EXAMPLE OF VERIFICATION
AND ACCEPTANCE OF OPERATOR DATA -
LOW ENRICHED URANIUM FABRICATION

Battelle

Pacific Northwest Laboratories
Richland, Washington 99352

AUGUST 1974

Prepared for the U.S. Atomic Energy
Commission under Contract AT(45-1):1830

NOTICE

The report was prepared as an account of work sponsored by the United States Government. Neither the United States nor the United States Atomic Energy Commission, nor any of their employees, nor any of their contractors, subcontractors, or their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness or usefulness of any information, apparatus, product or process disclosed, or represents that its use would not infringe privately owned rights.

PACIFIC NORTHWEST LABORATORY
operated by
BATTELLE
for the
U.S. ATOMIC ENERGY COMMISSION
Under Contract AT(45-1)-1830

Printed in the United States of America
Available from
National Technical Information Service
U.S. Department of Commerce
5285 Port Royal Road
Springfield, Virginia 22151
Price: Printed Copy \$5.45; Microfiche \$1.45

EXAMPLE OF VERIFICATION AND ACCEPTANCE OF OPERATOR DATA -
LOW ENRICHED URANIUM FABRICATION

C. G. Hough

R. A. Schneider

K. B. Stewart

J. L. Jaech (1)

C. A. Bennett (2)

Nuclear Technology Department
Safeguards Systems Section

August 1974

- (1) With the Exxon Nuclear Company, Inc.
Richland, Washington
- (2) With the Battelle Human Affairs Center,
Seattle, Washington

BATTELLE
Pacific Northwest Laboratories
Richland, Washington 99352

Work Performed under the Sponsorship of the U. S.
Arms Control and Disarmament Agency and
the U. S. Atomic Energy Commission

TABLE OF CONTENTS

I.	INTRODUCTION.....	1
II.	SUMMARY AND CONCLUSIONS.....	2
III.	DISCUSSION	5
A.	<u>Inspection Planning</u>	6
	. Treatment without Practical Constraints.....	7
	. Design Basis for Sampling Plans for Remeasurement.....	8
	. Summary of the Design of the Sampling Plan for Verification.....	14
B.	<u>Planning Basis for Probability Statements</u>	15
	. Falsification by Large Defects	16
	. Falsification by Biasing.....	18
	. Gross Defective Detection in the Stratified Situation.....	19
	. Bias Detection in the Stratified Situation under Random Variation.....	20
	. The Overall Test for Falsification.....	25
	. $Q_{\max}$ with no Induced Process Variance.....	25
	. $Q_{\max}$ under Induced Process Variance.....	28
	. $Q_{\max}$ in the Context of $\hat{D} + \text{MUF}$	29
	. Properties of the MUF, $\hat{D}$ and $\hat{D} + \text{MUF}$ Statistics	30
	. $Q_{\max}$ in a Different Inspector Testing Situation	39
C.	<u>Plans for Evaluation of Inspection Data</u>	42
	. Verification of Item Identification.....	44
	. Attributes Inspection	44
	. Variables Inspection	48
	. Random Error Variance	48
	. Unaccounted for Bias	51
	. Supplementary Statistical Tests.....	56
	. Evaluation of MUF	58
	. Operator's MUF (MUF Approach).....	60
	. Rules for Calculating the Variance of MUF....	60
	. MUF Adjusted for Bias ($\hat{D} + \text{MUF}$).....	63

D.	<u>Application to a Uranium Fabrication Plant.....</u>	65
. .	Introduction.....	65
. .	Facility Definition	65
. .	Evaluation of Design Information.....	67
. .	Agency Measurements	74
. .	Estimation of Verification Measurement Errors	75
. .	Inspection Sample Size	80
. .	Attribute Method Sample Size	81
. .	Variable Size for Crossover Point.....	82
. .	Variable Size for Bias Detection	82
. .	Sample Size for Variance Test	86
. .	Special Case for Waste Drums.....	86
. .	Sample Sizes for Implementation	86
. .	Summary of Sample Detection Procedure.....	89
. .	Probability of Detection.....	94
. .	Overall Power of Detection.....	95
. .	Final Evaluation of Sampling Plans.....	105
. .	Basis for the Values of M used in the Examples	106
E.	Glossary of Terms.....	107

TABLES, FIGURES AND GRAPHS

<u>Table</u>	<u>Title</u>	<u>Page</u>
1	Critical Values for Directional Errors	57
2	Design Information - Expected Annual Material Balance Data	69
3	Design Information - Operator Measurement Error Data	70
4	Rules for Finding the Coefficients in Calculating the Variance Sum of Element Weights	71
5	Design Information. Stated Variance Components of MUF.....	73
6	Single Measurement Standard Deviations for the Difference Statistic	77
7	Variance Components of the Difference Statistic $\hat{D}$	78
8	Sample Sizes for Attribute Type Measurements.....	83
9	Variable Sample Size for Detection of Bias.....	87
10	Inspection Sample Sizes for Implementation.	88
11	Values of $R(\theta)$	97
12	Probability of Nondetection for Verification Parameter Values	98
13	Probability of Nondetection when Single Defects, $\hat{D}$ and MUF are treated.....	100
14	Overall Probability of Detection Using $\hat{D}$ + MUF Approach	104

Graph 1 Solution to Equation

$$Z_{1-\alpha} \sqrt{1 + R} = T - Z_{1-\beta} \sqrt{1 + FR} \dots\dots\dots 24$$

Figure 1 Flowsheet of the UO_2 Fabrication Operation. 68

EXAMPLE OF VERIFICATION AND ACCEPTANCE
OF OPERATOR DATA-LOW ENRICHED
URANIUM FABRICATION

C. G. Hough, R. A. Schneider, K. B. Stewart
J. L. Jaech and C. A. Bennett

I. INTRODUCTION

The purpose of this example is to demonstrate the application of statistical methods to the planning of inspection activities and evaluation of inspection results for a hypothetical fabrication plant. The example concentrates on the statistical aspects of test criteria,¹ sample plans, data collection, acceptance testing, MUF evaluation and statements of results and conclusions of inspections. Thus, the effects of containment and surveillance inspection activities and complexities in inspection introduced by the operator's system of scheduling of receipts and shipments, incomplete measurements and by-difference accounting are not included. The example is based on data and statistical methods given in References 2, 3, 5 and 6 and represents an attempt to adapt the methods of the references to IAEA inspection activities that might be used in connection with the NPT.

The report is structured by first presenting a description of inspection objectives. For each of the various objectives a description is given of the statistical concept or approach to be applied for each phase of the inspection activity. The conceptual basis for sampling plans, data evaluation and probability statements are described. Following the conceptual parts, the various statistical methods are applied to the inspection of the hypothetical fabrication plant. Results are given in terms of sample sizes, probabilities of detection, and confidence intervals. For the most part, sufficient explanation of the various formulas and equations is given in the text.

It should be noted that although the ideas presented are illustrated for a fabrication facility, many of them are applicable to other types of nuclear facilities as well, the primary difference being the degree of emphasis or importance attached to the various planning, inspection, and evaluation activities for each type of plant.

The report itself is presented in three main sections - Introduction, Summary and Conclusions, and Discussion.

The Summary section presents the salient features and main conclusions. The body of the text is presented in the Discussion section. Parts A, B, and C of the Discussion present the objectives and conceptual approaches chosen for application. Part D of the Discussion gives the example of application to the hypothetical fabrication plant.

II. SUMMARY AND CONCLUSIONS

The statistical methods given in References 2, 3, 5 and 6 were applied to the case of IAEA inspection of a hypothetical low-enriched uranium fabrication plant. The example case concentrates on the statistical and measurement aspects of verified materials accounting. The safeguards aspects of containment and surveillance are not included.

For the setting of the problem, it is assumed that the Agency's inspection goal is to make a quantitative statement about the plant's material balance for a single accounting period. To provide a rigorous setting for demonstration of the concepts, the plant operator is assumed to be the adversary. It is also assumed that the operator's goal is to divert M kilograms of uranium from the total material available by any combination of diversions which would appear as MUF and diversions which are disguised by data falsification and would not appear as MUF. It is further assumed that removals if made from individual items (or equivalent falsifications) could range from small amounts up to the complete removal of the contents of a container. Diversion strategies are also assumed to include the possibility of the operator inflating the variance in the measurement data to reduce the power of statistical tests to detect bias.

To combat these possible strategies, a battery of tests are developed to cover all possibilities. These include a mixed attributes-variables sampling plan, a variance test, MUF evaluation, and special adaptation of the MUF + bias statistic. These approaches are designed to provide an overall plan which will detect with high probability the absence of a given quantity of material. The plan is developed under the assumption that the operator always employs his best strategy. In this sense, the approach is conservative. Other elements of conservatism are introduced in the sampling plans and in the calculation procedures such that the resulting probabilities of detection tend to be at least as high as stated.

From the results obtained on the example case, several particular and several general conclusions can be drawn. These include:

- The effective employment of fractional sampling plans in verification requires attribute type measurement methods which have a quantitative capability as well as accurate variable measurements. In addition to conventional methods of sampling and analysis, a non-destructive instrument which can detect with high probability the absence of 10 - 30% of the material in an item is required for the fabrication plant verification.
- Because of the high precision of the variable measurement methods assumed for the example case, only a few samples are required to detect a constant bias for a given stratum. By contrast, fairly large sample sizes (approximately 100 samples) are required to protect against partial data falsifications of a size that cannot be detected by portable instruments.
- The power to detect small differences between operator and inspector measurements is limited by systematic error variances. As a result, the sample size for bias detection reaches a practical limit corresponding to a region of diminishing return.

- The inspector's power to detect loss is considerably lower ($\sim 20 - 50$ percent) than that of the "honest" operator. There are several reasons for this difference in sensitivity. Firstly, the inspector who employs a fractional sampling plan makes fewer measurements. Secondly, the difference statistic employed by the inspector contains the measurement uncertainties of both the operator and the inspector. Thirdly, and of key importance, is the fact that the diversion strategy chosen by the operator influences the inspector's power of detection. The operator can, for example, partition the diversion of a quantity M between several diversion paths. He can divert a part of M so that it would appear as MUF. He can endeavor to conceal the diversion of another part of M by altering the source data (or selective removals). In this strategy, he can make large, medium-sized, and small falsifications. Both the split of M between the paths as well as the combinations of individual falsifications affect the inspector's power of detection. Since the inspector does not know what strategy the operator will adopt, he must protect against all possibilities. Although the inspector is forced to maintain a fixed strategy, he can compute the probability of nondetection for the operator's best strategy for a given inspection plan. On this basis, the inspector can adjust his sample size to obtain the desired probability of detection for the operator's best strategy. Less than optimum choices by the operator would increase the inspector's power of detection.
- The inspector can improve his overall power of detection considerably by making special use of a composite statistic, $(\hat{D} + \text{MUF})$, in the evaluation stage. In that statistic, the observed value of MUF is added to the value of $\hat{D}$, the difference between the operator and the inspector taken over all strata for the paired

comparisons made by variable measurements. This composite statistic has two advantages for evaluative purposes over the separate use of $\hat{D}$ and MUF in the evaluation process. Firstly, the $(\hat{D} + \text{MUF})$ statistic contains, ideally, only the inspector's systematic error variance whereas the $\hat{D}$ statistic contains both the operator's and the inspector's. Secondly, the $(\hat{D} + \text{MUF})$ statistic is an unbiased estimator of the total diversions by the two paths. It contains diversions which would appear as MUF as well as diversions which would be concealed by intentional or nonintentional subtle biases. Thus, in concept, the $(\hat{D} + \text{MUF})$ statistic increases the inspector's power of detection in two ways. It has a lower variance and it contains two diversion paths - $\hat{D}$ and MUF.

The example case is given in detail in Part D of the Discussion. Samples sizes are computed for several values of M and the probabilities of detection are computed for the various ways an operator could partition his diversion. The values for M were chosen for illustrative purposes to correspond to several levels of measurement capabilities and to an assumed value for the international standard of accountability. The approach to inspection is independent of the size of M, although specific sample sizes, of course, do depend on M.

III. DISCUSSION

In the remainder of the report, the statistical approaches to be applied are discussed and illustrated from several points of view. First, they are discussed in conceptual terms as to how they are used to meet Agency inspection objectives. The conceptual portrayal includes 1) the design of sampling plans for verification measurements and records examination, 2) the design basis for probability statements, and 3) the approaches to be used in the data evaluation stage. This manner of exposition is followed by the example of application to the hypothetical fabrication plant. In the example, the concepts are applied in a concrete manner.

The material balance model for the facility is developed as well as the corresponding error propagation model. Sample sizes for inspection and probabilities of detection are calculated for various values of M , the threshold amount.

The details of these aspects of inspection planning are discussed separately in the subsequent parts of this report.

A. Inspection Planning

Problem Setting

Before proceeding further, it is useful to describe the problem and the corresponding Agency (IAEA) objectives. It is assumed that the plant operator's goal is to divert M kilograms of total uranium from the material available at the plant during the one year accounting period. The inspector, on the other hand, wishes to be able to state with a high degree of confidence that M kilograms of uranium have not been lost or diverted. The inspector's approach is to verify the source data* for the material balance by independent measurements and other activities and to draw conclusions from inspection results and the operator's MUF.

The operator is assumed to be able to obtain his goal of M kilograms by any combination of direct diversions or data falsifications to conceal diversion. To structure the problem for mathematical treatment, alterations of source data or items such that those diversions would not appear as MUF are considered data falsifications. Removals of material which would appear as MUF are considered direct diversions. Thus, the quantity of material diverted, M , is equal to the sum of falsifications plus MUF.

It is further assumed that all of the items which comprise the material balance are available for random selection and verification of contents by the inspector. In the hypothetical inspector-operator game, the operator obtains his M by altering

* Source data used in the generic sense of the source of data upon which the material balance is based.

either the contents or the source data for an item such that his options for a single item range from removal of no material to the complete removal^{*}. Similarly, for those items for which an understatement is the best strategy, it is assumed for mathematical purposes that the understatement can range from zero to 100 percent of the nominal value. The inspector on the other hand makes independent measurements and would observe, assuming for the moment no measurement errors, the exact degree of alteration on any such items included in his random sample. An item having a falsified amount value is termed a defect.

Treatment Without Practical Constraints

To illustrate the statistical principles involved, the hypothetical plant is treated as an isolated case without ties to other facilities in the fuel cycle. This approach is taken to allow application of the methods to all strata in a material balance. In actual practice, fuel cycle redundancy such as international independence on shipments to the facility may reduce the sample size for receipt verification.

For the purpose of designing sampling plans for detecting medium to gross defects certain practical constraints are not applied. These include 1) the maximum amount of material a given container could hold, 2) the minimum loss a given process could sustain, and 3) the relationship between the inherent structure of a material balance and an adversary's strategy. For all strata, the maximum gross defect is taken as the entire contents of an item or an equivalent falsification. This approach is taken to set mathematical limits for design purposes. The concept of an empty container is correct for product items and items comprising ending inventory. However, it is artificial for receipts, since no operator would want to have to account for more material than he actually has unless he were covering a discrepancy of another facility. To set boundary limits, however, input containers

* Throughout this report "removal" includes replacement of nuclear material such that the net weight of the item does not change.

are treated for gross defects as ranging from empty to twice the nominal value. Similarly, items comprising the stated waste discards are treated as being potentially overstated by the nominal amount, e.g., potentially containing no material. This treatment is used only for setting limits. For small defects, conventional treatment is used so that small biases are assumed to accumulate in a manner similar to the MUF equation but always in a direction which would conceal diversion.

Design Basis for Sampling Plans for Remeasurement

To combat the various possible strategies a diverter may adopt for individual items, the inspector employs two types of measurement methods. The first method is an instrumental method (or equivalent compositing approach) in which the inspector can quickly check individual items for medium sized to gross discrepancies with a high degree of certainty. The other method is a more accurate measurement which is capable of detecting small differences. In the remainder of the report, these two methods are referred to as the attribute method and the variable method respectively.

