently, in July 2017, CB asked JBS to apply our understanding of the CFSRs and analytic expertise to conduct an additional, independent review and validation of the syntax used to calculate Observed Performance on the statewide data indicators.

Figure 1 illustrates key steps in calculating the statewide data indicators. SPSS syntax extracts needed data fields from AFCARS, NCANDS, and U.S. Census databases and creates Source Data files. SPSS syntax is used to conduct a wide variety of data quality checks on the source data and to calculate Observed Performance on each statewide data indicator nationally and for each state having acceptable data quality. Stata syntax is then run to calculate Risk-Standardized Performance, in an effort—new in Round 3—to adjust Observed Performance to account for differences in populations served with respect to elements known to affect risk (e.g., child age). Tableau is used to summarize output in user-friendly formats.

Figure 1: Key Steps in Calculating Statewide Data Indicators



To complete the assessment as quickly and efficiently as possible so the syntax can be made available for state use, the scope of our analytic support was confined to validating the SPSS syntax used to extract the data, perform data quality checks, and calculate Observed Performance for states on each of the seven statewide data indicators. This report does not include review and verification of the Stata code used to calculate Risk Standardized Performance; we will undertake that step following the submission of this report.

The syntax developed to calculate state performance on the data indicators and produce the Data Profiles is quite complex, encompassing several thousand lines of code flexible enough to handle variation across individual state systems and data practices. The syntax has also been written by a number of different data experts in the CB over time. Elements of the AFCARS and NCANDS data are selected, checked for quality, and used to calculate state performance through application of multipart syntax files. **Tables 2A, B, C, and D**, below, provide an overview of the syntax files and associated processes.

Table 2A: Pull Data From Appropriate Population Database and Create Merged Source Data Files

Task/Procedure	Syntax File Type	Notes
Child Populations by State	SPSS	-
Create AFCARS Source Data	SPSS	Available source data provided in 6-month reporting periods for fiscal years 2009-2017.
Create NCANDS Source Data	SPSS	Available source data provided in 12-month reporting periods for fiscal years 2011-2016.

Table 2B: Conduct Data Quality Checks at the Child, Record, and State Levels

Task/Procedure	Syntax File Type	Notes
Create AFCARS DQ Files	SPSS	-
Create NCANDS DQ Files	SPSS	-

Table 2C: Calculate State Observed Performance for Each Data Indicator at the Child, State, and National Levels

Task/Procedure	Syntax File Type	Notes
1. Observed Performance: Maltreatment in FC 2. Observed Performance: Recurrence of Maltreatment 3. Observed Performance: Permanency a. Entries b. 12-23 Months c. 24+ Months 4. Observed Performance: Re-Entry **5.** Observed Performance: Placement Stability	SPSS	Multiple syntax files are used to calculate state Observed Performance on each data indicator.

Table 2D: Calculate State Risk-Standardized Performance, Which Adjusts for Child and State-Level Characteristics

Task/Procedure	Syntax File Type	Notes
1. Risk-Standardized Performance: Maltreatment in FC 2. Risk-Standardized Performance: Recurrence of Maltreatment 3. Risk-Standardized Performance: Permanency a. Entries b. 12-23 Months c. 24+ Months 4. Risk-Standardized Performance: Re-Entry **5.** Risk-Standardized Performance: Placement Stability	Stata	Syntax for calculating Risk-Standardized Performance was not included at this stage of the review and verification. These syntax files will be reviewed and validated following the submission of this report.

Methods

The following section describes major components of the review conducted by senior JBS researchers to assess the existing Children's Bureau syntax that calculates Observed Performance and data quality checks related to CFSR indicators. It provides a chronology of steps taken to plan and conduct a comprehensive review and verification of the syntax files against standards established in the relevant FRN, technical bulletins, and operational definitions provided by CB.

Timeline, Workplan, and Approach

On August 8, 2017, the JBS team met with representatives from CB, including CFSR Unit leadership and ODARE analysts, to formally discuss key elements of the FRN and operational definitions for the statewide data indicators, the scope of the syntax review and verification task, the timeline for completing the task, and key points of contact and communication plans across organizations. In addition to the main kickoff meeting, the JBS team held internal meetings to confirm the project scope, tasks, timeline, staffing, and deliverables. The JBS team outlined processes for systematically verifying the syntax, with a focus on confirming that the syntax files not only address all components of the CFSR Data Indicators as defined in the FRN, technical bulletins, and operational definitions provided by CB, but that data processing and management also include some applicable validity checks. JBS reviewed online sources related to the data science literature and an array of standardized, external validity checks that can be used to maintain data quality (e.g., Essnet Validat Foundation, 2016[2]), and adapted these sources to

[2] https://ec.europa.eu/eurostat/cros/system/files/methodology_for_data_validation_v1.0_rev-2016-06_final.pdf

develop a data quality checklist for this review (more details about the items on the checklist are provided below, in the *Line-Level Verification of SPSS Syntax Files* section).

The JBS analytic team also consulted as needed with project child welfare experts with extensive experience in child welfare systems and CFSRs, to better understand practical aspects of the design and application of the statewide data indicators.

