JUSTIFICATION: PART B

Collections of Information Involving Statistical Methods

Introductory Note.

The design combines four different evaluative approaches. Some involve statistical sampling. As part of the "Program Review" methodology, small Acceptance Samples (AS) are drawn of various processes' outputs to confirm that the processes' internal controls work as intended to yield accurate results. The objective of the Program Review methodology, as of the program audits upon which it is based, is to make a judgment of reasonable assurance of accuracy--not to produce a point estimate of the accuracy/inaccuracy rate.

Acceptance Sampling differs considerably in concept from the more common estimation sampling. Estimation (or enumerative) sampling seeks to infer the size or rate of occurrence of something--in this case, some measurement of an attribute such as accuracy--within a universe or population. It usually implies a null hypothesis that the population value equals or exceeds a desired value for the attribute. For example, if the standard is that a program function be at least 95% accurate, a sample would be drawn with the objective of estimating the accuracy rate (percentage) for the population and specifying the lower limit of the confidence interval that includes the universe value at the given level of probability. The probability specified is the ability to avoid rejecting the null hypothesis when the hypothesis is true (statistically, this known as making a Type I error). The assumed population value, the estimated variance, the precision desired and degree of confidence determine the sample size. Estimation samples are designed to stand alone, often forming the beginning of a process of further investigating levels or causes of errors.

The objective of Acceptance Sampling, and the related procedure discovery sampling, is to indicate, very economically, whether or not certain events (usually, errors or exceptions) occur at or below some specified frequency referred to as the "acceptable quality level" (AQL). An initial step is to examine the process and assess its risk of producing errors. An acceptable Quality Level (AQL) is set to represent the upper level of the rate of exceptions produced by the process. Sample size is determined by the size of the population being inspected; the AQL (e.g., error rate or exception rate); and the degree of confidence desired. The design of Acceptance Sample balances the risk of rejecting (failing) a process that meets the AQL (Type I error), and accepting (passing) a process that produces exceptions above the AQL (Type II error).

B-1 Describe the potential respondent universe.

Samples are drawn from universes of completed actions (e.g., new employer status determinations, field audits, and benefit charging). The potential respondent universe and size for each AS appears in Table 1 pg B-3. The range is based on data from two states, Montana and California, which contain some of the smallest and largest employer populations, and so indicate the upper and lower limits for each.

Response rates per se are not relevant, because verification merely involves retrieving information relevant to a determination from primary source records, which are maintained by the state agencies. Occasionally, however, a sampled case cannot be verified because documentation cannot be located. Under such circumstances, instructions indicate that a replacement case can be drawn. Only one such replacement can be made, and if more documents are missing, the state can not claim a reasonable assurance of accuracy, and must provide further details in the hard copy annual report.

B-2. Description of procedures for collecting information.

a. **Methodology for Acceptance Sample selection**. States are given instructions on how to assemble the "transactions" (universe) files for each AS. If the sampling is to occur in an automated environment, the state has options for proceeding. It can use the COBOL program provided as part of the TPS software, or the state can select the sample in the same way using the

state's application software or a commercial statistical package. In both cases, the samples are drawn using a balanced systematic (interval) sampling method: the universe is arrayed according to a prescribed key (in most cases, employer account number); a sampling interval is obtained by dividing the universe by the number of cases to be selected; and a random start number is applied to pick the first case. The remaining cases are picked by applying the interval. Instructions are also provided for selecting samples manually, however, all states have an automated process.

b. **Methodology for Estimation Sample Selection.** For the Cashiering tax function, data are collected for the sole purpose of determining whether the state has met timely deposit requirement of 90% or more remittances deposited into the clearing account within three days or less of receipt. This is the only part of the TPS program which uses an estimation sample rather than an acceptance sample. This consists of computing a skip interval, k, which equals N/n, rounded to the nearest integer. The <u>first</u> selection, i, is randomly selected between 1 and k. Subsequent selections are: i + k, i + 2k, ... ,i + (n-1)k. Because the population size is unknown, the skip interval must be estimated. For example, a state estimates that the number of checks that will be received is 50,000. A sample of 500 checks will be selected, and the skip interval is computed: k=50000/500,k=100.