The sample sizes for the two measurement methods are designed as follows:

1. Attribute Measurement Sample

For each stratum, it is assumed that the operator could accumulate a quantity of material M by complete removals or falsification on a number of individual items. To combat this strategy, the inspector's sampling plan is designed to include, with a high probability, at least one such defect given that there are $M/\bar{X}$ defects in each stratum, where $\bar{X}$ is the nominal content of an item in an individual stratum. Hence, $\frac{M}{\bar{X}}$ is the minimum number of defects needed to accomplish the diversion of an amount M .

The attribute measurement method is selected on the basis that it have a quantitative range. That is, it will detect with 100 percent probability

discrepancies which range from 100 percent to some smaller percentage of the nominal value, e.g., 30 percent. In the example, a nondestructive assay instrument having a relative standard deviation of 5 percent of the nominal amount in an item is assumed for the attribute measurement. The lower range of the instrument is taken conservatively as six times the relative standard deviation or 30 percent. Thus, the attribute measurement is assumed to detect with essentially 100 percent certainty all removals (or falsifications) over 30 percent of the nominal amount of uranium in an item. The diverter's best strategy for obtaining M through large defects is either by taking the maximum he can per item, e.g., whole items, or by medium-sized removals which are below the detection range of the attribute measurement. A strategy of partial removals which can be detected by the attribute measurement is self defeating since it requires more removals to obtain a quantity M, each such removal being detectable by the attribute measurement.

2. Variable Measurement Sample Size

Three separate sample sizes are computed for the variable measurement method during the planning stage. However, only the largest of the three sample sizes is used in each stratum. Three different bases for computing sample sizes for variable measurements are needed to counter three different strategies the operator may use to attain M, the quantity of safeguards significance. These strategies are 1) medium sized removals (data falsifications) which are not readily detectable by the attribute measurement, 2) small removals from a number of items or by biasing by small amounts a larger number of items, and 3) variable biasing in order to inflate the variance of the difference statistic in order to conceal small differences. For each of these strategies, a variable sample size is designed as a counter measure. Taking

the largest of the sample sizes protects against all three strategies. The basis for each sample size design is discussed separately below.

a. Variable Sample Size for
Medium Sized Defects

To counter the strategy of medium-sized removals (or data falsifications) from an item which are just beyond the detection range of the attribute measurement, the more accurate variable measurement is employed. The variable sampling plan is designed to have a high probability of including at least one defect when the number of removals of medium size are such that the total amount removed in a stratum is M . For planning purposes, the possible number of defects in a stratum is determined from the range of the attribute measurement. In the example, the attribute measurement is assumed to be insensitive to removals which are less than 30 percent of the nominal amount of uranium in an item. Thus, the number of such removals (or data falsifications) needed to obtain an amount M for a single stratum in the example would be $M/0.3\bar{X}$.

For example, if the nominal item amount is 10 kilograms of uranium and the goal M is 100 kilograms, one would have to remove about 3 kilograms of uranium from each of 33 items to obtain 100 kilograms.

The concept of assigning a sharp cutoff point to a measurement method where the sensitivity of detection is imagined to drop instantly from 100 percent to zero is used in order to design a conservative sample size. The actual transition in sensitivity is much more gradual. However, by assuming a sharp line of demarcation in the design of the sampling plan, the inspector always takes at least enough samples to meet the inspection goal.

It should be noted, that the concept of protecting against medium-sized defects which are not readily

detectable by the attribute measurement is a vital necessity in achieving a quantitative over-all sampling plan. Sample sizes computed for bias detection alone are usually too small to provide adequate protection against such a diversion strategy.

The point where the sensitivity of the attribute measurement is assumed to fail is termed the "crossover" point, because it is at this point that the variable measurement must be employed in an attribute mode.

b. Variable Sample Size for
Small Defects (Bias)

A second variable sample size is computed in order to counter the strategy where small removals are made from a number of items or where the source data is biased by a small amount on a number of items. In this case, the diverter could obtain a quantity M from a single large stratum or he could obtain fractions of M from all or several strata.

The sample size for bias detection is based on the use of the difference statistic $\hat{D}$, where $\hat{D}$ is the overall difference between the operator and the inspector obtained by extrapolation over all strata from the individual paired comparisons, d_{ij} , on single items, e.g., $\hat{D} = \sum N_i \bar{d}_i$.

The sample size is chosen so as to detect with high probability in the presence of systematic error a total unexplainable bias or difference M over all strata. The differences from individual strata are assumed to accumulate in a direction such that the effect of deliberate bias would be to reduce the value of MUF.

The power to detect bias is directly related to the measurement uncertainty of the difference statistic. Since the measurement uncertainty is made up of both random and systematic errors, both components of error influence the power of detection. However, only the random component reduces with increasing sample size.

As a result, there is a point or region, where a further reduction in the size of the random error component has little effect on the over-all measurement uncertainty of the difference statistic. The schematic relationship

$$\sigma_{\text{total}} = \sqrt{\sigma_{\text{systematic}}^2 + \frac{\sigma_{\text{random}}^2}{n}}$$

where n is the inspector's sample size, shows in a conceptual way the manner in which the random error component decreases with increasing n and that the systematic variance component stays constant.

It is apparent from the expression that once the random variance term becomes about equal in magnitude to the systematic variance term, a region of diminishing return is reached. From this region to the region where the random variance component approaches zero, i.e., where an infinite sample size is needed; the total reduction in the over-all standard deviation is only about 40 percent. As a consequence, in this report the point of diminishing return is taken somewhat arbitrarily as the point where the standard deviation of the random error component is equal to one-half the standard deviation of the systematic error component. This point is within about 12 percent of the minimum, e.g.,

$$\sigma_{\text{total}} = \sqrt{(1)^2 + (1/2)^2} = \sqrt{1.25} = 1.12$$

Two other considerations are taken into account in designing the sample size for bias detection. First is the need to protect against the possibility that the observed random error of $\hat{D}$ will be larger than that anticipated from the design information. Second is the fact that the initial sample size for bias detection is calculated for the entire material balance using Stewart's minimum variance method. Allocation to individual strata is done as the second step in the calculation.

Regarding the first point on random error, there are several reasons why the observed random error could be larger than anticipated. Firstly, the error of either the operator or the inspector could simply be underestimated. Secondly, shifts in systematic errors could produce the same effect. Thirdly, the operator could induce a process variance in the difference statistic by varying the amount and direction of the bias in individual items. To protect against this possibility of concealment, the random error variance which is used in the calculation of the sample size is deliberately taken as being several times larger than the one based on design information. Larger variances will be detected by the test below.

c. Variable Sample Size for Variance Test

A third sample size is determined for the variable measurement in order to check the observed random error variance versus that based on design information. For the variance check, a variable sample size is computed for each stratum so as to be able to detect with high probability a difference of about 4.0 in the ratio of the true random variance to that based on design information. A chi-square test is employed using the statistic

$$\frac{(n_v - 1) S_{dr}^2}{\sigma_{Or}^2 + \sigma_{Ir}^2}$$

to test whether the true random variance is inflated. The test is designed to have about a 95 percent chance of detecting a ratio as large as:

$$\frac{\sigma_{Or, \text{ true}}^2}{\sigma_{Or, \text{ design}}^2} = 4.0 .$$

Summary of the Design of the Sampling Plan for Verification

To verify the operator's source data, sampling plans are designed for an attribute measurement and a variable measurement. The attribute sample size is chosen so as to include with high probability at least one gross defect if the operator were to obtain a quantity M by large removals (data falsifications) from a small number of items. For the variable measurement, three separate sample sizes are computed, but only the largest one is used. The first and usually the largest sample size is designed to protect against medium sized defects which are just beyond the detection range of the attribute measurement. The second sample size for variables is chosen to detect small biases. The third sample size is for a variance test. That sample size protects against encountering a greater random error in the difference statistic than anticipated from the design information. The sample size for the variance test is the same for all strata, since it is based only on the required degrees of freedom to detect a difference in ratios of variances.

The attribute sample size is also used for the records examination. For each sample selected for attribute measurement, the source data and corresponding calculational steps are checked for transcription errors, calculational errors, etc. Details are given in Part C, Plans for Evaluation of Inspection Data.

Sections B and C which follow describe the basis for the designs of probability statements and the evaluation of the data. For ease of understanding, it may be more meaningful for the reader to read first Section D (page 65) the example application before reading Sections B and C.

B. Planning Basis for Probability Statements

The objective of the inspector in the design and planning of the sampling and statistical analysis techniques is to detect with a probability of $100(1-\beta)$ percent or higher at least one of the out-of-control situations which can contribute to diversion. There are a multiplicity of techniques which can be employed by the operator in order to obtain a goal of M units of material. The material may be taken from any one of the strata which comprise the material balance. The removals may be disguised by falsification of the records. The material may also be diverted directly so as to appear as MUF. Furthermore the operator may not use just one of these techniques but may employ a combination of them to obtain M . A well-thought-out operator procedure using a combination of diversion techniques, partitioning the diversions and falsifications over the different strata and diversion modes will generally be harder for the inspector to detect than the technique of concentrating the diversion in one stratum or diversion mode alone. There are some exceptions to this last assertion as it pertains to gross defects which will be discussed later.

The null hypothesis is the controlled situation where no material has been diverted. The alternative hypothesis is that the goal quantity M is obtained by some operator technique. The alternative hypothesis includes the possibility of partitioning M among the different possible diversion paths. Restated, the inspector's objective is that the probability at least one of the paths will indicate an out-of-control situation when the alternative hypothesis is true, is $1-\beta$ or more. This probability is obtained in the following way. Let Q_i denote the probability of nondetection of diversion for the i th path. When the probabilities of detection for the different paths are independent, the probability of nondetection is $Q = Q_1 Q_2 \dots Q_k$.

If the material is not partitioned among the different modes of diversion but is concentrated in one path the probability of detecting at least one situation which is out-of-control is $1 - \beta$ or higher. The particular partitioning of M by the operator which makes the probability of nondetection a maximum is denoted as $Q_{\max}$. Against this operator procedure the probability of detection is $1 - Q_{\max}$. Generally the operator will not employ the perfectly optimum technique so that the probability of detection will be $1 - Q_{\max}$ or higher.

Falsification by Large Defects

Total or partial removal of material from an item in certain strata is a form of falsification because the material stays on the books and does not affect the MUF for the particular period. Biasing the data to decrease MUF is also, obviously, a form of falsification.

There is a range of amount values that the operator can take from an item. For example the operator may take all, 50 percent, 25 percent or so of the total amount of material. It is easily argued that the operator's objective of obtaining M units of material and minimizing the probability of detection is attained when the operator takes the material at the highest value of the range (total removal) or slightly below the crossover point (below the range of the attribute measurement). Generally the diverted material would be replaced by similar-appearing but degraded material. To be conservative the inspector uses the binomial distribution (rather than the hypergeometric) to determine sample sizes and solves the equations

$$\text{(Total Removals)} \quad \left(1 - \frac{n_a}{N}\right)^{r_1} = \beta \quad (1)$$

$$\text{(Crossover)} \quad \left(1 - \frac{n_{v1}}{N}\right)^{r_2} = \beta \quad (2)$$

for n_a and n_{v1} , where n_a and n_{v1} are the attribute and variable samples sizes and r_1 and r_2 are the number of removals of the amounts $\bar{X}$ and $\gamma\bar{X}$, ($\gamma < 1$), necessary to obtain M at these two points, i.e., $r_1\bar{X} = M$, $r_2(\gamma\bar{X}) = M$. Formula (1) implies that any particular total removal defect has a $\beta^{(1/r_1)}$ probability of not being included in the random sample with n_a items. Similarly formula (2) implies that any partial removal defect at the crossover value has a $\beta^{(1/r_2)}$ probability of not being included in the variable random sample of n_{v1} items with the variable instrument acting in the attribute mode. Now $r_1 = M/\bar{X}$, $r_2 = M/(\gamma\bar{X})$ so that a scheme to remove a total amount of material M by partitioning the goal quantity into the two amounts kM and $(1 - k)M$ from the two regions (points, strata) results in a

$$(\beta^{1/r_1})^{kM/\bar{X}} (\beta^{1/r_2})^{(1-k)M/(\gamma\bar{X})} = \beta^k \beta^{1-k} = \beta$$

probability or less of nondetection since $M/\bar{X} = r_1$, $M/(\gamma\bar{X}) = r_2$. This is a very important result for several reasons.

- a. It implies that if M is obtained by gross defects it is immaterial how the operator apportions M over the total and partial removal cases since the probability of nondetection is β or less.
- b. It implies that determining the number of samples in the strata in MUF by the binomial formula will yield a β or less probability of nondetection of at least one gross defective.
- c. A similar line of reasoning would show that if the operator decided to take only $100(1-\eta)$ percent of the amount M by total and partial removals the probability of nondetection will be $\beta^{1-\eta}$ or less.

Falsification by Biasing

Another method of falsification by the operator is to bias the results. The purpose of biasing in an ending inventory, for example, is to obtain credit for more material than is actually present. With relatively accurate measurements it would not require many inspector measurements to detect a constant bias using the statistical method of paired comparisons for analyzing results. The expression constant bias needs some explanation because bias itself means a constant deviation from a true value. Constant bias as used here means that every item is falsified by the same amount. Bias, as used here, means that the average amount value over all items varies from the true average amount value but the operator, in an attempt to camouflage the falsification, is selective as to which items are falsified, so that bias here means an average deviation over all items. The operator random variation of measurement as used in the statistics then includes this operator-induced process variation in the stated amount values. There are several constraints which are exerted on the operator's ability to induce such a process variance without being detected. In the first place the biases are constrained by outlier-type criteria on individual items in the paired-comparison analysis. In the second place the overall variance is constrained by the inspector's use of chi-square type statistical tests on whether the operator's measurement variance is consistent with design information.

There are implicit constraints which exist but which are not explicitly used in the calculations of the probability of detection of at least one out-of-control condition. Such restraints result from constraints on S/R shipments of material across international boundaries, the fact that ending inventories form the beginning inventories for the next period and there are technical considerations which delimit the stated amount of wastes which would be deemed credible.

Gross Defective Detection in the Stratified Situation

The previous sections have shown how to obtain adequate sample sizes to guard against total and partial removals and against biases. Further questions remain as the manner of adopting these techniques to the material balance and MUF situations when there are different strata within MUF.

A general proposition that has been established is that if the number of attribute samples are chosen such that the probability is $1-\beta$ of detecting at least one gross defective in a stratum when the goal quantity M implies that $M/\bar{X}$ defectives are needed by the operator in a stratum to obtain M , the probability is $1-\beta$ or higher of detecting at least one gross defective when the goal M is partitioned over different classes by the operator. This property can be used over the different classes to effect the condition that the probability is β or less of nondetection of at least one gross defective by the inspector.

The Variance Test Constraint

The inspector's test statistic

$$\frac{\sum_{i=1}^{n_v} (d_i - \bar{d})^2}{(\sigma_{Or}^2 + \sigma_{Ir}^2)}$$

is distributed as χ^2 with n_v-1 degrees of freedom if the operator's true random measurement variance agrees with the design information. If $(\sigma_{Or}^2)_T$, the true operator measurement variance, is three or four times as large as given in the design information, the discrepancy will be detected about 95 percent of the time by the aforementioned χ^2 test statistic if $\sigma_{Or}^2 = \sigma_{Ir}^2$, i.e., if the postulated operator and inspector measurement variances are the same. Considerations of this kind lead to a variable sample size of $n_{v3} = 13$ or so samples.

The Outlier Test Constraint

If Z_δ is the 100 δ percentile point of a distribution, the probability is δ^k that all k items in a random sample from the distribution will be equal to or less than Z_δ . If one sets $\delta^k = 1 - \eta$, with η suitably small, then $\delta = (1-\eta)^{1/k}$ and the probability is $1 - \eta$ that all k items in a random sample of size k will be less than or equal to $Z_\delta = Z_{(1-\eta)^{1/k}}$. Thus $\sigma_d Z_{(1-\eta)^{1/k}}$ leads to an outlier criterion for individual differences when the σ_d value is known. Considerations of this type lead to values of $3\sigma_d$ as outlier criteria for individual differences from random samples of 20 or so differences.