Following these meetings, the JBS team submitted to CB, and received approval for, an expanded workplan specifying time frames for completing the following milestones:

- Establishing data security and transfer plans
- Setting up secure infrastructure for data processing and management
- Initial reviewing of files received and cross-walk of the FRN, technical bulletins, and operational definitions provided by CB to specific code within each statewide data indicator syntax file
- Performing reliability checks by duplicating the file directory, running all syntax files without encountering program errors, and confirming statistical output against observed performance results in CB-produced state Data Profiles
- Performing line-level review validity checks of syntax for extracting data fields from AFCARS and NCANDS and creating data source files and data quality check files, and for calculating Observed Performance for each of the seven statewide data indicators
- Regular reporting of updates, questions, and findings to CFSR Unit leadership and the ODARE team at each stage of syntax review and verification through scheduled and ad hoc phone meetings, in-person meetings, and virtual meetings using screen-share

Data Transfer, Security, and Infrastructure

JBS fully adhered to the Data Use Agreement (DUA) with CB for this task, signed on September 22, 2017, along with other applicable laws and regulations governing the use of protected and sensitive data. AFCARS and NCANDS source data files were transferred onto an encrypted flash drive, carried by hand from ODARE to JBS offices, and uploaded to JBS secure systems for data management and processing. The encrypted flash drives were then stored in a secure area. JBS systems are housed in a secure server room accessible only to authorized JBS Information Technology (IT) personnel. Only JBS researchers, authorized under the DUA, were provided remote desktop connections to access AFCARS and NCANDS data through JBS-issued, password-protected computers. All data management and processing were performed through the JBS headquarters secure local area network protected by up-to-date firewalls.

A total of 22 electronic files were included in the data transfer. The data were formatted in IBM SPSS Statistics Data Document Files (Version 23). The combined size of all NCANDS and AFCARS source data files totaled approximately two gigabytes. The source data transfer included a total of 16 AFCARS 6-month data files spanning the reporting periods 9B (April 1, 2009, to September 30, 2009) through 17A (October 1, 2016, to March 31, 2017), and a total of 6 NCANDS 12-month data files spanning fiscal years 2011 (October 1, 2010, to September 30, 2011) through 2016 (October 1, 2015, to September 30, 2016).

In addition to source data files, a total of 12 SPSS syntax files with no sensitive or protected data were transferred to JBS by email. These 12 files were the main focus of this review and

verification task. The combined files totaled roughly 6,000 lines of code and comments. The syntax file transfer included a total of 2 syntax files for merging source data by AFCARS and NCANDS, 5 syntax files for conducting data quality checks, and 5 syntax files for calculating observed performance on the statewide data indicators.

Initial Review of Syntax and Data Files Received

Once we received the files, JBS researchers reviewed the syntax for completeness and functionality. This review involved a first read-through of the syntax to gauge the length and complexity for achieving its overall purposes (e.g., datafile preparation, DQ checks, indicator calculation), to identify the input datasets and constructed (output) datasets generated, and to understand specific functions of subsections of the code (e.g., opening and combining source files, generating flag variables for data quality checks, selecting cases based on child-level and state-level data quality, computing observed performance on statewide data indicators, saving datasets into different file formats for additional processing and computation). The researchers indicated their review findings by adding notes to copies of the syntax files themselves and in review logs/diaries, then convening to discuss and cross-check their foundational understanding of the syntax.

JBS researchers then reviewed the syntax to confirm that the files appeared complete for generating all the statewide data indicators according to the FRN, technical bulletins, and operational definitions provided by CB. We confirmed with ODARE that we had received the latest versions of these reference documents before beginning the review. The researchers completed a checklist noting the location of statewide data indicator code within files, and sections of code that applied requirements according to the FRN, technical bulletins, and operational definitions provided by CB.

After conducting this check, JBS requested and received additional code, not included in the original batch of syntax, to calculate statewide data indicators for Permanency in 12 Months for Children in Foster Care 12 to 23 Months, and Permanency in 12 Months for Children in Foster Care 24+ Months.

The JBS team then verified complete data files for performing validity checks. Researchers confirmed that each source data file contained data on all 52 states/territories for all time points and that record and report dates fell within the applicable date range for the period. The JBS team also checked the AFCARS and NCANDS data structures, verifying that the files contained those data fields that were referenced in the SPSS syntax and in supporting documentation available online. Researchers identified multiple cases per child identification number, indicating multiple reports and episodes for some children. Missing information was noted across fields, and researchers verified that the information was handled appropriately using functions to define, label, and accommodate missing values.

In addition to SPSS syntax, several sections of code using Python programming language are used primarily for merging large numbers of NCANDS and AFCARS source data files. Using frequency tables and spot checks, the JBS team verified that these procedures combined all data files correctly. As a supplementary step, the JBS team consulted with internal programming experts with particular knowledge of Python to cross-check their understanding of the

procedures executed in these code segments and to verify their accuracy. The Python experts identified no problems with those parts of the syntax.

Reliability Checks on SPSS Syntax Files

Once the initial review of files was complete, the JBS team executed all SPSS syntax files to verify that the code ran without any error messages or warnings. The syntax files were run in sequential order, as instructed by ODARE (i.e., source data construction, data quality file construction, statewide data indicators). Statewide data indicator syntax files included manual input sections for users to specify the 12-month cohort(s) of interest. Any cohort within the years of available source data could be used to test the code's basic functionality. The JBS team used the cohort already specified in the versions of the syntax files as transferred from ODARE for the review and verification. The cohorts specified in the syntax received via email were the ones used to test the code's basic functionality. One minor change was made to each data file, to edit the root directory specifying the location of data files on JBS computers. The code was run in sections of approximately 5 to 15 lines to check for any errors and warning messages and to briefly review output produced by the code.

