SUPPORTING STATEMENT FOR REQUEST FOR OMB APPROVAL UNDER THE PAPERWORK REDUCTION ACT

PART B –
COLLECTION OF INFORMATION EMPLOYING STATISTICAL METHODS

---

1. Description of Universe and Selection Methods Used

As described in Part A of the Supporting Statement, the data validation methodology consists of two parts:

1) **Report validation** evaluates the validity of aggregate reports submitted to ETA by 53 states or territories. It does so by checking the accuracy of the state's reporting software used to calculate the reports. The universe for report validation comprises all participant records included in the extract file. Report validation is accomplished by independently processing an entire file of participant records, providing validation counts, and comparing the validation counts to those reported by the state or grantee.

2) **Data element validation** assesses the accuracy of participant data records. For Workforce Investment Act (WIA) Title IB, the universe for data element validation comprises records of participants who exited between April 1st of the year prior to the program year through March 31st of the current program year. For Wagner-Peyser, the universe for data element validation is comprised of participants who have been placed and retained in employment. For Trade Adjustment Assistance (TAA), the universe for data element validation comprises all Trade Act participant records submitted to ETA during the prior fiscal year. For the National Farmworker Jobs Program (NFJP), the universe for data element validation comprises all NFJP records submitted to ETA during the prior program year. Data element validation is performed by reviewing samples of participant records against source documentation to ensure compliance with federal definitions.

The data validation process results in an estimate of the error rates for each data element and each reported count. Error rates are estimated separately for each state or grantee for the WIA, TAA, and NFJP programs. Error rates cannot be estimated for Wagner-Peyser because statistically valid samples are not used.

The methodology for data element validation employs sampling to improve the efficiency of the validation process. To minimize states' and grantees' burden in performing validation consistent with producing a reliable estimate of the error rates, the data element validation process is designed to compute a reliable error rate using the smallest possible sample size. To accomplish these objectives, three sampling techniques are used:

- Variable sampling rates among states are used to reduce the burden on small states and grantees as much as possible.

- Oversampling high-risk and high-importance cases is used to provide a more accurate estimate of the error rate.

- For WIA Title IB and TAA programs, multistage sampling is employed. Samples of offices are selected, and records are selected for validation only within sampled offices. This multistage design reduces the number of locations that state staff must visit to access supporting documentation.

These sampling methods balance the numbers of records and the numbers of locations so that the overall burden is reduced as much as possible, while still achieving a reliable estimate of error.

To reduce the burden on states and grantees, ETA provides validation software that calculates the validation values, imports the reported counts, draws the data element validation samples, produces online and paper validation worksheets, calculates error rates, and produces the validation reports.

Data validation relies on existing records from state and grantee management information systems and case files. Response rate issues do not arise in the data validation program.

2. Procedures for the Collection of Information

A. Statistical Methodology for Stratification and Sample Selection

- Item B.1 above indicates that report validation does not require states to use samples.

- For data element validation, multistage samples of participant records will be drawn. Independent samples are selected in each state for 7 groups – TAA, NFJP, and 5 WIA groups (dislocated workers, NEGs, adults, older youth, and younger youth). For TAA and NFJP, stratification is not employed within the samples either in the selection of offices or records. Stratification would not add substantially to the accuracy of error rate estimates. For WIA, stratification of offices is employed to reduce the burden of validation. Offices are stratified based upon the distribution of records of dislocated workers, NEGs, adults, older youth, and younger youth.

- To increase the efficiency of the process, records receive a risk weight of 1, 2, or 3 based upon two factors: whether the record is a success for calculating performance (i.e., whether the adult, dislocated worker, NEG, older youth, TAA participant, or NFJP participant was employed in the first quarter after exit or the younger youth received a diploma within one quarter after exit), and the risk that the data used to calculate performance are in error. In addition, WIA records receive a density weight that equals the number of elements to be validated per record. The two weights are added to determine a composite weight. The composite weights result in oversampling records that are more likely to contain

errors and are judged to be more substantively important (i.e., an error in such a record would be more important for a state or grantee).

- For WIA title IB programs, offices are selected at the first stage for the data element validation. Each office is assigned a weight equal to the sum of the weights for the individual records associated with that office. For each state's WIA validation, offices are stratified. Up to 5 strata may be created. The strata are based upon the number of offices in the state, the weight of each office, and the distribution of records across the 5 WIA groups. Up to 15 offices may be selected per stratum, leading to a maximum of 75 offices sampled for the state.

- For TAA programs, offices are selected at the first stage for the data element validation. Each office is assigned a weight equal to the sum of weights for the individual records associated with that office. For each state's TAA validation, a sample of offices is then selected using probability proportional to size (PPS) methods. The selection probability of an office is based upon the total number of offices in the state and the weight of that office. Samples of records are then drawn from within the selected offices. Each record has a probability of selection proportional to its weight. The expected number of offices selected in a state is based on the total number of offices shown in Table A.

- For NFJP, no offices are sampled. Records are sampled directly. Each record has a probability of selection proportional to its weight.