Bias Detection in the Stratified Situation Under Random Variation

The inspector wishes to guard against the biasing of the total goal quantity M over all the strata of the material balance. There are already certain minimum sample sizes which are required to be taken in each stratum in order for the variable instrument in the attribute mode to detect partial removals and for the variable measurements to detect understated design measurement variances.

The bias variance in $\hat{D}$ or MUF limits the reduction which can be accomplished in $\sigma_{\hat{D}}$ or σ_{MUF} by increased sampling by the inspector. The inspector gets a diminishing amount of information on the variability of MUF, or the significance of $\hat{D}$ with an increasing sample size due to the variance of the systematic error. Write $\sigma_{\hat{D}}$ in the symbolic manner

$$\sigma_{\hat{D}} = \sqrt{\sigma_{\hat{D}}^2 \text{ systematic} + \sigma_{\hat{D}}^2 \text{ random}} \quad (n)$$

where some departure is made from our usual symbolism to emphasize that the systematic error component in $\hat{D}$ does not decrease with sample size but the random component

does. The following table illustrates this in a simplified situation where $\sigma_D^2 \text{ systematic} = \sigma_D^2 \text{ random}(n) = A$, when $n = n_1$.

Sample Size	σ_{Dr}	σ_{Ds}	σ_D^2	σ_D
n_1	A	A	$2.00A^2$	1.41A
$4n_1$	A/2	A	$1.25A^2$	1.12A
$16n_1$	A/4	A	1.06A	1.03A
∞	0	A	1.00A	1.00A

It is apparent that a law of diminishing returns sets in with an increasing sample size in an effort to reduce the random component. A decision is made here that σ_{Dr} will not be reduced beyond A/2 since σ_D is within 12 percent of the limiting value when $n_1 = \infty$.

σ_D will be different under H_0 and H_1 but the systematic variance component will not change. The inspector in designing the sample sizes needed decides to take enough samples in order to induce the power $1-\beta$ in the statistical test of D if this can be done without reducing the random component standard deviation of $\hat{D}$ under H_1 to less than one-half of the standard deviation of the systematic error. Thus the total n required to test for a significant bias is the n which solves

$$z_{1-\alpha} \sqrt{\sigma_{Ds}^2 + \sigma_{Dr}^2} = M - z_{1-\beta} \sqrt{\sigma_{Ds}^2 + C^2 \sigma_{Dr}^2} \quad (3)$$

or

$$1/2 \sigma_{Ds} = C \sigma_{Dr} \quad \text{where: } C = \frac{\sigma_{Dr}|_{H_1}}{\sigma_{Dr}|_{H_0}}$$

according to where (3) can be solved without making $C \sigma_{Dr} < 1/2 \sigma_{Ds}$ or not. Thus it must first be determined

whether (3) can be solved and if it can be solved whether $1/2 \hat{\sigma}_{Ds}^2 > C \hat{\sigma}_{Dr}^2$ or not. Graph 1 is useful for this when $\alpha = \beta = 0.05$. Graph 1 will show the ratio $R = \hat{\sigma}_{Dr}^2 / \hat{\sigma}_{Ds}^2$ needed for (3) to be solved, if it can be solved.

In order to minimize the variance

$$C^2 \hat{\sigma}_{Dr}^2 = C^2 \sum N_i^2 \sigma_{dr_i}^2 / n_i$$

over the i classes and subject to the constraint $n = \sum n_i$, it can be shown that the relationship

$$n_i = n \left(\frac{N_i \sigma_{dr_i}}{\sum N_i \sigma_{dr_i}} \right)$$

must hold and the resulting variance is

$$C^2 \hat{\sigma}_{Dr}^2 = \left(C \sum N_i \sigma_{dr_i} \right)^2 / n .$$

Then set

$$\hat{\sigma}_{Dr}^2 |_{H_1} = \left(C \sum N_i \sigma_{dr_i} \right)^2 / n = \lambda^2 \hat{\sigma}_{Ds}^2$$

and solve for

$$n = \left(\frac{C \sum N_i \sigma_{dr_i}}{\lambda \hat{\sigma}_{Ds}} \right)^2$$

where $\lambda^2 = 4R$ or $\lambda^2 = (1/2)^2$. $\lambda^2 = 4R$ if (3) can be solved with $\sqrt{FR} > 1/2$, and $\lambda^2 = 1/2$ if not.

The number of variable samples required in a stratum is then equal to $n_i = n N_i \sigma_{dr_i} / \sum N_i \sigma_{dr_i}$ and is designated as n_{v2} for the stratum.

Examples are as follows:

Example 1

$$\begin{aligned}\sigma_{\hat{D}s} &= 313.7 \\ C &= 2, F = C^2 = 4 \\ \alpha &= \beta = 0.05 \\ M &= 833\end{aligned}$$

Then $T = 833/313.7 = 2.655$. Since T is less than $(Z_{1-\alpha} + Z_{1-\beta}) = 3.29$ it is not possible to solve equation (3).

Then $C\sigma_{\hat{D}r}$ is set equal to $1/2 \sigma_{\hat{D}s}$ so that

$$\begin{aligned}n &= \left(\frac{C \sum N_i \sigma_{dr_i}}{\lambda \sigma_{D_s}} \right) \\ &= \left(\frac{2 (1140)}{1/2 (313.7)} \right) = 214\end{aligned}$$

Example 2

The conditions are the same as in example 1 except that $M = 1140$. Then $T = 1140/313.7 = 3.634$. From the graph, $R = (\sigma_{\hat{D}r|H_0} / \sigma_{\hat{D}s})^2 = 0.09$. Then $\sigma_{\hat{D}r|H_1} / \sigma_{\hat{D}s} = \sqrt{2^2 (.09)} = 0.6$ and

$$n = \left(\frac{2 (1140)}{0.6 (313.7)} \right) = 149$$

If $N_i \sigma_{dr_i}$ in a stratum is 200, then

$$n_{v2} = 149 \left(\frac{200}{1140} \right) = 37.$$

Graph 1

The Overall Test for Falsification

Since falsification can consist of

- a. gross removals of the maximum amount in an item
- b. partial removals of an amount just below the attribute measurement range
- c. biases

in order to obtain the goal amount M from any of these diversion paths and from any combinations of these diversion paths, the inspector wants assurance that the sampling and analysis procedures allow the detection with an acceptably high level of probability of evidence from at least one of these paths whatever procedures the operator uses.

$Q_{\max}$ with no Induced Process Variance

It has already been established that if the inspector chooses sample sizes in the attribute mode at each of the total and partial removal levels, the probability of nondetection is $\beta^{1-\eta}$ or less where $1 - \eta$ is the proportion of M chosen at these levels. If the operator does not induce a process variance, where the χ^2 test of the observed variance against design variance and the outlier test provides some assurance in these matters, the alternative hypothesis for the bias test can be written as

$$M = Z_{1-\alpha} \hat{\sigma}_D + Z_{1-\beta} \hat{\sigma}_D = \hat{\sigma}_D (Z_{1-\alpha} + Z_{1-\beta}).$$

The probability of nondetection here under the alternative hypothesis is

$$\Phi \left(\frac{Z_{1-\alpha} \hat{\sigma}_D - (Z_{1-\alpha} \hat{\sigma}_D + Z_{1-\beta} \hat{\sigma}_D)}{\hat{\sigma}_D} \right)$$

$$= \Phi(-Z_{1-\beta}) = \Phi(Z_{\beta}) = \beta$$

assuming that $\hat{D}$ is normally distributed. If only ηM of the goal quantity M is obtained by biasing, the probability of nondetection of the bias is

$$\begin{aligned} & \Phi \left(\frac{Z_{1-\alpha} \sigma_{\hat{D}} - \eta (Z_{1-\alpha} \sigma_{\hat{D}} + Z_{1-\beta} \sigma_{\hat{D}})}{\sigma_{\hat{D}}} \right) \\ &= \Phi \left(Z_{1-\alpha} - \eta (Z_{1-\alpha} + Z_{1-\beta}) \right). \end{aligned}$$

If the operator divides M into two components $(1-\eta)M$ and ηM , and diverts from the gross defects and bias paths, respectively, the probability of nondetection is

$$Q = \beta^{1-\eta} \Phi \left[(Z_{1-\alpha} - \eta (Z_{1-\alpha} + Z_{1-\beta})) \right].$$

The operator wishes to determine the values $(1-\eta)M$ and ηM which will maximize the probability of nondetection. This is determined in the following way.

$$\ln Q = (1-\eta) \ln \beta + \ln \Phi \left[Z_{1-\alpha} - \eta (Z_{1-\alpha} + Z_{1-\beta}) \right]$$

$$\frac{d \ln Q}{d \eta} = 0 = - \ln \beta - (Z_{1-\alpha} + Z_{1-\beta}) \frac{\Phi' [Z_{1-\alpha} - \eta (Z_{1-\alpha} + Z_{1-\beta})]}{\Phi [Z_{1-\alpha} - \eta (Z_{1-\alpha} + Z_{1-\beta})]}$$

$$= - \ln \beta - (Z_{1-\alpha} + Z_{1-\beta}) R \left(Z_{1-\alpha} - \eta (Z_{1-\alpha} + Z_{1-\beta}) \right)$$

where $R(x) = \phi'(x) / \phi(x)$.

Then solve $R(\theta) = -\ln \beta / (Z_{1-\alpha} + Z_{1-\beta})$

and

$$\eta = \frac{Z_{1-\alpha} - \theta}{Z_{1-\alpha} + Z_{1-\beta}}.$$

($R(\theta)$ is related to Mills' ratio by the formula $R(\theta) = (S(-\theta))^{-1}$ where $S(\theta)$ is Mills' ratio.)

Following is a simple example to show what is meant by these results; two solutions are given, one by a straightforward but trial and error approach and another by the $R(\theta)$ table method.

Example:

$$\alpha = \beta = 0.05, Z_{1-\alpha} = Z_{1-\beta} = 1.645.$$

The objective is to find out what value of η maximizes Q , and to determine what this value $Q_{\max}$ is. The step by step solution is as follows:

η	$\beta^{1-\eta}$	$\phi(Z_{1-\alpha} - \eta(Z_{1-\alpha} + Z_{1-\beta}))$	Q
0.1	0.06746	0.9057	0.0611
0.2	0.09102	0.8382	0.0763
0.3	0.12282	0.7447	0.0915
0.4	0.16572	0.6289	0.1042
0.5	0.22361	0.5000	0.1118
0.6	0.30171	0.3711	0.1120
0.7	0.40709	0.2553	0.1039
0.8	0.54928	0.1618	0.0889
0.9	0.25973	0.0943	0.0698

$Q_{\max}$ appears to lie between 0.1118 and 0.1120. Try $\eta = 0.55$. Then $\beta^{1-\eta} = 0.25973$, $\Phi = 0.4327$, $Q = 0.1124$.

The solution from the R tables is as follows.

$$R(\theta) = \frac{-\ln 0.05}{2 (1.645)} = 0.9105$$

$$\theta = -0.1719$$

$$\eta = \frac{1.645 - (-0.1719)}{3.29} = 0.5522$$

$$Q = (0.26146)(0.43176) = 0.1129$$

Two significant places in Q is probably satisfactory for all practical purposes. The extra significant places are useful primarily to show the working of the procedure.

$Q_{\max}$ Under Induced Process Variation

Sometimes it is not possible to attain a desired standard deviation of $\hat{D}$ such that $M = \sigma_{\hat{D}}(Z_{1-\alpha} + Z_{1-\beta})$ because $\sigma_{\hat{D}}$ is limited by its inherent systematic error standard deviation. In other instances the inspector must assume that there is an induced process variation in excess of the natural measurement accuracy which, though limited by inspector tests on the operator measurement variance, is still larger than would be detected most of the time by the variance tests. In this case call $\sigma_{\hat{D}|H_0}$ the standard deviation of $\hat{D}$ under the controlled and $\sigma_{\hat{D}|H_1}$ the standard deviation under the quasi-controlled situation. The probability of nondetection then may be written as

$$Q = \beta^{G/M} \Phi \left(\frac{Z_{1-\alpha} \sigma_{\hat{D}|H_0} - (M - G)}{\sigma_{\hat{D}|H_1}} \right)$$

where G is the amount obtained by gross defectives and $D = M - G$ is the amount obtained by biasing. Here the problem of maximizing Q goes through in a similar manner.

$$\begin{aligned} \frac{d \ln Q}{dG} &= \frac{\ln \beta}{M} + \frac{1}{\hat{\sigma}_D | H_1} R(\theta) = 0 \\ \theta &= R^{-1} \left(\frac{\ln \beta}{M / \hat{\sigma}_D | H_1} \right) \end{aligned} \quad (4)$$

and

$$G = M + \theta \hat{\sigma}_D | H_1 - Z_{1-\alpha} \hat{\sigma}_D | H_0. \quad (5)$$

The relationship is written with $M / \hat{\sigma}_D | H_1$ in the denominator in order to emphasize that the determining consideration is the magnitude of the M to $\hat{\sigma}_D | H_1$ ratio.

$Q_{\max}$ in the Context of $\hat{D} + \text{MUF}$

Write

$$Q = \beta^{G/M} \Phi \left(Z_{1-\alpha} - \frac{(M-G)}{\hat{\sigma}_{\hat{D}+\text{MUF}}} \right). \quad (6)$$

Q is maximized as indicated before so that

$$R(\theta_{\hat{D}+\text{MUF}}) = \frac{-\ln \beta}{(M / \hat{\sigma}_{\hat{D}+\text{MUF}})}$$

is solved, and

$$M - G = (Z_{1-\alpha} - \theta_{\hat{D}+\text{MUF}}) \hat{\sigma}_{\hat{D}+\text{MUF}}$$

and

$$G = M - (Z_{1-\alpha} - \theta_{\hat{D}+\text{MUF}}) \hat{\sigma}_{\hat{D}+\text{MUF}}.$$

The significance here is that when these values of G and M-G are substituted into (6) the operator's maximum probability of nondetection is obtained when the inspector tests for gross defectives using the attribute tester and the variable measurement in the attribute mode and uses the statistic $\hat{D}+MUF$ to test for combined falsification and diversion into MUF.

Properties of the MUF, $\hat{D}$ and $(\hat{D}+MUF)$ Statistics

MUF is the operator's statistical estimate of the amount of material missing in a MBA (facility, MBA system) based on the methods of mass balance accounting. $\hat{D}$ is an estimate of the inspector MUF bias minus the operator MUF bias. The inspector component of $\hat{D}$ is usually based on a random, fractional sample of all possible item amount measurements in MUF and is extrapolated to include all items. The creation of this statistic is possible since $\hat{D}$ is based on amount differences of individual items and the process variation in amounts is thus eliminated. $\hat{D}$ thus also checks falsification since natural and purposely-induced biases behave the same in these statistics. If the inspector bias is relatively small in comparison to the operator's naturally-occurring plus artificially-induced biases, the statistic $\hat{D}$ will estimate the negative of the inspector's bias in MUF. Thus $(\hat{D}+MUF)$ is essentially the operator's MUF with the operator's bias replaced by the inspector's bias when the fraction of inspector measurements is small, and this is a statistic free of falsified, artificially-induced operator biases. In the adversary situation, the bias in MUF by the operator would reduce MUF. If the inspector duplicates every operator measurement $(\hat{D}+MUF)$ then becomes an inspector MUF since all operator measurements will then be replaced by inspector measurements.

Some of the statistical relationships between MUF, $\hat{D}$ and $(\hat{D}+MUF)$ will be shown in the following development.

The relationships are shown for only one stratum but the relationships are true over all strata in MUF. Let

z_i denote the operator's stated amount value for the i th item

z'_i denote the inspector's stated amount value for the i th item

n denote the number of items in the inspector random sample size

N denote the number of items in MUF.

For the sake of convenience the subscripting is arranged to include the first n items in the random sample.

The purpose of the next several sections is to develop relationships between the MUF, D and $D+MUF$ statistics. For simplification these relationships are developed for only one stratum but the general conclusions as shown are true over all strata.