Several steps were taken to improve the systematic verification of all syntax files. For example, source data files were transferred to JBS using an identical folder structure to that used by ODARE. All syntax files used for the statewide data indicators begin by establishing a root directory with file handles to subfolders for importing and exporting files. Mirroring folder structures allowed JBS to perform validity checks while replicating the data organization method used by ODARE and minimizing necessary changes to the code to run the syntax. Additionally, all review and verification activities were performed with IBM SPSS Statistics Version 23 so that all syntax and validity checks could be run using the same statistical package and version as ODARE.

During the verification procedure, the SPSS syntax file for Placement Stability that JBS received initially produced an error message. A list of variables retained by the code omitted one variable that was referenced later in the syntax, which generated the error and terminated the syntax file. ODARE recognized this to be a versioning issue and provided JBS researchers the up-to-date, correct version of the syntax file. After receiving the correct syntax file, JBS researchers were able to run all SPSS syntax files without error codes or warnings.

Verifying Observed Performance in State Data Profiles

With all syntax files running without warning or error messages, JBS researchers ran the statewide data indicator syntax files specifying an array of 12-month cohorts in the user input sections to generate Observed Performance results for all data indicators. The researchers checked these numbers against a single state's PDF Data Profile as well as Data Profile source data provided by CB. The researchers found replication between JBS-generated observed performance results and state Data Profile results, including instances where results were not provided for a state excluded due to data quality flags. Thus, JBS researchers were able to successfully recreate the statewide data indicator results generated by the SPSS syntax files and included in state Data Profiles. This reliability check and subsequent validity checks added further verification that the NCANDS and AFCARS source data files transferred to JBS were

complete and accurate. Additionally, reliability checks completed by the JBS team demonstrated the relative feasibility of having states run statewide data indicator syntax files with only a few additional steps.

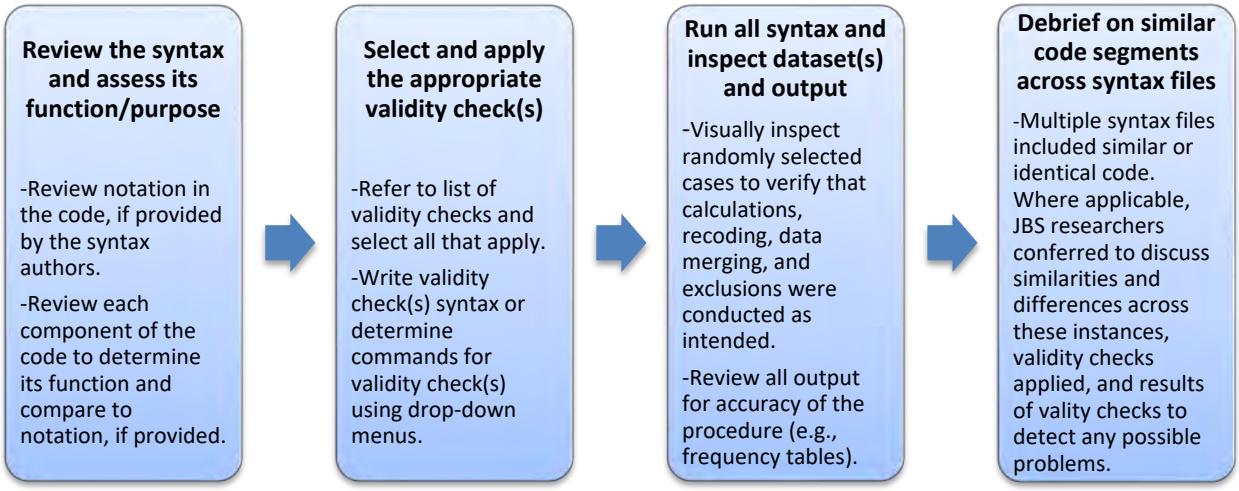
Line-Level Verification of SPSS Syntax Files

JBS researchers used the following list of validity checks as a guide to assess the types of validity checks performed in the syntax itself and for the team to conduct where applicable to verify the syntax procedures:

- Data Characters and Ranges:
  - Allowed Characters: Check for any invalid values on string or categorical variables.
  - Check Digits: Check dates and other numerical data for appropriate numerical characters.
  - Data Type: Check that formatting is appropriate for string versus numerical variables.
  - Date Format Check: Check that dates are formatted and utilized appropriately, accounting for potential challenges (e.g., leap years).
  - Upper/Lower Range Limit: Check that data do not surpass a meaningful limit.
    - Unidirectional checks (lower or upper limit only)
    - Bidirectional checks (lower and upper limits)
- Batch Totals: Check for any notable missing data patterns that may indicate incomplete data files.
- Cardinality: Check for any sign that dates or instances align chronologically.
- Consistency:
  - Within Dataset: Check that inter-related or dependent data points confirm one another.
  - Cross-System: Check that information in two separate datasets confirm one another.
- File Existence: Check that source data files were included in the data transfer and that syntax files correctly generate additional data files.
- Computations:
  - Hash Totals: Check that the intended variables are being added or subtracted.
  - Logic Check: Check for valid computations (e.g., can't divide a number by zero).
- Within-Cell Data Presence: Check for missing information within cases.
- Referential Integrity: Check for correct merging of files.
- Uniqueness: Check for duplicate cases or variables.