Because it is unlikely that the actual population is 50,000, the sample size will not be exactly 500, but will vary according to the actual size of the population. The true population size is estimated by k*n', where n' is the sample produced by the estimated skip interval k. For example, if the actual population is 52,000, the skip interval will produce a sample of 520, not the targeted 500, and k*n'=100*520 or 52,000.

Table 1: Potential Respondent Universe

Type of Completed Action	Universe Minimum	Universe Maximum	Sample Size	Exception Rate
Status - New Determinations	2092 / year	36,676 / year	60 per year	5 percent
Status - Successor Determinations	860 / year	2,384 / year	60 per year	5 percent
Status - Inactive/ Termination Determinations	1900/year	30,448 / year	60 per year	5 percent
Report Delinquency - Delinquent Accounts	1055 / quarter	26,981 / quarter	60 in one quarter	5 percent
Collections - Accounts Receivable	634 / at given point in time	55,421 / at given point in time	60 at the point in time	5 percent
Field Audit – Audits	1120/year	12,228/ year	60 per year	5 percent
Contribution Reports	23,896 / quarter	758,040/ quarter	60 in one quarter	5 percent
Billings - Contributory Employers	221 / quarter	19,710 / quarter	60 in one quarter	5 percent
Billings - Reimbursing Employers	7 / quarter	3,661 / quarter	up to 60 in one quarter	5 percent
Credits / Refunds	98 / quarter	13,906 / quarter	up to 60 in one quarter	5 percent
Benefit Charging - Statements	5,882 / quarter	282,127/ quarter	60 in one quarter	5 percent
Tax Rates – Notices	18,050 / year	530,625 / year	60 in a year	5 percent

Several states separate large remittances, for example through separate post office boxes. Different cutoff points are set for large remittances (e.g., Iowa considers checks over $100.00 as large, while Illinois sets the cut off at $2,500.00), and states must insure that the sample is representative with respect to these separately collected remittances.

The sampling instructions indicate that the sampling selection period will cover the time during which the state receives the peak mail for the quarter's reports. Whether this is a span of weeks or days, the reviewer is to identify, based on past historical data, the five peak days and to sample a total of 500 from those five days. Included is a chart that gives the critical values for various sample sizes for the percentages estimated from the samples. Unless the population estimate is grossly inaccurate, the samples fall within the range shown in the table, and the appropriate critical values are used to determine if the state has met the 90 percent standard.

Sample Is Between	Value To Pass
375 and 405	87.5
406 and 441	87.6
442 and 481	87.7
482 and 527	87.8
528 and 579	87.9
580 and 640	88.0

Value to pass (p*):

$p^* = 90 - [100 * (1.645 * \sqrt{\text{var}\ (P)/n})]$,

where:

var (P) = P * (1-P) = .9 * .1 = .09,
n = sample size, and
1.645 is the value of the standard normal deviate (z), appropriate for 95 percent of the cumulative standard normal distribution.

Analyses of data from the Benefit Accuracy Measurement (BAM) survey, which uses a similar systematic selection algorithm, indicate that the design effect (actual computed sampling error / srs sampling error) averaged 1.02 for the 52 UI agencies

conducting BAM. Documentation concerning the design effect of the BAM stratified systematic sample design was included in Part B of the OMB Justification for BAM (OMB No. 1205-0245) with clearance through August 31, 2009.

c. **Degree of accuracy needed for the purpose described in the justification.** As noted, the objective of accuracy investigations is to establish reasonable assurance of accuracy, taking into account findings of both the reviews of procedures and system controls ("Systems Review") and the AS. The meaning of "reasonable assurance" was discussed with a variety of persons. Particularly significant among them were top-level tax administrators, who were asked what level of inaccuracy in a given tax function would induce them to take corrective action. As a result of these discussions, an exception rate of 5% for all samples except remittances and accounts of active contributory employers were chosen, and 90% power was determined to be sufficient. The Department has decided to use an AQL of 5% for all functions. Samples of 60 cases, with up to 2 exceptions allowed, will be used to minimize the risks of penalizing states with acceptable systems. For the Cashiering sampling process, the following table shows the critical values for the test of the null hypothesis that the population percentage is greater than or equal to 90 percent (H_0: $P \geq .9$), with the risk of a type I error of 5 percent and the risk of a type II error of 10 percent. The results are stated as percentages.