The Model and Variance for MUF

The model for the operator's stated amount value for the i th item is $z_i = \mu_i + b + \epsilon_i$ where μ_i is the true amount value, b is the bias in the operator's measurement and ϵ_i is a random measurement error. b is fixed for N items in the MUF but is a random component in the sense of occurring from a population of biases which vary over time and facilities. (This particular point could be developed in considerable detail but would lead us afield from our main line of development here.) For our simple representation here one has

$$MUF = \sum_{i=1}^N z_i = \sum_{i=1}^N \mu_i + Nb + \sum_{i=1}^N \epsilon_i.$$

The variance in MUF due to random variation is

$$\sigma_{MUF_r}^2 = N\sigma_{Or}^2.$$

The variance in MUF due to the systematic error is

$$\sigma_{MUF_s}^2 = N^2\sigma_{Os}^2.$$

The total MUF variance then is

$$\sigma_{\text{MUF}}^2 = \sigma_{\text{MUF}_s}^2 + \sigma_{\text{MUF}_r}^2 = N^2 \sigma_{\text{Os}}^2 + N \sigma_{\text{Or}}^2.$$

(There are some occasions in this document when a distinction is made between the variance of MUF and $\hat{D}$ under the null and alternative hypotheses but these distinctions will not be considered here.)

The Model and Variance for $\hat{D}$

The model for the inspector's amount value for the i th item is

$$z'_i = \mu_i + b' + \epsilon'_i$$

where μ_i is the true amount value, b' is the bias in the inspector's measurement and ϵ'_i is a random measurement error. The model for the difference between the inspector's amount value and the operator's amount value for the i th item is

$$d_i = z'_i - z_i = b' - b + \epsilon'_i - \epsilon_i$$

and for the average difference

$$\bar{d} = \sum_{i=1}^n d_i / n = b' - b + \sum_{i=1}^n (\epsilon'_i - \epsilon_i) / n.$$

The variance of $\bar{d}$ based on random variation is

$$\sigma_{\bar{d}_r}^2 = (\sigma_{\text{Ir}}^2 + \sigma_{\text{Or}}^2) / n.$$

The variance of $\bar{d}$ based on systematic error is

$$\sigma_{\bar{d}_s}^2 = \sigma_{\text{Is}}^2 + \sigma_{\text{Os}}^2.$$

The total variance of $\bar{d}$ based on random and systematic errors is

$$\sigma_{\bar{d}}^2 = \sigma_{\bar{d}r}^2 + \sigma_{\bar{d}s}^2.$$

The statistic $D = N\bar{d}$ is an unbiased estimate of the inspector bias minus the operator bias in MUF because the sample items are randomly chosen even though, for convenience sake in the exposition, the sample subscripts are chosen to run from 1 to n. The variance in $\hat{D}$ due to random variation is

$$\sigma_{\hat{D}r}^2 = \sigma_{\bar{d}r}^2 = N^2 \sigma_{\bar{d}r}^2 = N^2 (\sigma_{Ir}^2 + \sigma_{Or}^2) / n$$

and due to systematic error is

$$\sigma_{\hat{D}s}^2 = \sigma_{\bar{d}s}^2 = N^2 \sigma_{\bar{d}s}^2 = N^2 (\sigma_{Is}^2 + \sigma_{Os}^2)$$

and the total variance in $\hat{D}$ is

$$\sigma_{\hat{D}}^2 = \sigma_{\hat{D}r}^2 + \sigma_{\hat{D}s}^2 = N^2 (\sigma_{Ir}^2 + \sigma_{Or}^2) / n + N^2 (\sigma_{Is}^2 + \sigma_{Os}^2).$$

If the operator and inspector error variances are the same, then

$$\sigma_{\hat{D}}^2 = 2N^2 (\sigma_s^2 + \sigma_r^2) / n.$$

The Model for $(\hat{D} + \text{MUF})$

The model for the statistic $(\hat{D} + \text{MUF})$ is obtained as follows,

$$\text{MUF} = \sum_{i=1}^N \mu_i + Nb + \sum_{i=1}^N \epsilon_i$$

$$\hat{D} = N(b' - b) + N \sum_{i=1}^n (\epsilon_i' - \epsilon_i) / n$$

$$\begin{aligned}
(\hat{D}+MUF) &= \sum_{i=1}^N \mu_i + Nb' + N \sum_{i=1}^N (\epsilon_i' - \epsilon_i) / n + \sum_{i=1}^N \epsilon_i \\
&= \sum_{i=1}^N \mu_i + Nb' + \frac{N}{n} \sum_{i=1}^N \epsilon_i' + \sum_{i=1}^N \epsilon_i (1 - N/n) + \sum_{i=1}^N \epsilon_i
\end{aligned}$$

This illustrates how the statistic $(\hat{D}+MUF)$ replaces Nb , the MUF operator bias by Nb' , the MUF inspector bias where $\sum_{i=1}^N \mu_i$ is the true MUF value. The variance of $(\hat{D}+MUF)$ is

$$\sigma^2_{(\hat{D}+MUF)} = N^2 \sigma_{Is}^2 + \frac{N^2}{n} \sigma_{Ir}^2 + N \left(\frac{N}{n} - 1 \right) \sigma_{Or}^2$$

If $n = N$, then

$$(\hat{D}+MUF) = \sum_{i=1}^N \mu_i + Nb' + \sum_{i=1}^N \epsilon_i'$$

which is the model for the inspector's MUF. The variance $\sigma^2_{(\hat{D}+MUF)}$ can be obtained directly from the statistical variation properties of MUF and $\hat{D}$ since $(\hat{D}+MUF)$ is a linear combination of the statistics MUF and $\hat{D}$.

The Variance, Covariance Relationships Between MUF, $\hat{D}$ and $(\hat{D}+MUF)$

The covariance in MUF and $\hat{D}$ due to the random component is

$$\begin{aligned}
\sigma_{(\hat{D}, MUF)}_r &= E \left\{ \sum_{i=1}^N \epsilon_i - E \left(\sum_{i=1}^N \epsilon_i \right) \right\} \left\{ N \sum_{i=1}^N (\epsilon_i' - \epsilon_i) / n - E \left(N \sum_{i=1}^N (\epsilon_i' - \epsilon_i) \right) \right\} \\
&= E \left\{ \sum_{i=1}^N \epsilon_i \right\} \left\{ N \sum_{i=1}^N (\epsilon_i' - \epsilon_i) / n \right\} \\
&= N E \left\{ \frac{-(\sum_{i=1}^N \epsilon_i)^2}{n} \right\} = N \left(\frac{n \sigma_{Or}^2}{n} \right) = N \sigma_{Or}^2.
\end{aligned}$$

The covariance in MUF and $\hat{D}$ due to the systematic component is

$$\sigma_{(MUF, \hat{D})}_s = E \left\{ Nb - E(Nb) \right\} \left\{ N(b' - b) - E(N(b' - b)) \right\}.$$

These "naturally-occurring" components which include bias are sampled from an infinite population of biases such that $E(b) = E(b') = 0$. Then

$$\sigma_{(MUF, \hat{D})_s} = - E(N^2 b^2) = - N^2 \sigma_{Os}^2.$$

Thus

$$\sigma_{MUF, \hat{D}} = - (N \sigma_{Or}^2 + N^2 \sigma_{Os}^2) = - \sigma_{MUF}^2.$$

The variance of $(\hat{D} + MUF)$, then, is

$$\begin{aligned} \sigma_{(\hat{D}+MUF)}^2 &= \sigma_{MUF}^2 + 2\sigma_{MUF, \hat{D}} + \sigma_{\hat{D}}^2 \\ &= \sigma_{MUF}^2 - 2\sigma_{MUF}^2 + \sigma_{\hat{D}}^2 \\ &= \sigma_{\hat{D}}^2 - \sigma_{MUF}^2 \end{aligned}$$

This last relationship indicates an easy way of calculating $\sigma_{(\hat{D}+MUF)}^2$ if σ_{MUF}^2 and $\sigma_{\hat{D}}^2$ have already been calculated. It also provides a good way of checking calculations.

The covariance of $(\hat{D}+MUF)$ and MUF is

$$\begin{aligned} \sigma_{(\hat{D}+MUF), MUF} &= \sigma_{MUF}^2 + \sigma_{MUF, \hat{D}} \\ &= \sigma_{MUF}^2 - \sigma_{MUF}^2 \\ &= 0 \end{aligned}$$

This illustrates a very interesting and important property of the statistic $(\hat{D}+MUF)$, viz., its zero covariance, zero correlation with the statistic MUF . If $(MUF, \hat{D}+MUF)$ have a bivariate normal density then the statistics MUF and $(\hat{D}+MUF)$ are also statistically independent.

The covariance of $\hat{D}$ and $(\hat{D}+MUF)$ is

$$\begin{aligned}\sigma_{\hat{D},(\hat{D}+MUF)} &= \sigma_{\hat{D},MUF} + \sigma_{\hat{D}}^2 \\ &= -\sigma_{MUF}^2 + \sigma_{\hat{D}}^2 = \sigma_{(\hat{D}+MUF)}^2\end{aligned}$$

This last relationship illustrates a convenient way of calculating the covariance term for $\hat{D}$ and $(\hat{D}+MUF)$.

The tables on the following page illustrate some easy ways of summarizing these results.

Any of these tables taken as matrices are singular due to the linear dependence of the statistical estimates therein.

It has already been mentioned that the operator generally does better in avoiding detection by partitioning the goal quantity M over the different diversion paths of

- A. diversion into MUF
- B. falsification
 - a. gross defects
 - b. biases

by choosing certain advantageous methods of partitioning. The inspector has a very convenient way of analyzing the data using the $(\hat{D}+MUF)$ statistic.

The evaluation of $\hat{D}$, MUF and $(\hat{D}+MUF)$ are somewhat academic for the inspector if the attribute measurement and/or the variable measurement in the attribute mode have detected one or more gross defects. If the amount G is removed by gross defects and the prescribed inspector sample sizes n_a and n_v are used, the probability of not detecting one or more gross defects is $\beta^{G/M}$ or less. The probability that a $(\hat{D}+MUF)$ statistic will fall below the acceptance level $Z_{1-\alpha} \sigma_{(\hat{D}+MUF)}$ is given by the area under the distribution curve from $-\infty$ to the acceptance level. Using the cumulative normal distribution symbolism then

The Variance-Covariance Matrix of the Statistics MUF, $\hat{D}$, and $(\hat{D}+MUF)$

	MUF	$\hat{D}$	$(\hat{D}+MUF)$
MUF	$N\sigma_{Or}^2 + N^2\sigma_{Os}^2$	$-(N\sigma_{Or}^2 + N^2\sigma_{Os}^2)$	0
$\hat{D}$	$-(N\sigma_{Or}^2 + N^2\sigma_{Os}^2)$	$\frac{N^2}{n}\sigma_{Or}^2 + N^2\sigma_{Os}^2 + \frac{N^2}{n}\sigma_{Ir}^2 + N^2\sigma_{Is}^2$	$N(\frac{N-1}{n})\sigma_{Or}^2 + \frac{N^2}{n}\sigma_{Ir}^2 + N^2\sigma_{Is}^2$
$(\hat{D}+MUF)$	0	$N(\frac{N-1}{n})\sigma_{Or}^2 + \frac{N^2}{n}\sigma_{Ir}^2 + N^2\sigma_{Is}^2$	$N(\frac{N-1}{n})\sigma_{Or}^2 + \frac{N^2}{n}\sigma_{Ir}^2 + N^2\sigma_{Is}^2$

Which may be written as

	MUF	$\hat{D}$	$(\hat{D}+MUF)$
MUF	σ_{MUF}^2	$-\sigma_{MUF}^2$	0
$\hat{D}$	$-\sigma_{MUF}^2$	$\sigma_{\hat{D}}^2$	$\sigma_{\hat{D}}^2 - \sigma_{MUF}^2$
$(\hat{D}+MUF)$	0	$\sigma_{\hat{D}}^2 - \sigma_{MUF}^2$	$\sigma_{\hat{D}}^2 - \sigma_{MUF}^2$

The correlation matrix is as follows.

	MUF	$\hat{D}$	$(\hat{D}+MUF)$
MUF	1	$\rho_{MUF, \hat{D}}$	0
$\hat{D}$	$\rho_{MUF, \hat{D}}$	1	$\sqrt{1 - \rho_{MUF, \hat{D}}^2}$
$(\hat{D}+MUF)$	0	$\sqrt{1 - \rho_{MUF, \hat{D}}^2}$	1

$$\text{where } \rho_{MUF, \hat{D}} = \frac{\sigma_{MUF, \hat{D}}}{\sigma_{MUF} \sigma_{\hat{D}}} = \frac{-\sigma_{MUF}^2}{\sigma_{MUF} \sigma_{\hat{D}}} = -\frac{\sigma_{MUF}}{\sigma_{\hat{D}}}$$

$$Q_2 = \Phi \left(\frac{Z_{1-\alpha} \sigma_{(\hat{D}+MUF)} - E(\hat{D}+MUF)}{\sigma_{(\hat{D}+MUF)}} \right)$$

is the probability that a random $(\hat{D}+MUF)$ value will be less than the acceptance level. Since $E(MUF) = MUF_T$, $E(\hat{D}) = D$, the bias in MUF, the expected value of $(\hat{D}+MUF)$ is $MUF_T + D$, the true amount of material unaccounted for aside from the material obtained from gross defects. Then

$$Q = \beta^{G/M} Q_2$$

is the approximate probability of nondetection of one of the three paths since Q_2 is, for all practical purposes, statistically independent of $\beta^{G/M}$. Thus the maximum value that Q can attain, subject to the constraint $M = G + D + MUF_T$ will yield the maximum probability of nondetection attainable by the operator.

In the development of the properties of and the relationships between the statistics MUF , $\hat{D}$ and $(\hat{D}+MUF)$, slight mention is made of how an induced operator bias and process variation in order to falsify data are handled. The operator's artificially-induced bias does not affect the variance of the statistics because this bias is directional and does not affect the natural variability of the statistics under either the null or alternative hypotheses. The induced process variation, however, affects σ_{Or}^2 , the operator random error variance which occurs in $\hat{D}$ and $(\hat{D}+MUF)$, and its value, under the alternative hypothesis of induced process variation, is usually treated as a multiple of σ_{Or}^2 under the null hypothesis. The factor of multiplication depends upon the constraints imposed on the operator by the inspector variance and outlier tests of operator data. F_C , the finite population under induced process variance, is ignored. (See page 113.)

A $Q_{\max}$ Situation in a Different Inspector Testing Situation

It was previously shown how the amount diverted should be partitioned into two components G and $M-G$ in order to maximize the probability of nondetection when the inspector uses the tests for gross defects (complete and partial) and the $(\hat{D}+MUF)$ statistic to test for bias plus diversion into MUF. The following Q maximization is done under the assumption that the inspector uses the tests for gross defectives, and tests for bias and diversion into MUF separately by the $\hat{D}$ and MUF statistics. Define D , G , and MUF_T as follows.

D - The amount that the operator MUF is biased.
(This can contain operator bias which is not purposely induced.)

G - The amount the operator obtains by gross defects.

MUF_T - The true amount diverted into MUF.