The JBS team performed an intensive line-level verification of roughly 6,000 lines of code to more closely ascertain that the procedures dictated by the code aligned accurately with the FRN, technical bulletins, and operational definitions provided by CB. Each researcher kept a review diary to describe the purpose and intent of each section of code; document validity checks performed on that code; and record any questions, concerns, or other important observations that emerged during the review and verification process. Generally, the JBS researchers assessed each syntax procedure separately (e.g., a "compute" or "aggregate" command) using a series of standard steps illustrated in **Figure 2**, below.

Figure 2: Standard Steps for Line-Level Review



Regular Discussions and Meetings With CB and ODARE

The JBS team met regularly with CB and ODARE to report progress and updates, discussing questions and findings at each stage of the syntax review and verification process through scheduled and ad hoc phone meetings, in-person meetings, and virtual meetings using screen-share. In particular, meetings with ODARE provided very important context and clarification of key concepts underlying the syntax, time to confirm that code was indeed conducting procedures accurately and as intended, and an opportunity to learn about the processes that have been used to validate the syntax over time. Where JBS identified any potential problems or inaccuracies, ODARE reviewed the syntax and the full group discussed all relevant considerations, implications, and potential solutions. These discussions resulted in suggestions that could be used to improve the syntax but did not yield any cases of error or inaccuracy in the statewide data indicator calculations. In addition to these specific meetings with ODARE, JBS researchers attended regular team meetings with CB, CFSR, and ODARE team members to discuss ongoing progress, reiterate goals and timelines, and provide a high-level overview of the review work and its findings to date.

Findings

This examination of those syntax files used to select and validate AFCARS and NCANDS data and calculate Observed Performance on the seven Data Indicators was conducted primarily to determine whether any elements of the syntax threaten the accuracy of the resulting state performance figures. This section describes findings from this review and verification effort.

Syntax for Observed Performance Yields Valid and Reliable Calculations of Statewide Data Indicators

The JBS team confirmed that the syntax files incorporate all required data elements and accurately calculate Observed Performance of the seven statewide data indicators in the FRN. Line-level validity checks conducted by the JBS team (see **Table A1**, below, for examples of validity checks) did not identify any errors that would cause inaccuracies in the calculation of Observed Performance of statewide data indicators per the FRN definitions.

The review and verification of the syntax found that the procedures calculate the Observed Performance for the statewide data indicators, as defined by CB and without technical errors in the syntax. JBS did not independently assess the quality of the data submitted by states through AFCARS and NCANDS. Although the syntax addresses several common data quality errors, some unforeseen errors in state-submitted data could still affect the accuracy of the statewide data indicators, as calculated using the syntax. As JBS understands, it is the responsibility of states to ensure that accurate data are submitted for use in the indicator calculations.

Conclusion and Next Steps

Using a systematic review and verification process, JBS researchers determined that the SPSS syntax used to develop source data files, ensure data quality, and calculate state Observed Performance can validly and reliably calculate the seven statewide data indicators according to the definitions set forth in the Federal Register Notices (FRN) and associated CB guidance.

The process of examining the syntax thus verified that there are no errors according to the standards provided by the CB in using the syntax to report on the indicators. The JBS team also demonstrated the relative feasibility for states to run statewide data indicator syntax files with only a few additional steps. The syntax files have evolved over multiple iterations, written by multiple programmers, over multiple years. This has produced some inefficiencies in the syntax that do not affect overall functionality. As a next step, we will expand this review and verification to include syntax for calculating risk-standardized performance on statewide data indicators.

Table A1: Data Validity Processes in Syntax or Verified by JBS—Examples

Process	Syntax File/JBS Check Example 1	Description	Syntax File/JBS Check Example 2	Description
Data Characters and Ranges				
Allowed Characters	AFCARS Source Data/ NCANDS Source Data	Verify dates for valid characters (i.e. numerical only)	AFCARS Source Data	Non-alphabetical and numerical characters included in AFCARS ID (recnum), such as $, %, and &—JBS verified that these IDs are unique
Check digits	AFCARS Source Data/ NCANDS Source Data	JBS verified that numerical variables did not contain non-numerical values	-	-
Data type	AFCARS Source Data/ NCANDS Source Data	All dates formatted in the AFCARS and NCANDS using 'adate' function in SPSS	NCANDS	ID numbers are formatted
Upper/Lower Range Limit	Permanency (AFCARS), stability	LOS calculation results in negative values; cleaned when cases where LOS < 8 days are deleted	Maltreatment in Care	Age at first day of the week results in negative ages output from the calculation—these are cleaned/deleted later in the code
Batch Totals	AFCARS Source Data	Syntax generates frequency tables to identify cases for each state across the merged datasets	NCANDS Source Data	Syntax generates frequency tables to identify cases for each state across the merged datasets
Cardinality	Permanency Indicator (entries)	SPSS syntax verifies that intake dates and discharge dates align before calculating LOS	-	-
Consistency				
Within-Dataset	Permanency Indicator (entries)	SPSS syntax conducts procedures based on the fact that date of discharge and date of prior discharge should be the same for subsequent episodes within a window of observation	Maltreatment in care	Birthdate and date of 18th birthday during the window of observations; entry and discharge flags are consistent with birthdays and age (line 363)