Table A
TAA OFFICE SAMPLING

	Number of Offices in State (N)	Number of Offices Sampled (n)
	A	**B**
1	250 or more	30
2	200-249	25
3	100-199	20
4	75-99	15
5	30-74	10
6	7-29	7
7	Fewer than 7	All

To reduce the relative burden on small states and grantees, their samples are smaller. After standards are established, it is reasonable to implement a design that holds smaller states to a lower standard of precision (hence, a smaller sample) because smaller states have fewer resources to devote to validation and they have a smaller impact upon national performance. The sample sizes are set so that the samples drawn have a maximum 95 percent confidence interval of +/- 3.5 percentage points for larger states and grantees and +/- 4 percentage points for smaller states and grantees, given certain assumptions. These assumptions are that the error rate is 5.0

percent and that the use of multistage sample selection and unequal sampling rates will result in a design effect of 2.0; that is, they will increase the variance of sample estimates by a factor of 2.0 compared to the variance of simple random samples of the same size. Where a substantial proportion of the universe of records will be sampled, allowances are also made for the finite population correction. Tables B and C show the ranges of overall sample sizes for states for WIA and TAA. Table D shows the ranges of overall sample sizes for NFJP grantees.

Table B
WIA EXITER RECORD SAMPLING

	A	B	C
	# of Exiters	Half-Length of the Confidence Interval	Range of Sample Size
1	500 or greater	3.5%	187-350
2	0-499	4%	0-187

Table C
TAA EXITER RECORD SAMPLING

	A	B	C
	# of Exiters	Half-Length of the Confidence Interval	Range of Sample Size
1	300 or greater	3.5%	100-180
2	0-299	4%	0-113

Table D
NFJP EXITER RECORD SAMPLING

	A	B	C
	# of Exiters	Half-Length of the Confidence Interval	Range of Sample Size
1	300 or greater	3.5%	100-150
2	0-299	4%	0-83

B. Estimation Procedure

For report validation, the states and grantees compare their annual reported values to the validation values to determine if the error rate is within an acceptable range. In the future, ETA will set standards for acceptable error rates for report validation.

For data element validation, estimation will include computing sample weights and estimates of error rates. Validators compare the data from the samples to source documentation. Once all the data have been evaluated, error rates are calculated for each data element. These error rates are estimated using data weighted to account for differences in probability of selection. During the initial year, states will not produce estimates of the precision of estimated error rates. In later years, after standards for the states have been established, precision estimates will be used to evaluate whether the sample error rate estimates indicate that a state has failed to meet the established standards. The validation software computes the sampling errors for each state or grantee, taking into account the multistage design and the use of unequal weights.

C. Degree of Accuracy Needed for Purpose Described in the Justification

For data validation to be effective and to allow for continuous improvement, ETA is establishing acceptable levels for the accuracy of reports and data elements in phases. Error rates for report validation and data element validation will be established independently of one another based on the analysis of validation efforts currently underway. For report validation, the first three validation years focused on detecting and resolving any issues with state and grantee data and reporting systems. Error rates collected in these years will be analyzed and, based on this information, standards for accuracy will be established for the Program Year (PY) 2007 report validation. The implementation of a set of common performance measures has delayed the establishment of standards for data element validation until states have had at least two years to validate the same data elements. Once accuracy standards are established, states and grantees will be held accountable for meeting those standards and will be required to address any issues concerning data accuracy.

D. Unusual Problems Requiring Specialized Sampling Procedures

The discussion above indicates that the methodology uses specialized sampling procedures. Strictly speaking, none of these are required. By using sampling techniques, however, the burden that data element validation imposes upon the states and grantees is significantly reduced.

3. Response Rates

As mentioned in Part 1, response rate issues do not arise in the data validation program. Data validation relies on existing records from state and grantee management information systems and case files. Through the use of valid sampling techniques, the validation

process results in estimates of data accuracy that can be generalized to the universe of data reported to ETA on program performance and activities.

4. Tests of Procedures or Methods

WIA Title IB, Wagner-Peyser, and TAA program staff have been conducting data validation for three years; the NFJP has been conducting validation for two years. The states and grantees received training prior to beginning validation and receive ongoing training and technical assistance from ETA's data validation contractor throughout the validation process. Results of these data validation activities indicate that the methodology has functioned as intended and has enabled states to identify and address reporting errors.

5. Individuals Consulted on Statistical Aspects of the Design

William S. Borden Senior Fellow Mathematica Policy Research, Inc. (609) 275-2321	Donsig Jang Senior Statistician Mathematica Policy Research, Inc. (202) 484-4246
John Eltinge Assistant Commissioner for Survey Method Research Bureau of Labor Statistics U.S. Department of Labor (202) 691-7404	Jonathan Ladinsky Senior Program Analyst Mathematica Policy Research, Inc. (609) 275-2250
John Hall Senior Sampling Statistician Mathematica Policy Research, Inc. (609) 799-3535	