Of the three tests the ones on $\hat{D}$ and MUF involve correlated statistics so that the following argument must be interpreted with this cautionary proviso. Under the assumption of statistical independence the probability of nondetection is

$$Q = \beta^{G/M} \Phi \left(\frac{z_{1-\alpha} \sigma_{\hat{D}|H_0} - D}{\sigma_{\hat{D}|H_1}} \right) \Phi \left(z_{1-\alpha} - \frac{MUF_T}{\sigma_{MUF}} \right)$$

$$= \beta^{G/M} \Phi(\theta'_D) \Phi(\theta'_{MUF_T})$$

where $M = G + D + MUF_T$. Lagrangian Multipliers are useful here. Write

$$H = \ln Q + \lambda (M - G - D - MUF_T).$$

Then

$$\frac{\partial H}{\partial G} = 0 \text{ implies } \frac{\ln \beta}{M} = \lambda$$

$$\frac{\partial H}{\partial D} = 0 \text{ implies } - \frac{1}{\sigma_{\hat{D}|H_1}} R \left(\frac{Z_{1-\alpha} \sigma_{\hat{D}|H_0} - D}{\sigma_{\hat{D}|H_1}} \right) = \lambda$$

$$\frac{\partial H}{\partial \text{MUF}_T} = 0 \text{ implies } - \frac{1}{\sigma_{\text{MUF}_T}} R \left(Z_{1-\alpha} - \frac{\text{MUF}_T}{\sigma_{\text{MUF}}} \right) = \lambda$$

These equations yield the solution

$$\Theta_D = R^{-1} \left(\frac{-\ln \beta}{M / \sigma_{\hat{D}|H_1}} \right)$$

$$\Theta_{\text{MUF}_T} = R^{-1} \left(\frac{-\ln \beta}{M / \sigma_{\text{MUF}}} \right)$$

$$D = Z_{1-\alpha} \sigma_{\hat{D}|H_0} - \Theta_D \sigma_{\hat{D}|H_1}$$

$$\text{MUF}_T = (Z_{1-\alpha} - \Theta_{\text{MUF}_T}) \sigma_{\text{MUF}}$$

$$G = M - D - \text{MUF}_T.$$

There is a caveat here, however, since a negative G as a solution does not fulfill the boundary condition that $G \geq 0$. If the first attempt at solution yields a negative G value, G is set equal to zero. The resulting maximization condition is

$$\frac{1}{\sigma_{\hat{D}|H_1}} R \left(\frac{Z_{1-\alpha} \sigma_{\hat{D}|H_0} - D}{\sigma_{\hat{D}|H_1}} \right) = \frac{1}{\sigma_{\text{MUF}}} R \left(Z_{1-\alpha} - \frac{\text{MUF}_T}{\sigma_{\text{MUF}}} \right) = 0$$

where $MUF_T = M - D$. This approach is exactly true but the relationship to solve for D , and subsequently $MUF_T = M - D$, is implicit. This can be solved by trial and error but an iterative and systematic procedure using the table of $R(\theta)$ values is Newton's method. Let $D^{(k)}$ denote the k th approximation to D . Write

$$F(D^{(k)}) = \frac{1}{\sigma_{\hat{D}|H_1}} R\left(\frac{z_{1-\alpha} - \frac{\sigma_{\hat{D}|H_1}}{\sigma_{\hat{D}|H_1}} D^{(k)}}{\sigma_{\hat{D}|H_1}}\right) - \frac{1}{\sigma_{MUF}} R\left(\frac{z_{1-\alpha} - (M - D^{(k)})}{\sigma_{MUF}}\right).$$

$$\text{Let } D^{(1)} = M\sigma_{\hat{D}|H_1} / (\sigma_{\hat{D}|H_1} + \sigma_{MUF})$$

$$D^{(k+1)} = D^{(k)} - \frac{F(D^{(k)})}{F'(D^{(k)})}$$

$F'(D^{(k)})$ is easily approximated by

$$F'(D^{(k)}) = \frac{R'(\theta_{D^{(k)}})}{\sigma_{\hat{D}|H_1}} - \frac{R'(\theta_{MUF_T})}{\sigma_{MUF}}$$

where $R'(\theta_{D^{(k)}}) = \left\{ R(\theta_{D^{(k)}} + 0.01) - R(\theta_{D^{(k)}}) \right\} / 0.01$ for example.

C. Plans for Evaluation of Inspection Data

Through records and reports submitted to the Agency at periodic intervals and at the time of an inspection, the inspector has a picture of the state of the operator's material control. The inspector has a twofold interest in this and wants answers to the following questions:

- (1) Do the records and reports correctly represent the operator's material balance status?
- (2) Is the operator's material control adequate when judged against certain criteria?

It will be noted later that under certain conditions to be discussed, the inspector may seek an answer to (2) and pass judgment on the state of control without having to reach a final conclusion as to whether or not the operator's records correctly reflect the material balance status. The advantages of this approach will be discussed. Nevertheless, it is helpful to think of the inspector's evaluation as proceeding in the two distinct stages because evidence that the operator has presented a distorted picture of his state of control, either through deliberate falsification or by a poor measurement system, (defined in its broad terms to include all errors and mistakes that affect a given reported value) may well trigger corrective action of a kind different from that triggered by a negative answer to question (2). The important point is that although the inspector would hope to obtain a positive answer to the first question, thereby giving him reason to believe the operator's data, he can proceed to obtain an answer to (2) irrespective of the final conclusion reached on question (1).

In seeking answers to these questions, the inspector has at his disposal the operator's records and reports, both as submitted and including the detailed source data

and calculations. Further, he has the capability to perform his own measurements on a statistical sampling basis. The design of his inspection activities was presented in Section A. In this section, the statistical evaluation of the inspection data, as compared with the operator's data, is discussed.

The evaluation methodology is conveniently divided into the following topics, using the terminology of Section A.

- (1) Verification of item identification
- (2) Attributes inspection
- (3) Variables inspection
 - (a) for random error variance
 - (b) for unaccounted-for bias
- (4) Supplementary statistical tests
- (5) Evaluation of MUF
 - (a) operator's MUF
 - (b) MUF adjusted for bias

In passing judgment on the correctness of the operator's data, it is tacitly assumed that rejection of any part of his data corresponds to "detection of falsification." This assumption is made when a given plan for inspection is evaluated at the design stage in the sense that the key criterion is how effective is a given plan at detecting a given amount of falsification. At the planning stage, it is reasonable and in a sense necessary to make this assumption. However, as implemented, it is not reasonable to discard the operator's data in total if any aspect of it should fail one of the statistical tests. Rather, the aim should be to correct the data to the extent practicable. This may involve 100% re-inspection in certain instances. At what point the entire data base is discarded is left open to question, for the statistical tests can only provide the bases for making probabilistic statements

about the operator's data. The final decision on the acceptability of the operator's data is left to the inspector, who is influenced by the results of these tests to a large extent in making his judgment. However, it is difficult to specify the actions to take at each stage of the evaluation except in a general sense.

Verification of Item Identification

Item identification is an important first step in an inspection. This task is not statistical in nature since it should normally consist of a 100 percent attributes inspection effort. The aim is to provide assurance that all items listed can be located, and that there are not items that are physically accounted for but do not appear on the records. In the case of flow strata, the item identification will have to be scheduled to coincide with receipts and shipments.

Attributes Inspection

In designing for attributes inspection within a given stratum, two sample sizes were developed, n_a and n_v . The first sample size, n_a , is the number of samples to measure with an attributes tester, which may be defined to be a measurement method that will detect gross discrepancies. The sample size, n_v , gives the number of samples to measure by a variables measurement method that has the ability to detect relatively small discrepancies.

From an attributes viewpoint, a decision is reached on each item sampled as to whether it is a defect or acceptable, where a defect is an item with a discrepancy between the operator and inspector value in excess of some reject limit, RL.

Consider the definition of RL for the attributes tester. From a design viewpoint, in determining n_a , a reject was regarded as a discrepancy of size $\bar{X}$, the average amount in a given container. This was done to

yield a minimum sample size that would combat the operator's best strategy, one corresponding to the complete removal of material (or falsification) from a small number of items. From an evaluation viewpoint, RL may be defined differently, taking advantage of the method's ability to detect discrepancies smaller than $\bar{X}$.

Before defining RL for the attributes measurement, consider the two different kinds of gross defects that the inspector would like to detect. One kind of defect may occur as a result of recording and/or calculation mistakes. That is, the source data are correct, but the reduction of the source data to a reported value is incorrect. To detect this, the inspector need make no remeasurements, but need only check the calculations against the source data. On the other hand, the calculations may be correct with the source data in error. The latter condition can only be determined by remeasurement.

Although one may regard these as two distinct problems and draw separate samples, from an inspection viewpoint, there is no need to do this. Rather, it is logical to combine the two tests in the sense of drawing a single sample.

It remains to define RL, to specify inspector actions in the event discrepancies less than RL are found, and in the event the stratum is rejected. As pointed out previously, the sample size n_a was based on the assumption that the entire contents of a given item are missing, i.e., that the discrepancy is of size $\bar{X}$. Clearly, the discrepancy need not be that large before it will be detected by the inspector. In fact, for the check on calculational and recording mistakes, any discrepancy is detectable, there being no measurement errors involved (other than rounding errors).

For the check on the source data, the attributes measurement will almost surely detect a discrepancy in excess of $4\delta\bar{X}$, where δ is the coefficient of variation for the method (a typical value for σ might be 0.05 e.g., 5 percent).

It is reasonable to choose $RL = 4\delta\bar{X}$ for the attributes measurement. Since the check on records and calculations is made on the same sample items, RL would logically apply to the total reported value, rather than to the source data itself. However, it is not desirable to permit compensating errors to occur in the system, and calculational mistakes of any kind should not be permitted. With these points in mind, the evaluation activity within a given stratum for the attributes inspection for gross defects consists of the following steps.

- (1) Randomly select the n_a items and measure each item with the attributes instrument.
- (2) For each sampled item, check the operator's calculation from the source data. Record and correct any discrepancy smaller than $4\delta\bar{X}$. If a discrepancy exceeds $4\delta\bar{X}$, make a 100 percent check of all calculations in the stratum in question before proceeding further with the attributes instrument.
- (3) Having checked the calculations and corrected mistakes either on an item basis or for all items in the stratum, compare the result of the attributes instrument with the operator's corrected value. If one or more of the sampled items produces a discrepancy in excess of $4\delta\bar{X}$, that stratum is rejected. This should lead to remeasurement of all items in the stratum, an action that is possible at this phase of the inspection because of the immediate availability of results.

Next, consider the use of the variable method as an attributes measurement, where the sample size was chosen to be n_v . The variables method, capable of making precise measurements and of producing data that can be analyzed on a variables basis is also used on an attributes basis in order to combat the operator's strategy of partially removing the SNM content of items by amounts just under those that the attributes instrument can detect. Thus, on an attributes basis, a defect for the variables tester was set at $6\delta\bar{X}$, the detection level for the attributes instrument. Employing reasoning similar to that used for the attributes instrument, the steps in the evaluation process are as follows:

- (1) Randomly select the n_v items and measure each item with the variables instrument.
- (2) For each sampled item, check the operator's calculation from the source data. Record and correct all discrepancies.
- (3) Compare the inspector result with the operator result, as corrected for calculational mistakes. If one or more of the sampled items produces a discrepancy in excess of $RL = 4\sigma d_i$, that stratum is rejected. Here, σd_i is the standard deviation of the difference between the operator's and inspector's value using the variables method, and including the systematic error variances.

One other distinction is made between the two attributes tests. With the attributes instrument, the inspection results are immediately available and necessary actions involving 100 percent inspection of certain strata can take place during the inspection. With the variables instrument, however, the results must await completion of the laboratory analyses, and the validity of the operator's data cannot be verified until some time after the inspection has been completed. Thus, rejection

of a stratum in this latter case may lead to different kinds of inspector actions depending on the circumstances. These possible actions may vary from a partial or complete reinspection to an adjustment of the operator's data before proceeding with the MUF calculation.

Variables Inspection

Once the attributes inspection results using both the attributes and variables instrument have demonstrated that the frequency and size of misstatements in flows and inventories are such that the calculated MUF is not unduly affected thereby, the data using the variables instrument are evaluated on a variables basis to see if reported measurement variances are supported by the data, and to detect small unaccounted-for biases.

Random Error Variance

The operator provides values for his measurement error variances, systematic and random. It is assumed that these values are acceptable as determined in the review of design information. Based on inspection results, it is possible to verify his random error variance but not his systematic. This is an important step in the evaluation because it combats the operator's strategy of misrepresenting his variance in order to decrease the sensitivity of the statistical tests. Further, and perhaps more important, the operator's strategy of small removals may be reflected in the data as a large random error variance, either by design or as a consequence of his diversion strategy. Whatever the cause, the inspector must be able to detect when stated variances are not supported by the data.

Within a given stratum, the statistical test on the validity of the random error variance is based on the difference statistic, or the d-statistic, d_i , which is the difference between the operator's and inspector's

value (element or isotope weight) for the i th item in the sample. The following steps form the evaluation:

- (1) Calculate s_d^2 , the sample variance of d_i values.

$$s_d^2 = \frac{n \sum d_i^2 - (\sum d_i)^2}{n(n-1)}$$

where n is the sample size ($n = n_v$).

- (2) Calculate the statistic

$$\chi_{n-1}^2 = \frac{(n-1)s_d^2}{\sigma_{Or}^2 + \sigma_{Ir}^2}$$

where σ_{Or}^2 is the stated random error variance for the operator, and σ_{Ir}^2 is the inspector's random error variance.

- (3) Under the hypothesis that σ_{Or}^2 and σ_{Ir}^2 are correctly stated, χ_{n-1}^2 is distributed as chi-square with $(n-1)$ degrees of freedom. The calculated value is compared with the appropriate lower and upper critical values in a chi-square table to see if s_d^2 supports the hypothesized values, σ_{Or}^2 and σ_{Ir}^2 . If it does not, conclude that the operator's value σ_{Or}^2 is not correct, assuming that the inspector can properly assess the size of σ_{Ir}^2 .

Suppose the hypothesis is rejected and it is concluded that σ_{Or}^2 is incorrectly represented. What follows then depends on the circumstances. If σ_{Or}^2 is understated, this could mean that the operator was overly optimistic and that his measurement system is not as good as represented. Perhaps the operator understated his variance in order to make his design information acceptable when judged against Agency standards. The variance s_d^2 would also be inflated leading to an under-

statement on the size of σ_{Or}^2 if the operator biases a number of items by differing amounts.

Understating the size of σ_{Or}^2 can represent a serious flaw in the operator's system, and the action to take in the event this is detected is not clear. Depending on the circumstances, this can lead to outright rejection of the operator's data and the conclusion that the inspection results are unacceptable. Or, it may require further inspection. On the other hand, if the stated σ_{Or}^2 is well below acceptable standards, and if there is no suspicion that the operator is intentionally inflating the size of σ_d^2 , all that may be necessary would be to replace σ_{Or}^2 in the later MUF evaluation by an estimate, given by

$$\hat{\sigma}_{Or}^2 = s_d^2 - \sigma_{Ir}^2$$

In this event, it is presumed that $\hat{\sigma}_{Or}^2$ is not too different from acceptable standards of performance.

Suppose, now, that σ_{Or}^2 is overstated, i.e., the computed chi-square value is too small. This would indicate that the operator is too pessimistic, perhaps intentionally so in order to make it difficult for the inspector to detect small biases. In this case, simply replace σ_{Or}^2 by $\hat{\sigma}_{Or}^2 = s_d^2 - \sigma_{Ir}^2$ in the later evaluations, and call the attention of the operator to the fact that his measurement system is better than represented. This would likely occur only if σ_{Or}^2 is rather large relative to acceptable standards in the first place, but not so large as to lead to rejection of the operator's design information.

Unaccounted-for Bias

An important function of the variables instrument inspection is to detect small, unaccounted-for biases. These are biases that are introduced by the operator either intentionally or not and that affect his MUF. The expression "unaccounted-for" biases is used in order to emphasize that some differences are expected to occur between the operator and the inspector because of the recognized presence of systematic errors. It is the bias over and above those explained by stated systematic error variances that the inspector would like to detect.

Within a given stratum, the bias test is simple to implement, and consists of the following steps:

(1) Calculate $\bar{d}$, the average of the d_i values.

(2) Compute $\sigma_{\bar{d}} = \sqrt{(\sigma_{Os}^2 + \sigma_{Is}^2) + (\sigma_{Or}^2 + \sigma_{Ir}^2)/n_v}$

where σ_{Os}^2 and σ_{Is}^2 are the systematic error variances for the operator and inspector respectively, and σ_{Or}^2 and σ_{Ir}^2 are the corresponding random error variances. σ_{Or}^2 is based on design information unless the variance test just discussed invalidates this value.

(3) Divide $\bar{d}$ by $\sigma_{\bar{d}}$. If this ratio exceeds three in absolute value, conclude that there is statistical evidence of unaccounted-for bias.

In the event the statistical test indicates that there is a bias, then the inspector must use his judgment to decide on the course of action. If he attributes this to an understatement of σ_{Os}^2 , then for the later calculation of LEMUF, σ_{Os}^2 should be appropriately adjusted. If he feels that σ_{Os}^2 is realistically appraised then it is necessary to correct the operator's data for bias. If the bias is too large, the operator's data

may be declared suspect and rejected outright, with further inspection required before the final evaluation of the operator's material balance performance can be completed.