Process	Syntax File/JBS Check Example 1	Description	Syntax File/JBS Check Example 2	Description
Cross-System	AFCARS/NCANDS DQ files	JBS verified that DQ indicators at the state level were properly referenced and merged, and operations successful to remove states flagged for DQ removal for the specified observational period	-	-
File Existence	AFCARS Source Data/NCANDS Source Data	JBS verified that all cases are merged successfully by Python code	AFCARS/NCANDS DQ files	SPSS process files to construct NCANDS DQ checks at the record level and child level
Computations				
Hash Totals	Maltreatment in Care	Total length of stay across episodes for each Child ID (Den_Child) is calculated across entire window of observation (line 764)	-	-
Logic Check	All indicators	JBS verified that no denominator was 0 by running indicators for all observation windows valid as of March 2017 (17A); no other rate or percentage calculated noted	-	-
Within-Cell Data Presence	All indicators	If something was missing, did a computation based on the variable that had missing cases…18th birthday, if age is missing then date of 18th birthday should also be missing…if date of discharge/adjusted date of discharge is missing, LOS should be missing	-	-
Referential Integrity	All indicators	Clean state ID for merging in DQ checks and Census data correctly	Maltreatment in Care	Merging of AFCARS and NCANDS data paying attention to differences in reporting periods/source data periods
Uniqueness	AFCARS Source Data	Syntax identifies and excludes duplicate cases based on AFCARS ID	-	-

JBS International, Inc.
Child Welfare Reviews Project
Contract No. HHSP233201500038C

REVIEW AND VALIDATION OF THE ANALYTIC DO-FILES USED TO PRODUCE CFSR STATEWIDE DATA INDICATORS

ADDENDUM FOR RISK-STANDARDIZED PERFORMANCE

June 2018

Submitted to:
Linda Mitchell, CFSR Unit Leader, Children's Bureau
James Gregory, Contracting Officer's Representative, Children's Bureau

Submitted by:
JBS International, Inc.
5515 Security Lane, Suite 800
North Bethesda, MD 20852

This information shall not be disclosed outside the Government or be duplicated in whole or in part for any purpose other than as determined by the Children's Bureau.

Contents

Executive Summary ..... **i**
Overview i
Introduction i
Methods i
Findings i
Next Steps ii
Introduction **1**
Risk-Standardized Performance Indicators Background **1**
Methods **2**
Timeline, Workplan, and Approach 2
Do-File Transfer, Security, and Infrastructure 2
Initial Review of Stata Do-Files and Data Files Received 2
Reliability Checks on Stata Do-Files 3
Verifying Risk-Standardized Performance in State Data Profiles 3
Line-Level Verification of Stata Do-Files 3
Regular Discussions and Meetings With CB and ODARE 4
Findings **5**
Conclusion and Next Steps **6**

Executive Summary

Overview

This report presents findings from a systematic review and verification by JBS International, Inc. (JBS) of the Risk-Standardized Performance (RSP) Stata do-files used to produce Child and Family Services Reviews (CFSR) statewide data indicators for the Children's Bureau (CB). Using a systematic review and verification process, the JBS team verified that there are no errors according to the standards provided by the CB in the do-files calculating the RSP indicators and determining states' performance against national performance. Thus, JBS determined that the Stata do-files' calculations for the RSP indicators yielded valid and reliable results as to whether the state performed statistically better than national performance, no different than national performance, or worse than national performance, according to the definitions set forth in the related Federal Register Notices (FRN) and associated CB guidance. The review demonstrated that it was feasible for an external party to execute the statewide data indicator do-files.

Introduction

The RSP indicators, as communicated in the CB's Data Profiles, are important for informing states and the public of statewide performance related to child welfare Safety and Permanency outcomes measured in the CFSRs. The RSP indicators are calculated using data that states submitted in the Adoption and Foster Care Analysis and Reporting System (AFCARS) and the National Child Abuse and Neglect Data System (NCANDS). JBS reviewed and verified the RSP do-files against standards established in the relevant CB-provided FRN, technical bulletins, notices, and operational definitions.

Methods

JBS researchers used a multi-step process to systematically review and verify the RSP indicator do-files:

1. Reviewed do-files for completeness and functionality on systems external to CB
2. Performed reliability checks against RSP indicators and Data Profiles
3. Performed line-level validity checks of Stata do-files for accurate task execution and data quality verification
    a. Reviewed the do-files and assessed the commands' function/purpose
    b. Selected and applied validity checks as needed
    c. Ran all do-files and inspected dataset and output
    d. Debriefed on similar code segments across do-files
4. Held regular meetings with Office of Data, Analysis, Research and Evaluation (ODARE) staff for reviewing do-files, discussing findings and questions

Findings

The JBS team successfully executed the statewide data indicator do-files and confirmed that they incorporate all required data and do-files commands to accurately calculate RSP for all seven statewide data indicators in the FRN. Line-level validity checks that the JBS team conducted did not identify any errors that would cause inaccuracies in the calculation of RSP of

statewide data indicators per the CB-provided FRN, technical bulletins, and operational definitions.

Next Steps

The JBS review noted that the do-files might be improved for the future by changing how the national performance estimate is managed in the do-files to make it identical for all states when the current and national performance cohorts are the same. Further, JBS recommended amending the do-files to facilitate reviews by external parties and improve the usability of the do-files when running RSP indicators in the future. The do-files used to calculate RSP indicators in the future will be edited in response to these suggestions.