Sample	Value To Pass	Minimum Pct. Passed
400	87.5	85.3
500	87.8	85.8
600	88.0	86.2

Value to pass (p*):

$$p^* = 90 - [100 * (1.645 * \sqrt{\text{var}\ (P)/n})],$$

where:

var (P) = P * (1-P) = .9 * .1 = .09,
n = sample size, and
1.645 is the value of the standard normal deviate (z), appropriate for 95 percent of the cumulative standard normal

distribution.

Ninety-five percent of the samples of the indicated size selected from a population in which timeliness is equal to or greater than 90 percent will be equal to or greater than the percentage in the "Value To Pass" column. These samples will pass the test.

Five percent of the samples will be below the value to pass and will fail the test, even though the actual percentage is 90 percent or greater.

Ten percent of the samples of the indicated size selected from a population in which timeliness is equal to the percentage in the "Minimum Percent Passed" column will be equal to or greater than the percentage in the "Value To Pass" column. These samples will pass the test. Ninety percent of the samples will be below the value to pass and will fail the test.

The minimum percent passed (p') is the minimum value that satisfies the condition:

$$p' + [100 * (1.282 * \sqrt{var\ (p')/n})] \geq p^*$$

where:

var (p') = p' * (1-p'),
n = sample size, and
1.282 is the value of the standard normal deviate (z), appropriate for 90 percent of the cumulative standard normal distribution.

A state is not required to obtain a sample point estimate of 90 percent to "pass" the test of whether they have met the standard. Because sample estimates are used, they are subject to sampling variance (as well as nonsampling error). The point estimates for 50 percent of the samples obtained from a process performing at the 90 percent level will be below 90 percent and will "fail" the test. In fact, setting the value to pass at 90 percent, with a sample size of 500 and type I error risk of .05, implies that the population percentage is $\geq$ 92 percent.

d. **Unusual problems requiring specialized sampling procedures.** Not applicable.

e. **Use of less frequent sampling to reduce burden.** It has been decided that AS need to be drawn annually to monitor the health of the various tax functions, since systems reviews will only be done every 4 years, unless a problem was discovered in the year before or the state introduced a system change

B-3. Methods to maximize response rates.

The acceptance samples will be drawn from existing agency records; therefore non-response is not an issue. Should documentation for an entire employer's file be missing, instructions allow for one such case to be replaced. No more than one such case can be replaced (see Section B-1).

B-4. Tests of procedures or methods to be undertaken.

Various parts of the design have been tested at least once. The systems reviews were pretested in 6 States; their comments on the workability of the design led to considerable modification of the questions. (No AS were drawn nor data results submitted to the Department during the pretest). A full-scale pilot test, including AS and computed measures, was conducted in 8 other states. This test gathered data on the results of systems reviews and AS, the degree that they confirmed one another, and the time required to program and collect the various kinds of information. The test also refined the questions further.

B-5. Names, addresses, telephone numbers of persons consulted to collecting/analyzing data for the agency.

a. Consulted on Statistical design.

Dr. Charles K. Fairchild, 5615 Nevada Ave., NW, Washington, DC 20005 (202) 244-2493

Dr. Michael Battaglia, Abt Associates, Inc., 55 Wheeler St., Cambridge, MA 02138 (617) 492-7100

Mr. Steven Marcus, Sparhawk Group, Inc., 1375 Commonwealth Ave., Suite 7, Alston, MA 02134 (617) 787-0388

Mr. Andrew Spisak, U.S. Department of Labor, Office of Workforce Security, 200 Constitution Avenue, NW, Room S-4522, Washington, DC 20210 (202) 693-3196

b. Collecting/Analyzing Data

Dr. Burman Skrable, U.S. Department of Labor, Office of Workforce Security, 200 Constitution Avenue, NW, Room S-4522, Washington, DC 20210 (202) 693-3197

Dr. Charles K. Fairchild, 5615 Nevada Ave., NW, Washington, DC 20005 (202) 244-2493

Mr. Steven Marcus, Sparhawk Group, Inc., 1375 Commonwealth Ave., Suite 7, Alston, MA 02134 (617) 787-0388