Attention is next directed at the accumulation of biases over all strata. Within individual strata, biases may be so small as to avoid detection, but summed over all strata, they could accumulate to an amount of concern to the inspector. An evaluation is made of the combined differences between the operator's and inspector's measurements as they relate to the operator's calculated MUF.

This evaluation is conducted by forming an appropriate linear combination of the $\bar{d}_j$ values, where $\bar{d}_j$ is the average difference for stratum j . This linear combination is denoted by $\hat{D}$ and may be written in the form

$$\hat{D} = N_1 \bar{d}_1 + N_2 \bar{d}_2 + \dots - N_{k-1} \bar{d}_{k-1} - N_k \bar{d}_k$$

where N_j is the total number of items in stratum j . There are a total of k strata, and in the equation for $\hat{D}$ these strata are so ordered that the first m strata are those affecting MUF in a positive way, while the last $(k-m)$ strata affect MUF negatively.

The variance of $\hat{D}$ must be computed. This involves propagating the errors that affect each $\bar{d}_j$, where the errors for the operator are based on design information unless the variance tests do not support them. The errors may be propagated by writing the variance of $\hat{D}$ as the following sum, where it is assumed that the comparisons are made on an element basis rather

rather than for isotope weights.

$$\begin{aligned} \text{var } \hat{D} = & \sum_{k=1}^K A_{1k} \sigma_{\Omega_k}^2 + \sum_{k=1}^K A_{2k} \sigma_{\xi_k}^2 + \sum_{p=1}^P A_{3p} \sigma_{\Gamma_p}^2 + \sum_{p=1}^P A_{4p} \sigma_{\tau_p}^2 \\ & + \sum_{t=1}^T A_{5t} \sigma_{\Lambda_t}^2 + \sum_{t=1}^T A_{6t} \sigma_{\zeta_t}^2 \end{aligned}$$

In this equation,

$\sigma_{\Omega_k}^2$ = systematic error variance due to weighing on scale k,

$\sigma_{\xi_k}^2$ = random error variance due to weighing on scale k,

$\sigma_{\Gamma_p}^2$ = systematic error variance due to measurement of element factor with technique p (includes sampling and analytical),

$\sigma_{\tau_p}^2$ = random error variance due to measurement of element factor with technique p (includes sampling and analytical),

$\sigma_{\Lambda_t}^2$ = systematic error variance due to making NDA measurement with equipment t,

$\sigma_{\zeta_t}^2$ = random error variance due to making NDA measurement with equipment t.

The sums in the equation for $\text{var } \hat{D}$ include components for both the operator and the inspector. That is, there are K scales in total, etc.

Once the above variances, expressed on an absolute rather than a relative basis, are calculated, then the A_{ij} values must be calculated before $\text{var } \hat{D}$ can be computed. These are computed as follows:

- A_{1k} :
- (1) Multiply N_i by the average element factor.
 - (2) Sum these terms over all the strata in which scale k is used, being careful to preserve the proper sign in the sum.
 - (3) Square this sum. This is A_{1k} . There is one such value for each of the K scales.

- $\underline{A_{2k}}$:
- (1) Multiply N_i by the average element factor.
 - (2) Square this quantity.
 - (3) Divide by n_i , the number of weighings made in the stratum.
 - (4) Sum over all the strata in which scale k is used. In this instance, all terms in the sum are positive. This sum is A_{2k} , and there is one such value for each of the K scales.
- $\underline{A_{3p}}$:
- (1) Multiply N_i by the average item net weight.
 - (2) Sum these terms over all the strata in which analytical technique p is used, being careful to preserve the proper sign in the sum.
 - (3) Square the sum. This is A_3 . There is one such value for each of the P analytical techniques.
- $\underline{A_{4p}}$:
- (1) Multiply N_i by the average item net weight.
 - (2) Square this quantity.
 - (3) Divide by the number of analyses that relate to the factor(s) for the items inspected in the stratum.
 - (4) Sum over all the strata in which analytical technique p is used. In this instance, all terms in the sum are positive. This sum is A_4 , and there is one such value for each of the P analytical techniques.
- $\underline{A_{5t}}$:
- (In this case, and for A_6 , the element values are reported directly without the intermediate steps of separately determining the weights and the element factors.)

- (1) Sum the N_i values over the strata in which NDA equipment t is used, being careful to preserve the proper sign in the sum.
- (2) Square this sum. This is A_{5t} . There is one such value for each of the T NDA instruments.

- A_{6t} :
- (1) Square N_i .
 - (2) Divide by the number of NDA analyses.
 - (3) Sum over all the strata in which NDA equipment t is used. In this instance, all terms in the sum are positive. This sum is A_{6t} , and there is one such sum for each of the NDA instruments.

Once var D is computed, the following ratio is formed:

$$\frac{\hat{D}}{\sqrt{\text{var } \hat{D}}}$$

This is assumed to be normally distributed with zero mean and unit standard deviation under the hypothesis that there is no unaccounted-for bias. If the ratio is less than 2 in absolute value, accept the operator's data as verified. If it exceeds 2, the data are rejected. Depending on the circumstances, rejection of the data may be followed by further inspection or by adjustment of the operator's data in the MUF evaluation.

Supplementary Statistical Tests

The preceding sections have presented the statistical tests that the inspector will use to evaluate the quality of the operator's data. Although these are the tests that are essential to a complete evaluation and probabilities of detection at the design stage are based on them, the inspector should not limit himself to their use in his evaluation. There are other tests at his disposal that will be useful in detecting anomalies in the operator's data, or that will provide further information about the operator's measurement error structure. Some of particular potential importance are discussed briefly.

- (1) With respect to recording or calculational errors, the attributes test detects the gross mistakes and corrects smaller ones that may occur in the sampled items. A diverter, recognizing that the inspector has no real basis for distinguishing between honest mistakes and attempts at diversion, may choose to commit an inordinately large number of small "mistakes" in order to mask attempts at diversion. If these are indeed honest mistakes, there should be an equal number that affect MUF positively and negatively on the average. A statistical sign test can be made to detect when a suspiciously large number of mistakes are made in a preferential direction. Table 1 gives a test criterion that can be used. Here c is the critical number of mistakes such that if the number of mistakes in a given direction is equal to or less than c for a given total number of mistakes k , there is statistical evidence that mistakes are being made in a preferential direction.

TABLE 1 Critical Values for Directional Errors

<u>k = Total</u> <u>Number of Mistakes</u>	<u>c = Critical Value</u>
6-8	0
9-11	1
12-14	2
15-16	3
17-19	4
20-22	5
23-24	6
25-27	7
28-29	8
30-32	9
33-34	10
35-36	11
37-39	12
40-41	13
42-43	14
44-46	15

For larger k, c is the largest integer in $0.5(k-1) - 0.98\sqrt{k}$.

- (2) In the analysis of the variables data, if the operator is introducing small biases in selected items, this may be detected more readily by a test of normality on the difference data than by the variance test. Such a test would tend to indicate that the observed differences come about not as a result of a controlled measurement process, but because of a perturbation introduced in the data. A simple frequency distribution plot of the data may disclose this immediately without applying a formal statistical test. Otherwise, the W-test for normality may be applied. (See Reference 7.)

- (3) In III-B-(3) (a), the variance test is applied only to an individual stratum. The inspector should be alert to the possibility of extracting useful information from the entire set of $\bar{d}_j$ data. For example, information on short-term systematic error variances may be derived from the receipts data if a given stratum or sub-stratum consists of all receipts measured at a given point in time and hence, presumably under the same set of conditions. (See Reference 8.) In a more general sense, because of the wide variety of situations that might be imagined, it is difficult to define statistical estimation procedures that would produce useful information from a given set of $\bar{d}_j$ values. The inspector should, however, be alert to the possibility that the variance between $\bar{d}_j$ values may be of use to him on particular occasions.

Evaluation of MUF

The inspection effort is directed ultimately at determining the significance of the operator's MUF for the material balance period in question. There are two approaches that might be used. On the one hand, the inspection data may be used to verify that the operator's data are valid and then, having been verified, the operator's data are used in the MUF calculations with the inspector's data playing no further role. This is called the MUF approach. On the other hand, the inspector's data may be regarded as a body of data whose role is to correct or adjust the operator's data for biases, regardless of whether such biases are judged to be significant or not. This is called the $\hat{D} + \text{MUF}$ approach. With both approaches, it is assumed that the incidence of gross discrepancies has been found to be acceptably low. Otherwise, there is little point in proceeding with the MUF evaluation.

Before giving the details of both approaches, their advantages and disadvantages are cited. The MUF approach will normally be more sensitive in detecting a given real or true MUF because the variance of MUF is in general smaller than the variance of $(\hat{D} + \text{MUF})$. This is so because the random error variances of the inspector affect $(\hat{D} + \text{MUF})$, but of course, they do not affect MUF. Since the inspector normally only performs measurements on a sample basis, with the sample representing a small fraction of the total number of items, this inflation in the random error variance of the statistic can be quite large. This advantage of the MUF approach may not be as great as it may appear at first glance, however, because the uncertainty in the MUF statistic is in many cases dominated by systematic error variances rather than by random error variances.

Consider the advantages of the $(\hat{D} + \text{MUF})$ approach. First, the use of this statistic eliminates the effect of biases introduced by the operator to obscure diversions. In mathematical terms, the expected value of $(\hat{D} + \text{MUF})$ is the true MUF, while the expected value of MUF is the true MUF plus operator induced biases.

As a second advantage, the variance of $(\hat{D} + \text{MUF})$ is not affected by the systematic error variances of the operator. This is a decided practical advantage because on the one hand, it may prove to be difficult on occasion to obtain from the operator his assessment of the systematic error variances. On the other hand, the intelligent operator bent on diversion will inflate these variances to reduce detection probabilities since he recognizes that the inspector will have a difficult time demonstrating that his systematic error variances are not realistic. Anything that can be done to reduce the reliance on the operator's data and information works to the advantage of the inspector, and the use of the $(\hat{D} + \text{MUF})$ statistic works in that direction.

In a summary evaluation of the two approaches, it would appear that the MUF approach is superior in an "honest" environment. However, since Agency safeguards practices are by their very nature designed to combat diversion attempts, it can scarcely be said that there is an honest environment. Therefore, the authors advocate the use of the ($\hat{D}$ + MUF) approach, although both approaches are discussed in the following sections.

Operator's MUF (MUF approach)

With the MUF approach, it is assumed that the operator's data have been verified, and therefore, they are used in determining MUF. For the error variances, the data in question are the design data, i.e., the pre-inspection values submitted by the operator. If adjustments are made to some of these variances, it is assumed that they will have minor impact on the variance of MUF, and such adjusted variances are also treated as known constants.

The approach to MUF evaluation will involve the construction of a confidence interval on the true MUF. This requires that the variance of MUF be computed. This can be done for element weight using the rules given in Table 6.2 of Reference 2. These rules are simplified and restated below where the simplification for purposes of this discussion consists in ignoring the effects of short-term systematic errors, and in combining the effects of sampling and analytical.

1. Rules for Calculating the Variance of MUF

Rule (1) Systematic Error in Weighing

Sum the element weights for all items weighed on scale i . Square this sum. Multiply by the systematic error variance for scale i , and sum over all scales.

Rule (3) Random Error in Weighing

Square each element weight for items weighed on scale i . Sum these squares over all items. Multiply the results by the random error variance for scale i , and sum over all scales.

Rules (4) and (7) Systematic Error in Sampling and Analysis

Sum the element weights for all items with element factors based on a given method of sampling and analysis. Square this sum. Multiply by the systematic error variance for the method in question, and sum over all methods.

Rules (6) and (9) Random Error in Sampling and Analysis

Sum the element weights for all items that have a common element factor based on a given method of sampling and analysis. Square this sum. Divide by the number of samples (analyses) on which each factor is based, and sum over all factors. Multiply the result by the random error variance for the method in question, and sum over all methods.

(The user should be alerted to the fact that Rules (1)-(9) of Reference 2, and repeated above, apply to variances expressed on a relative basis, i.e., as squared coefficients of variation. The result is on an absolute basis, i.e., in terms of variance of MUF.)

When NDA methods are used in one or more of the flow or inventory strata, the above rules do not apply. See Section 3.3.8 or 3.3.10 of Reference 2 in this connection. Assuming for simplicity that all items measured have the same element weight, that the systematic error variance has been calculated for that weight, and that the random error variance is given, then for k items in the stratum sum, the systematic error variance of the sum is k^2 times the systematic error variance for a given item, and the random error variance of the sum

variance of $(\hat{D} + MUF)$, and then constructing the confidence interval of the form $(\hat{D} + MUF) \pm k\sigma_{(\hat{D} + MUF)}$. The only remaining problem, therefore, is to give rules for calculating the variance of $(\hat{D} + MUF)$. It turns out that this is a simple exercise using the results already calculated. The key equation is

$$\text{variance } (\hat{D} + MUF) = \text{variance } (\hat{D}) + \text{variance } (MUF).$$

The rules for calculating variance $(\hat{D})$ were given previously in this Section, pages 52-55, while those for calculating variance (MUF) were given on pages 60-62. Thus, variance $(\hat{D} + MUF)$ is readily calculated.

D. Application to a Uranium Fabrication Plant

Introduction

In this section, the statistical concepts described earlier are applied to a uranium fabrication plant. The example begins with a description of the hypothetical plant and the pertinent design information. The design information includes an accountability model for the facility and the operator's estimates of his measurement uncertainties.

From the design information for the facility and assumed measurement capabilities for the Agency, concrete examples are developed. These are presented in the following order:

- Error propagation for σ_{MUF} .
- Error propagation for $\sigma_{\hat{D}}$.
- Sample sizes for attributes and variables measurements and for the variance test.
- Power of detection for verification, verification and MUF evaluation, and the $(\hat{D} + \text{MUF})$ approach.

To illustrate the relationship between M , the threshold quantity, and the power of detection, three values of M are used in the examples.

Facility Definition

The hypothetical plant is a low-enriched uranium (3 percent) fuel fabrication plant processing 240 MTU per year from UO_2 powder to pellets for loading into fuel rods. The assumed flow sheet is given in Figure 1. Six percent of the feed is assumed to be scrap that is processed and recovered internal to the MBA. About 5.5 percent of the feed in the form of recovered scrap is recycled leaving 0.5 percent solid waste as a measured discard. Laboratory samples are also recycled.

The plant consists of one material balance area (MBA). The flow and inventory key measurement points (KMP) are:

KMP 1 - UO_2 powder feed (20 kg U per container)

KMP 2 - UO_2 pellet product (5 kg U per container) based on measurements made at inspection and rod loading station and item counting of rods and bundles.

KMP 3 - Measured discards (0.433 kg U per drum) as solid waste.

KMP 4 - Inventory locations for each type of material. (4 kg U per container.)

Unmeasured losses and hidden inventory accumulation are assumed to be negligible so that the expected value of material unaccounted for (MUF) would be zero.

A batch or lot of nuclear material is assumed to be approximately 3,000 kg U, except for the measured discards and inventory. The approximate number of items per batch is given in Table 2 together with average item weights, element weights, uranium factors and the number of samples taken per batch for analytical determination of the uranium factor.

An annual physical inventory is determined by clean-out of the process holdup. The accumulated scrap powder, pellets and grinder sludge are divided into batches according to material type. Each batch is sampled and analyzed to determine an average element factor for each batch and each item is weighed. All batches are assumed to be recycled during the next material balance period. All feed, product and waste material is shipped out of the MBA to the customer or to a central storage MBA before the material balance is closed. The physical inventory on hand is assumed to be three percent of the feed of which one percent is recycled scrap and two percent is the normal process holdup.