Introduction

This addendum presents findings from a systematic review and verification JBS International, Inc. (JBS) conducted of the Risk-Standardized Performance (RSP) Stata[1] do-files used to produce Child and Family Services Reviews (CFSR) statewide indicators for the Children's Bureau (CB). It supplements the systematic review and verification that JBS conducted of the SPSS syntax to produce the Observed Performance (OP) for CFSR statewide data indicators. The addendum briefly describes background, methods, results, conclusions, and next steps for the RSP component of the systematic review and verification. Readers should refer to the full report, *Review and Validation of the Analytic Syntax Used to Produce CFSR Statewide Data Indicators*, for a comprehensive explanation of the systematic review and verification of the OP indicators.

Risk-Standardized Performance Indicators Background

The RSP indicators, as communicated in the CB's Data Profiles, are important for informing states and the public of statewide performance related to child welfare Safety and Permanency outcomes measured in the CFSRs. The RSP indicators are calculated using data that states submitted in the Adoption and Foster Care Analysis and Reporting System (AFCARS) and the National Child Abuse and Neglect Data System (NCANDS). The do-files employ a multilevel modeling approach in Stata to construct a performance estimate for each state that is comparable to the national OP in the national standard years while accounting for child age, and the state's foster care entry rate for two indicators (Permanency in 12 months for children entering foster care, and Re-entry to foster care in 12 months) that could affect maltreatment and foster care outcomes. JBS reviewed and verified the RSP do-files against standards established in the relevant CB-provided Federal Register Notices (FRN), technical bulletins, notices, and operational definitions.[2]

An expert panel supported the RSP approach specifically for Round 3 of the CFSR Indicators to provide a precise measure of a state's performance against national performance. The approach was an alternative to constructing national performance using principal components analysis, which was conducted in Round 2. Expert consultants developed the RSP do-files for Round 3, wherein each indicator is calculated using a unique do-file. Administration for Children and Families (ACF) and Office of Data, Analysis, Research and Evaluation (ODARE) staff have modified the do-files over time to facilitate the process of reporting and reviewing.

---

[1] Stata is a statistical package for managing, analyzing, and graphing data, commonly used for social science research. More details can be found at https://www.stata.com/manuals/u2.pdf. Electronic files for using syntax, or code, to run Stata commands are called "do-files."

[2] (1) Notice of Statewide Data Indicators and National Standards for Child and Family Services Reviews. 79 Fed. Reg. 22604 (April 23, 2014). (2) Final Notice of Statewide Data Indicators and National Standards for Child and Family Services Reviews. 79 Fed. Reg. 61241 (Oct. 10, 2014). (3) Final Notice of Statewide Data Indicators and National Standards for Child and Family Services Reviews; correction. 80 Fed. Reg. 27263 (May 13, 2015). Information on legislation, policy, and technical bulletins related to the CFSRs is at https://www.acf.hhs.gov/cb/resource/cfsr-legislation-policy-tb.

Methods

The following section describes major components of the review senior JBS researchers conducted to assess the existing Children's Bureau do-files that calculate RSP related to CFSR indicators.

Timeline, Workplan, and Approach

During regular check-in meetings with CB representatives, including CFSR Unit leadership and ODARE analysts, JBS confirmed the scope and timeline for the RSP do-files review and verification task. The JBS team also held internal meetings to confirm the tasks, timeline, and deliverables. The task plan outlined processes for systematically verifying the do-files that were as consistent as possible with those used to review and verify the RSP do-files. A parallel objective of the task plan was to confirm that the do-files fully and accurately addressed all components of the CFSR data indicators as defined in the CB-provided FRN, technical bulletins, notices, and operational definitions. The JBS team submitted to CB, and received approval for, an expanded workplan specifying time frames for completing the following milestones:

- Conducting initial review of files and crosswalk of the CB's FRN, technical bulletins, and operational definitions within each statewide data indicator do-files file
- Requesting amended/supplementary files
- Performing reliability checks by duplicating the file directory, running all do-files without encountering program errors, and confirming statistical output against RSP results in CB-produced state Data Profiles
- Performing line-level review validity checks of do-files for calculating RSP for each of the seven statewide data indicators
- Regularly reporting updates, questions, and findings to CFSR Unit leadership and the ODARE team at each stage of do-file review and verification through scheduled and ad hoc phone meetings, email, and virtual meetings using screen-sharing

Do-File Transfer, Security, and Infrastructure

ODARE sent the JBS team emails containing a total of seven do-files. Each do-file calculated RSP for a single indicator. JBS loaded these do-files onto secure desktop computers that also held OP data files that were generated by the SPSS syntax and had been reviewed and verified previously.

Initial Review of Stata Do-Files and Data Files Received

After receiving each of the files, JBS researchers reviewed the do-files for completeness and functionality. This review involved a first read-through to identify the input datasets and constructed (output) datasets generated, and to understand specific functions of subsections of the code. The review highlighted commands intended to open and combine source files, conduct multilevel modeling, calculate RSP and 95% confidence intervals, and save values into different file formats for additional processing. The researchers recorded their findings from this preliminary review in review logs, then convened to discuss and cross-check their foundational understanding of all do-files.

JBS researchers then reviewed the do-files in more detail to confirm that the files appeared complete for generating all the statewide RSP data indicators according to the CB-provided FRN, technical bulletins, and operational definitions. Researchers completed a checklist noting the sections of code that appeared to apply all major requirements for the indicators. Requirements that were specific to the RSP indicators included risk adjustments for individual indicators and determinations of states' performance on indicators compared to the national performance.