The measurement methods used by the operator and the associated random and systematic error standard deviations are given in Table 3. The random and systematic error standard deviations due to sampling and analysis to determine the element factors are combined in Table 3. It is assumed the operators will normally maintain a quality

control program that is designed to estimate periodically, for each stratum, the analytical error, sampling error and process variation from batch-to-batch by application of and analysis-of-variance methods. The symbols σ_{Os} and σ_{Or} denote the respective absolute systematic and random error standard deviations of the operator and δ_{Os} and δ_{Or} denote the corresponding relative systematic and random error coefficients of variation: $\delta = \sigma/\mu$. The data in Table 3 are based on Reference 2 and represent actual experience which is a high level of performance when compared to Reference 4.

The standard deviation of MUF (σ_{MUF}) obtainable by the operator can be calculated from the data of Table 3 using the rules given in Section C. These rules are summarized in Table 4 with the parameter values of Table 3 indicated. The results of applying these values are summarized in Table 5. In stratum 4, nondestructive assay is used and this is treated as a special case. The stated σ_{MUF} for the hypothetical plant is 231 kg of U which is about 0.10 percent of the annual feed. This is a very good level of accountability. Other plants may have larger values for the stated σ_{MUF} . As a result, it is advisable to consider levels of accountability for this type of plant which are based on international standards of accountability ($ISOA\sigma_{MUF}$) as well as the stated σ_{MUF} for this particular example.

Evaluation of Design Information

During the planning stage, a review and evaluation of the design information is made. The objective of this review is to determine the level of accountability of the plant and to establish a quantitative basis for the planning of inspection activities. Those parts of the planning activity which relate to the example application include:

Figure 1 Flowsheet of the UO₂ Fabrication Operations

TABLE 2 Design Information - Expected Annual Material Balance Data

	2	3	4	1,5
	Feed	Product	Measured	Physical
	UO ₂	UO ₂	Discards	Inventory
	Powder	Pellets	Waste	Scrap UO ₂
Expected total element weight, T _i (kg U)	240,000	238,800	1,200	7,200
Percent of feed	100	99.5	0.5	3.0
Nominal batch size (kg U)	3,000	3,000	---	1,200
Number of batches, P _i	80	80	1	6
Average element weight, $\bar{X}_i$ (kg U)	20	5	0.433	4
Number items (weighings)/batch, Q _{ij}	150	597	2,770	300
Number items (weighings), N _i = P _i Q _i	12,000	47,760	2,770	1,800
Average item weight, $\bar{W}_i$ (kg UO ₂)	22.8	5.7	---	6
Average element factor, $\bar{U}_i$	0.876	0.881	---	0.67
Number samples (analyses)/batch, L _{ij}	3	5	2,770	5

$$\text{MUF} = T_1 + T_2 - T_3 - T_4 - T_5 = 7,200 + 240,000 - 238,800 - 1,200 - 7,200 = 0 \text{ kg U}$$

TABLE 3 Design Information - Operator Measurement Error Data

Stratum	Material	Measurement Averages	Systematic ⁽¹⁾		Random	
			$\frac{\sigma_{Os}}{(\text{abs})}$	$\frac{\delta_{Os}}{(\text{rel})}$	$\frac{\sigma_{Or}}{(\text{abs})}$	$\frac{\delta_{Or}}{(\text{rel})}$
2	<u>Feed (UO₂ powder)</u>					
	Scale #1, N ₂ = 12,000 weighings	22.8 kg UO ₂	0.010	0.00044	0.015	0.00066
	U-factor #1, P ₂ L ₂ = 240 samples	<u>0.876</u>	0.0005	0.00057	0.0006	0.00068
	Element weight	20.0 kg U				
3	<u>Product (UO₂ Pellets)</u>					
	Scale #2, N ₃ = 47,760 weighings	5.7 kg UO ₂	0.001	0.000175	0.005	0.000877
	U-factor #2, P ₃ L ₃ = 400 samples	<u>0.881</u>	0.0003	0.00034	0.0005	0.00057
	Element weight	5.0 kg U				
4	<u>Measured Discards (solid waste)</u>					
	Element weight, N ₄ = 2,770 NDA	0.433 kg U	0.0173	0.04	0.0433	0.10
1,5	<u>Physical Inventory (Scrap UO₂)</u>					
	Scale #3, N _i = 1,800 weighings	6.0 kg UO ₂	0.010	0.00167	0.015	0.0025
	U-factor #3, P _i L _i = 30 samples	<u>0.670</u>	0.006	0.0090	0.022	0.033
	Element weight	4.0 kg U				

(1) Absolute units.

Relative fractional coefficient of variation $\delta = \frac{\sigma}{u}$

TABLE 4 Rules for Finding the Coefficients Needed in Calculating the Variance of the Algebraic Sum of Element Weights. (See Table 6-2 of Reference 2)

Notation (See also Table 2)

$i = 1, 2, \dots, R$ strata ($R = 5$ in example)

$j = 1, 2, \dots, P_i$ batches ($P_1=6, P_2=80, P_3=80, P_4=1, P_5=6$)

$k = 1, 2, \dots, Q_{ij}$ items per batch

$Q_{1j} = 300$ for all j batches $= Q_1$

$Q_{2j} = 150$ " " " " $= Q_2$

$Q_{3j} = 597$ " " " " $= Q_3$

$Q_{4j} = 2770$ " " " " $= Q_4$

$Q_{5j} = 300$ " " " " $= Q_5$

N_i = Total number of items in stratum i

$= P_i Q_i$ ($N_1 = 1800, N_2 = 12,000, N_3 = 47,760,$
 $N_4 = 2770, N_5 = 1800$)

L_i = Number of samples (analyses) per batch in stratum i
 by operator (same for all batches in a stratum)

($L_1 = 5, L_2 = 3, L_3 = 5, L_4 = 2770, L_5 = 5$)

$\bar{U}_i$ = Average element factor for given batch in stratum i
 (same for all batches in stratum)

($\bar{U}_1 = 0.67, \bar{U}_2 = 0.867, \bar{U}_3 = 0.881, \text{---}, \bar{U}_5 = 0.67$)

$\bar{X}_i$ = Average element weight per item in stratum i
 (same for all items in stratum)

($\bar{X}_1 = 4, \bar{X}_2 = 20, \bar{X}_3 = 5, \bar{X}_4 = 0.433, \bar{X}_5 = 4$)

T_i = Total element weight in stratum i

$= N_i \bar{X}_i$

($T_1 = 7200, T_2 = 240,000, T_3 = 238,800$

$T_4 = 1200, T_5 = 7200$)

Rules of Table 6-2, Reference 2 (See also Section C)

(In this example, it is assumed that short-term systematic error variances are zero)

Rule 1. Systematic error in weighing (σ_{sw}).

(Assume 1 scale per stratum; also assume that the scale used for inventory measurements is recalibrated between inventories.)

TABLE 4 (Contd.)

"Sum the element weights for all items weighed on scale i. Square this sum."

$$C_{sw(i)} = T_i^2 (C_{sw(1)} = 51.84 \times 10^6, C_{sw(2)} = 5.76 \times 10^{10}, C_{sw(3)} = 5.703 \times 10^{10}, \text{---} \\ C_{sw(5)} = 51.84 \times 10^6)$$

Rule 3. Random error in weighing (σ_{rw}).

"Square each element weight for items weighed on scale i Sum the terms ... over all items."

$$C_{rw(i)} = N_i \bar{X}_i^2 (C_{rw(1)} = 2.88 \times 10^4, C_{rw(2)} = 4.80 \times 10^6, C_{rw(3)} = 1.194 \times 10^6, \text{---} \\ C_{rw(5)} = 2.88 \times 10^4)$$

Rules 4 and 7 Combined. Systematic errors in sampling and Analysis (C_{su})

"Sum the element weights for all items with element factors (based on given method of sampling and analysis) and square this sum."

$$C_{su(i)} = T_i^2 \text{ (Same as rule 1)}$$

Rules 6 and 9 Combined. Random errors in sampling and analysis (C_{ru})

"Sum the element weights for all items that have a common element factor (based on a given method of sampling and analysis). Square each sum. Divide ... by the number of samples on which each factor is based. Sum over (batches)."

$$C_{ru(i)} = P_i (Q_i \bar{X}_i)^2 / L_i = N_i \bar{X}_i^2 / L_i \\ (C_{ru(1)} = 1.728 \times 10^6, C_{ru(2)} = 2.40 \times 10^8, \\ C_{ru(3)} = 1.426 \times 10^8, \text{---} C_{ru(5)} = 1.728 \times 10^6)$$

Note: In combining rules 4 and 7 and rules 6 and 9, it is assumed that a given combination of sampling and analytical methods is used within each stratum. Thus, the error standard deviations in Table 1-2 for uranium factors represent the combined effects of sampling and analytical errors.

TABLE 5 Design Information. Stated Variance Components of MUF (See Tables 3 and 4)

Stratum	Rule	Source	σ_{si}^2		σ_{ri}^2	
			$C_s \sigma_s^2 (\text{kgs})^2$		$C_r \sigma_r^2 (\text{kgs})^2$	
1	1	$(51.84 \times 10^6) (0.00167)^2$	=	144.58		
	4 & 7	$(51.84 \times 10^6) (0.0090)^2$	=	4199.00		
	3	$(2.88 \times 10^4) (0.0025)^2$	=		0.18	
	6 & 9	$(1.728 \times 10^6) (0.033)^2$	=		1881.79	
2	1	$(5.76 \times 10^{10}) (0.00044)^2$	=	11151.36		
	4 & 7	$(5.76 \times 10^{10}) (0.00057)^2$	=	18714.24		
	3	$(4.80 \times 10^6) (0.00066)^2$	=		2.08	
	6 & 9	$(2.40 \times 10^8) (0.00068)^2$	=		110.98	
3	1	$(5.703 \times 10^{10}) (0.000175)^2$	=	1746.40		
	4 & 7	$(5.703 \times 10^{10}) (0.00034)^2$	=	6592.14		
	3	$(1.194 \times 10^6) (0.000877)^2$	=		0.92	
	6 & 9	$(1.426 \times 10^8) (0.00057)^2$	=		46.32	
4	-	$(1200)^2 (0.04)^2$	=	2301.73		
	-	$2770 (0.433)^2 (0.10)^2$	=		5.19	
5	1	Same as Stratum 1	=	144.58		
	4	" " " "	=	4199.00		
	3	" " " "	=		0.18	
	6	" " " "	=		1881.79	
		$\sigma_{MUF_s}^2$	=	49193.03		
		$\sigma_{MUF_r}^2$	=		3929.43	
		σ_{MUF_s}	=	221.79		
		σ_{MUF_r}	=		62.69	
		σ_{MUF}^2		53122.46 kg ² U		
		σ_{MUF}		230.5 kg U		

- Preparation of a material balance model for the facility,
- Preparation of an error propagation model, and
- Preparation of the measurement error variances and calculations of the operator's σ_{MUF} .

Examples of this type of planning activity were shown previously in Tables 2, 3, 4, and 5. Table 2 shows the material balance model. Table 3 shows the operator's measurement error data and Table 4 shows the error propagation model and rules of error propagation. Table 5 shows the propagated measurement error variances for all strata and the material balance as a whole.

The standard deviation of MUF as shown in Table 5 is 231 kg U. This is a very good level of accountability. For purposes of comparison, an international standard of accountability ($\text{ISOA}\sigma_{\text{MUF}}$) for σ_{MUF} of 0.3 percent of annual feed or 720 kg U is assumed.

Using this assumed standard, the plant material accounting system readily meets the criterion for levels of accountability.

Agency Measurements

The next step in the planning stage is the establishment of a measurement plan for independent measurements by the Agency. Two basic measurement methods are assumed for the inspection activity - an attribute method and a variables method. For the attribute measurement, a non-destructive assay device is assumed. The device is assumed to have the capability of measuring the total ^{235}U of inventory, feed, and product items with a standard deviation of 5 percent relative. For waste items the inspector's device is assumed to have the same measurement uncertainties as the operator's devices. For the variables measurement it

is assumed that the inspector uses the same methods and has the same measurement uncertainties as does the operator.

From a detection standpoint, the assumption of a non-destructive device for the attributes measurement is not absolutely necessary. Composite sampling techniques and destructive methods of analysis could provide the same detection sensitivity. However, the nondestructive approach is preferred if a suitable instrument is available. In this case, further testing could be done immediately if defects are found. By contrast, the time delay due to sample transit and analysis would preclude such immediate steps when conventional methods are used.

Estimation of Verification Measurement Errors

Estimation of Agency measurement errors is the next item of concern in the planning stage. In the example, the nondestructive assay instrument used for attribute measurement is assumed to have a standard deviation of 5 percent relative to the item amount. It is further assumed that this error is the combined random and systematic error or the error which corresponds to a single measurement.

For purposes of establishing a conservative sample size (at least enough samples), the detection limit of the attribute instrument is taken as 6 times the standard deviation or in this case 30 percent of the item amount.

The variables measurement capability needs to be considered from two standpoints. First, is for the use of the variables measurement when used in the attribute mode as a single measurement test. Second, is for the use of the variables measurements in bias detection when sums of differences are compared.

When used in the attributes mode the variable measurements must detect, with nearly 100 percent certainty by a single measurement, medium-sized defects which are below the detection limit of the attributes measurement. To

test whether or not this requirement can be met, the combined (random plus systematic) standard deviation of a single difference measurement is computed. This value is then compared to the value taken as the lower limit for the attribute measurement. In the example, the variable measurement uncertainty of the inspector is assumed to be the same as that of the operator. Thus, the standard deviations for the difference statistic are simply $\sqrt{2}$ times the operator's standard deviations. The test of adequacy consists simply of comparing the standard deviation of the difference statistic to the crossover point, e.g., to 30 percent of the item value. This comparison is shown in Table 6 for the example case. As shown by the data in the table, the ratio of the assumed defect size to the standard deviation of the difference statistic is quite large for all strata except the stratum of waste drums. Thus, all medium sized defects would be detected nearly 100 percent of the time if included in the random sample (except for waste drums).

The waste drums are a special case in that there is only one measurement method - the nondestructive assay instrument. This situation is discussed later.

Also shown in the table are 4σ values for a single difference. These values represent the reject limits for a single difference measurement. Thus, all removals or falsifications above this level would be detected with high probability by a single measurement with the variable instrument. This point illustrates the fact that the operator's best strategy for obtaining M by the large defect approach is to remove material or falsify data in amounts just below 30 percent of the item values or in amounts equivalent to 100 percent of the item value. This strategy presents, for a given M, the fewest number of defects available for detection by either method .

TABLE 6 Single Measurement Standard Deviations for the Difference Statistic

Stratum	Combined σ ⁽¹⁾ Rel. %	4σ Rel. %	Ratio of 30% Bias to σ
1 Inventory	3.1	12.4	9.7
2 Feed	0.15	0.60	200
3 Product	0.14	0.56	214
4 Waste	15.2	60.8	1.97
5 Inventory	3.1	12.4	9.7

- (1) Combined random and systematic error for a single paired comparison for total U. U factor based on two inspector assays and operator replicate assays as shown in Table 4.

The next step on the planning stage is to estimate the variance of the difference statistic, $\hat{D}$, when considered as a sum of differences taken over a single stratum or over the whole material balance.

The error propagation procedure is similar to that used in calculating the operator's σ_{MUF} . Under the assumptions taken for the example of identical uncertainties for Agency and operator, the systematic error variance for $\hat{D}$ is simply two times the operator's value. The random error propagation procedure is more complicated since it involves n , the number of variable samples. It is also complicated by the use of the same uranium assay factor for a number of items. Thus, the random error variance of the difference statistic for the assay factor reduces with increasing n in a nonuniform way. The error propagation formulas for the random error variance of $\hat{D}$ are given in Table 7 along with the calculated systematic error variances.

The formulas given in Table 7 are given in simplified form for the special case where the operators and inspectors' measurement uncertainties are the same. They are given using relative errors to show the relationship to Table 5 for σ_{MUF} .

TABLE 7 Variance Components of the Difference Statistic, $\hat{D}$.