Reliability Checks on Stata Do-Files

Once the initial review was complete, the JBS team executed all do-files to verify that the code ran without any error messages or warnings. The statewide data indicator do-files include manual input sections for users to specify the 12-month cohort(s) and state(s) of interest. The JBS team specified two to five cohorts for the reliability check, until the team was satisfied that its calculations were replicating the Data Profile results for a given indicator. By confirming that the Data Profile results matched the output generated when JBS executed the do-files, JBS researchers assessed whether this system was functioning properly. The accuracy of the indicators was verified next, during the line-level review. For this initial reliability assessment, a single state's Data Profile was used and so only that state was manually specified at this stage. More states were manually specified during the line-level verification, described below.

Verifying Risk-Standardized Performance in State Data Profiles

With all seven do-files running without warning or error messages on a single state and 12-month cohort, JBS researchers ran do-files specifying up to three additional cohorts in the user input sections to generate additional RSP results for each of the data indicators. Researchers also amended the user input to generate RSP values for multiple states within the same 12-month cohort for each indicator. The researchers checked outputted RSP numbers (e.g., the national performance, upper and lower 95% confidence interval values, and "Better Than," "No Different Than," and "Worse Than" performance status) against CB-provided Data Profile source data and PDF Data Profiles. The JBS team found replication between its results and values listed in the state Data Profiles, including instances where results were not provided for a state excluded due to data quality problems.

Thus, JBS researchers were able to successfully recreate the statewide RSP indicator results generated by the Stata do-files and included in state Data Profiles. These reliability checks demonstrated that it was feasible for an external party to execute the statewide data indicator do-files.

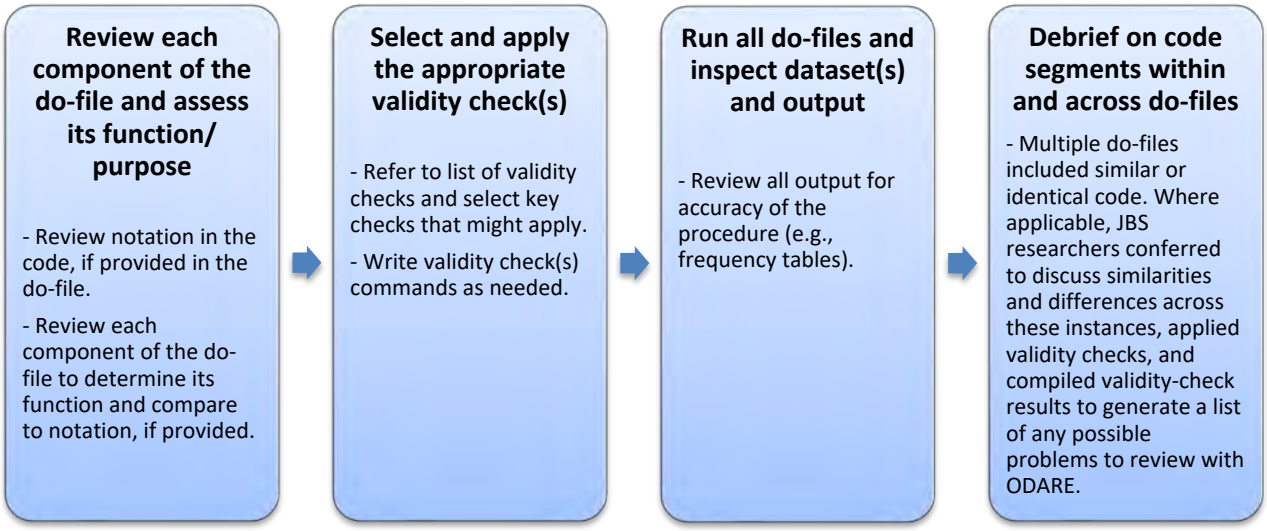
Line-Level Verification of Stata Do-Files

After the reliability of the do-files had been assessed and verified, the JBS team amended the code so that it could be run one procedure at a time to verify the accuracy of each do-files command. To complete this verification, it was necessary to amend the code in two ways.

First, researchers removed loops executing blocks of code iteratively, and then replaced commands defining and evaluating local macros as global macros. Once the do-files could be executed procedure-by-procedure, the JBS team performed a line-level verification for each do-

file in full to closely assess whether the code aligned accurately with the CB-provided FRN, technical bulletins, and operational definitions. JBS researchers conducted their own checks to verify the accuracy of the do-file commands, as needed. Line-level review and verification processes were completed for two states per each statewide data indicator RSP do-file. JBS used one state that failed the data quality (DQ) check and one state that passed it in the OP syntax because the RSP do-files have code to address the situation of a state failing the DQ, which necessitates testing the code with both a state the fails DQ and a state that passes DQ. JBS's validity checks included monitoring batch totals (or sample sizes) and data merging, reviewing for accurate computations, and assessing missing data. Each researcher kept a review diary to describe the purpose and intent of each section of code; document validity checks performed on that code; and record any questions, concerns, or other important observations that emerged during the review and verification process. JBS researchers assessed each do-file command using a series of standard steps illustrated in **Figure 1**, below.

Figure 1: Standard Steps for Line-Level Review



Regular Discussions and Meetings With CB and ODARE

Throughout the review and verification process, the JBS team met regularly (four times, in total) with ODARE to gather information, report progress and updates, and discuss questions and findings through scheduled virtual meetings using screen-share. In the first meeting, ODARE provided an overview of one of the do-files and introduced the RSP Indicator calculation procedures. In the second meeting, JBS presented its understanding of the do-file commands, and ODARE confirmed or amended this understanding as needed. In the final two meetings, JBS and ODARE reviewed the do-files to jointly confirm that they contained no errors, and discussed recommendations for making the do-files more usable for future RSP calculations. JBS team members also communicated with the ODARE team as needed via email, confirming that the adjustments JBS researchers made to the do-files to enable the line-level verification (i.e., removing loops and replacing local macros with global macros) were appropriate. In addition to specific meetings with ODARE, the JBS team attended regular team meetings with

the CFSR Unit and ODARE team members to discuss ongoing progress, reiterate goals and timelines, and provide a high-level overview of the review work and its findings to date.

Findings

JBS examined the do-files—which are used to select OP data indicators and national performance data before calculating RSP indicator values—to determine whether any elements of the do-files threaten the accuracy of the resulting RSP figures and state performance determinations included in the Data Profiles. This section describes findings from this review and verification effort.

The do-files for RSP yield valid and reliable calculations of statewide data indicators and comparison against national performance. The JBS team confirmed that the do-files incorporate all required data and commands to accurately calculate RSP for all seven statewide data indicators in the FRN. Line-level validity checks that the JBS team conducted did not identify any errors that would cause inaccuracies in the RSP calculations for statewide data indicators per the CB-provided FRN, technical bulletins, and operational definitions. The team noted one minor improvement that might be made to the do-files for future RSP calculations:

- **Differences in national performance estimate in current cohort and historical/national standard cohort**. The RSP do-files append the most recent data for the state and cohort of interest for a particular statewide data indicator to data from the fixed national performance from a specific historical/national standard cohort for that statewide data indicator. When appending the data when the current and the national performance cohorts are the same, there is a very slight difference in the values of the national performance variable starting at the 17th decimal point, when they should be identical. JBS confirmed that this floating point inconsistency in the data did not compromise the accuracy of the RSP indicators by manually adjusting the data to make the national performance values consistent across all the states, and observed that the outputted RSP results did not change.

During the review, the JBS team made some recommendations to ODARE regarding portions of the do-files that were not directly related to the RSP calculation. The recommendations concerned commands that formatted or outputted data after RSP calculations were made, or were labeled as miscellaneous or “starred-out,” so they were not executed as part of calculating the RSP indicators. ODARE concurred with these recommendations during phone meetings, and do-files used to calculate future RSP indicators will be edited in response to these suggestions. The edits will facilitate the review of these files by additional external parties, and will improve the usability of the do-files when running statewide RSP indicators in the future.

Specifically, JBS recommended that the do-files be amended to address:

- **Inconsistent labeling and statistical procedures**. For example, in output for some indicator files, states are not always consistently labeled with a specific state name. ODARE noted that they changed or added labels where it facilitated reporting and

review, but that had not been done throughout all the do-files. In the Permanency (24+ Months) do-file, the median reference category for the age variable in the multilevel model is specified differently compared to other do-files. In the Permanency (entries) do-file, reviewers noted that slightly different code was used for rounding, although the output numbers were not affected.

- **Miscellaneous or "starred-out" do-files incorrectly generating state ranks on OP and RSP**. For example, certain sections of the RSP do-files include code that is not used for official purposes, including code that assigns a numerical rank to each state in terms of its OP and RSP. Further, some starred-out code for outputting data into output files is not consistent with other do-files.

- **Limited iterations for estimation of multilevel model for RSP of Placement Stability statewide data indicator**. The RSP includes multilevel model estimation for each statewide data indicator. Whereas the default in Stata is to allow for up to 16,000 iterations in the estimation procedure, the iterate option in the Placement Stability syntax restricts iterations to 20, at which point the model outputs an error message and presents the last results from the last iteration completed. The JBS team and ODARE agreed that it would be ideal for all estimation code to be consistently specified across all syntax, and JBS recommended that the limitation on the iterations be removed from the Placement Stability syntax.

As with the OP indicator review and verification, JBS did not independently assess the quality of the data that states submitted through AFCARS and NCANDS. Errors in state-submitted data could still affect the accuracy of the OP and RSP data indicators and national performance calculated using the syntax/do-files. Further, any manual adjustments made to the do-files (e.g., removing loops to assess them line-by-line) may be more subject to human error if someone not familiar with the adjustments executes the commands/procedures.

Conclusion and Next Steps

After using a systematic review and verification process, JBS determined that there are no errors according to the CB standards in the do-files used to report on the RSP indicators. Thus, the do-files' calculations for the RSP indicators yielded valid and reliable results as to whether the state performed statistically better than national performance, no different than national performance, or worse than national performance according to the definitions set forth in the FRNs and associated CB guidance.

The review also demonstrated that it was possible for an external party to run the statewide data indicator do-files with a few additional steps.

The JBS team noted that the do-files might be improved for future use by changing how the national performance estimate is managed in the do-files to make it identical for all states when the current and national performance cohorts are the same. Further, JBS recommended amending the do-files to facilitate reviews by external parties and improve the usability of the do-files when running RSP indicators in the future. These recommendations will lead to consistent command specification and output formatting, as well as removal of all non-essential

functions. The do-files used to calculate RSP indicators in the future will be edited in response to these suggestions.