Stratum	Systematic Error Variance, $\sigma_{\hat{D}_s}^2$ (kgs U) ²		
1 - Inventory	2(4343.6)	=	8687.2
2 - Feed	2(29865.6)	=	59731.2
3 - Product	2(8338.5)	=	16677.0
4 - Waste	2(2301.7)	=	4603.4
5 - Inventory	2(4343.6)	=	8687.2
	TOTAL $\sigma_{\hat{D}_s}^2$	=	98386 (kgs U) ²

$$\sigma_{\hat{D}_s} = 313.7 \text{ kgs U}$$

Stratum	Random Error Variance $\sigma_{\hat{D}_r}^2$ (kgs U) ²		
---------	---	--	--

$$\begin{aligned}
 1 - \text{Inventory} \quad \sigma_{\hat{D}_{r1}}^2 &= \frac{1.037^{(1)} \times 10^8}{n_1} (0.0025)^2 \\
 &+ 5.184 \times 10^7 \left[\frac{(0.033)^2}{5n_1^*} + \frac{(0.033)^2}{2n_1} \right] \\
 &\text{where } n_1^{*(2)} \leq 6
 \end{aligned}$$

$$\begin{aligned}
 2 - \text{Feed} \quad \sigma_{\hat{D}_{r2}}^2 &= \frac{1.152 \times 10^{11}}{n_2} (0.00066)^2^{(3)} \\
 &+ 5.76 \times 10^{10} \left[\frac{(0.00068)^2}{3n_2^*} + \frac{(0.00068)^2}{2n_2} \right] \\
 &\text{where } n_2^* \leq 80
 \end{aligned}$$

(1) The Coefficients for the random error variance terms are calculated from the values for N_i and $\bar{X}_i$ given in Table 3. The first coefficient in each expression is $2 N_i^2 \bar{X}_i^2$ and the second is $N_i^2 \bar{X}_i^2$, e.g., stratum 1: $(1800)^2 (4)^2 = 5.184 \times 10^7$.

TABLE 7 (Contd.)

Stratum	Random Error Variance $\sigma_{\hat{Dr}}^2$ (kgs U) ²
3 - Product	$\sigma_{\hat{Dr}_3}^2 = \frac{1.141 \times 10^{11}}{n_3} (0.000877)^2$ $+ 5.703 \times 10^{10} \left[\frac{(0.00057)^2}{5n_3^*} + \frac{(0.00057)^2}{2n_3} \right]$ <p style="text-align: center;">where $n_3^* \leq 80$</p>
4 - Waste	$\sigma_{\hat{Dr}_3}^2 = \frac{2.88 \times 10^6}{n_4} (0.10)^2$
5 - Inventory	Same as 1.

- (2) n_1, n_2, n_3, n_4, n_5 are the inspection sample sizes for variables measurements on strata 1, 2, 3, 4 and 5. The n^* is used to denote the number of batches in a stratum. Because of the operator's use of an average U factor for each batch, the random error part of $\hat{Dr}$ for the operator's U assay does not reduce after n^* is equal to the number of batches or after each batch has been sampled once. For the inspector, each sample is assayed in duplicate except for stratum 4 which is done by nondestructive assay.
- (3) The error propagation is done using the relative individual errors given in Table 3 to show correspondence to Table 5 for the error propagation for σ_{MUF} . A detailed procedure using absolute errors is given in Section C.

For more general use, the procedure given in Section C for the error propagation of α_D^2 should be used which uses absolute errors.

The random error variance formulas given in Table 7 require further explanation. In the example, the operator uses the same uranium assay factor for all the items in a batch for strata 1, 2, 3 and 5. For example, in stratum 2 (feed) three assays are done on each batch by the operator. The material is assumed to be uniform with no sampling error. The average of the three assays is applied to all 300 items in the batch. Thus the operator's random error of assay which is part of the difference statistic does not reduce with further inspector samples within a given batch. It reduces, of course, between batches. Thus, the random error of $\hat{Dr}_i$ due to uranium assay reduces within increasing sample size only until all batches have been sampled. Further sampling after that point reduces only the inspector's assay error. The random error due to weighing, however, does reduce with increasing sample size in a uniform way, since each item is weighed.

Inspection Sample Sizes

As was noted earlier, the determination of inspection sample sizes involves four separate calculational procedures. For each stratum these are:

1. Calculation of a sample size for the attribute measurement, n_a ,
2. Calculation of a sample size for use of the variable instrument in the attribute mode, n_{v1} ,
3. Calculation of a variable sample size for detecting a bias of size M taken over the entire material balance, n_{v2} , and
4. Determining (from tables of the Chi-square test) a variable sample size for the variance test, n_{v3} .

For actual implementation, only two sample sizes are used. One for the attributes measurement, n_a , and one, n_v , for the variables method. That variables sample size is the largest of the three sample sizes indicated above.

Attribute Method Sample Size

The attribute sample size, n_a , is designed such as to include, with high probability in the random sample, at least one gross defect. Here, it is assumed that the operator could accumulate a quantity M by removals (or falsifications) each of which is equal to the average amount of material in an item in the stratum. The removals (or falsifications) which accumulate to M could come from a single stratum, or be spread out over several or all strata. The procedure is as follows:

1. Determine for each stratum the number of whole items (or equivalent falsification) which would have to be removed to equal M. This number of defects is denoted by r_1 where $r_1 = M/\bar{X}_i$ and $\bar{X}_i$ is the average item amount in total U.
2. Compute the sample size for each stratum by the binomial approximation as follows:

$$n_a = N_i \left[1 - (0.05)^{\frac{1}{r_1}} \right]$$

where: N_i = the number of items in the stratum
 $0.05 = \beta$ = the probability of not including in the random sample at least one defect where there are r_1 defects among the N_i items.

It should be noted the above procedure is valid only for stratum of fairly uniform item amounts. When the item amounts differ by about 30 percent or more, more uniform substrata should be formed. The substrata are then treated exactly as before using the same value for M

and the new values for $\bar{X}_i$. Sample sizes for three values of M are given in Table 8.

Variable Sample Size for Crossover Point

The first sample size for the variable method, n_{v1} , is to protect against an accumulation of M by medium sized defects not detected by the attribute measurement. In the example, the crossover point for the attribute method was taken as 30 percent of the item amount. The procedure is identical to the one used for n_a except the number of defects, r_2 , is now 3.33 times larger. The formula is as follows:

$$n_{v1} = N_i \left[1 - (0.05)^{\frac{1}{r_2}} \right]$$

$$\text{where } r_2 = \frac{M}{0.3\bar{X}_i}$$

and N_i and $0.05 = \beta$ are as before.

Sample sizes for three values of M are given in Table 8.

Variable Sample Size for Bias Detection

The second variable sample size, n_{v2} , is designed so as to detect over the whole material balance a bias of size M in the presence of systematic error. The difference statistic is used in this case to detect such a bias. The calculational procedure is a three step process. First, a goal for the random error standard deviation of $\hat{D}$ is determined for detecting a bias M in the presence of systematic error. That value for $\sigma_{\hat{D}}^2$ random is then used in the minimum variance formula to determine a sample size for the whole material balance. Allocation is then made to the individual strata to obtain n_{v2} .

The procedure used to estimate the desired value for $\sigma_{\hat{D}}^2$ random is complicated by the presence of systematic error. For small values of M, the desired probability of detection often cannot be reached even with large sample

TABLE 8 Sample Sizes for Attribute Type Measurements

Stratum	N_i	$\bar{X}_i$, kgs	M, kgs Uranium							
			833				1140			
			r_1	n_a	r_2	n_{v1}	r_1	n_a	r_2	n_{v1}
1 Inventory	1800	4	208	26	694	8	285	19	950	6
2 Feed	12000	20	42	826	139	256	57	614	190	188
3 Product	47760	5	167	849	555	257	228	623	760	188
4 Waste	2770	0.433	1924	4	(1)	--	2633	3	--	--
5 Inventory	1800	4	208	26	694	8	285	19	950	6

83 Stratum	M = 2370 kgs Uranium						
	r_i	n_a	r_2	n_{v1}			
1	1800	4	593	9	1800	3	
2	12000	20	119	298	395	91	
3	47760	5	474	301	1580	91	
4	2770	0.433	2770	3	--	--	
5	1800	4	593	9	1800	3	

(1) For waste drums, it is assumed that the operator and inspector both use a nondestructive assay instrument for the variable measurement. This sample size is treated as a special case in the text.

sizes because of the dominance of systematic error. In this case, the point of diminishing return is taken as the point where the random error standard deviation is equal to one-half the standard deviation of the systematic error component of $\hat{D}$. For larger values of M , the desired probability for detection of bias can often be obtained with larger values $\left(\sigma_{\hat{D}}^2 \text{ random} > \frac{1}{2} \sigma_{\hat{D}}^2 \text{ systematic}\right)$ for the random error component.

In this latter case, the goal value for the random error standard deviation of $\hat{D}$ for use in the sample size determination may be estimated with the following formula:

$$z_{\beta} = -z_{1-\beta} = \frac{z_{1-\alpha} \sqrt{\sigma_{\hat{D}s}^2 + \sigma_{\hat{D}r|H_0}^2} - M}{\sqrt{\sigma_{\hat{D}s}^2 + \sigma_{\hat{D}r|H_1}^2}}$$

where $\sigma_{\hat{D}s}^2$ = systematic error variance of $\hat{D}$
 $\sigma_{\hat{D}r|H_0}^2$ = random error variance goal of $\hat{D}$ under the null hypothesis
 $\sigma_{\hat{D}r|H_1}^2$ = random error variance goal of $\hat{D}$ under the alternative hypothesis.

For the sample size determination, $\sigma_{\hat{D}r|H_1}^2$ is taken as $4\sigma_{\hat{D}r|H_0}^2$. This is to ensure that at least 1 enough samples are taken. Using the relationship between the alternative and null hypothesis, the desired value of $\sigma_{\hat{D}r|H_1}$ can be obtained using the above equation. Trial and error methods, graphical methods, or Newton's method may be used to solve the equation. Results are given in Table 9 for the three values of M used in the example.

Having obtained the value of $\sigma_{\hat{D}r|H_1}$ which corresponds to either $\frac{1}{2}\sigma_{\hat{D}s}$ or, if larger, the desired probability of detection, the sample size is then calculated for the whole material balance. The minimum variance formula of Stewart

is as follows:

$$n = \frac{4 \left(\sum N_i \sigma_{dr_i} \right)^2}{\sigma_{\hat{D}r|H_1}^2}$$

where σ_{dr_i} = the random error standard deviation of an individual difference.

4 = random error variance factor for H_1 .

Allocation of the total sample size to each stratum is given by:

$$n_i = n \left(\frac{N_i \sigma_{dr_i}}{\sum N_i \sigma_{dr_i}} \right)$$

Sample sizes and completed calculations for total n_1 stratum n_i , and $\sigma_{\hat{D}r|H_1}^2$ are given in Table 9.

The values for n_{v2} given in Table 9, represent in an approximate way the sample sizes which minimize the random error variance of $\hat{D}$ for a given total n . The minimum variance formulation assumes that the random error variances of the individual differences reduce in a uniform way with increasing sample size. This condition is true as long as the sample size does not exceed the number of batches in a stratum because of the use of a uranium assay factor. In the example, only the inventory stratum batch limit is exceeded, e.g., $n_{v2} > 6$. For sample size estimates, this effect can be ignored. However, in computing the actual variance of $\hat{D}$ for the probability of detection, the exact formulas given in Table 7 are used.

Sample Size for Variance Test

The sample size for the variance test is chosen to detect with high probability a ratio of observed random variance to design variance of 4. For α and $\beta = 0.05$, the desired sample size is 13 samples per strata for variables measurements. This sample size is denoted by n_{v3} .

Special Case for Waste Drums

As was noted earlier, only one measurement method is applied to the waste stratum. This is a nondestructive assay method having a random error standard deviation of 10 percent of the item amount and a long-term systematic error standard deviation of 4 percent. In the example, the inspector is assumed to use a similar method such that the difference statistic has a combined random plus systematic error standard deviation of about 15 percent. Thus, the crossover point for this stratum would be 6×15 or 90 percent of an item amount. For this reason, no sample size is given for, n_{v1} , the detection of medium-sized defects. However, the large error itself makes the sample size for bias detection sufficiently large so as to include with high probability any medium sized defects should that diversion route be chosen.

Sample Sizes for Implementation

For actual implementation only two sample sizes are used - one for attributes measurement and one for variables measurements. The final sample sizes for implementation are given in Table 10. As was noted earlier, n_v the variable sample size for the stratum is the largest of n_{v1} , n_{v2} , and n_{v3} . The random error variances of the difference statistic under the stated variable sample sizes are also given.

The values for the random error standard deviation of $\hat{D}$ were calculated using the formulas given in Table 7 and the sample sizes shown in Table 10. As an aid to the reader, a summary of the sample size selection process is given next.

TABLE 9 Variable Sample Size for Detection of Bias

Stratum	M, kgs Uranium					
	833		1140		2370	
	N_i	σ_{dr_i} kgs U	$N_i \sigma_{dr_i}$	$\hat{\sigma}_{Dr H_1}^{(1)}$ n total	n_{v2}	$\hat{\sigma}_{Dr H_1}$ n total
1 Inventory	1800	0.111	200	157	214	37
2 Feed	12000	0.022	264		188	149
3 Product	47760	0.0066	315		26	854
4 Waste	2770	0.061	170		26	34
5 Inventory	1800	0.111	200		40	22
					22	1
					26	1
SUM			1149			

87

Example for M = 833

$$n = \frac{4 \left(\sum N_i \sigma_{dr_i} \right)^2}{\sigma_{Dr|H_1}^2} = \frac{4 (1149)^2}{(157)^2} = 214$$

$$n_{v2} = n_1 = 214 \left(\frac{200}{1149} \right) = 37$$

$$n_{v2} = n_2 = 214 \left(\frac{264}{1149} \right) = 49$$

(1) The random error goal ($\hat{\sigma}_{Dr|H_1}$) to be achieved by the random size n.

$\hat{\sigma}_{Dr|H_1}$ For M = 1140 kgs

$$\frac{1140}{1.645} = \sqrt{(313.7)^2 + (188)^2} + \sqrt{(313.7)^2 + (94)^2}$$

$$693 = 365.7 + 327.5$$

$$693 = 693.2$$

where 313.7 = $\hat{\sigma}_D$ systematic

$$\frac{\hat{\sigma}_{Dr|H_1}}{\hat{\sigma}_{Dr|H_0}} = 2, -Z_{1-\beta} = -1.645$$

$$Z_{1-\alpha} = 1.645$$

TABLE 10 Inspection Sample Sizes for Implementation

Stratum	M, kgs Uranium							
	833				1140			
	n_a	n_v	$\sigma_{\hat{D}r H_O}^{(1)}$	n_a	n_v	$\sigma_{\hat{D}r H_O}$	n_a	n_v
1 Inventory	26	37	83.7 (kgs U)	19	26	91.1 (kgs U)	9	13
2 Feed	826	256		614	188		298	91
3 Product	849	257		623	188		301	91
4 Waste	4	32		3	22		3	13
5 Inventory	26	37		19	26		9	13

(1) Values obtained by using error propagation formulas of Table 7 and the sample sizes shown in the column under n_v .

Summary of Sample Selection Procedure

Attribute Sample Size, n_a

Step 1: Compute for each stratum the ratio $\frac{M}{\bar{X}_i}$, this ratio is r_1 , the number of whole $\bar{X}_i$ items which would have to be diverted to acquire an amount M.

Step 2: Compute the sample size for each stratum by:

$$n_{a_i} = N_i \left[1 - (0.05)^{\frac{1}{r_1}} \right]$$

Example: $M = 1140$, $N_i = 1800$, $\bar{X}_i = 4$

Step 1: $r_1 = \frac{1140}{4} = 285$

Step 2: $n_{a_1} = 1800 \left[1 - (0.05)^{\frac{1}{285}} \right]$

$$n_{a_1} = 18.8 = 19$$

Repeat for n_{a_2} , n_{a_3} , etc., for all stratum.
See Table 8.

Variable Sample Size for Medium-Sized Defects, n_{v1}

Step 1: Compute for each stratum the ratio of $\frac{M}{\gamma \bar{X}_i}$. This ratio is r_2 , the number of fractional falsifications of size $\gamma \bar{X}_i$ needed to accumulate to an amount M.

Step 2: Compute the sample size for each stratum by:

$$n_{v1_i} = N_i \left[1 - (0.05)^{\frac{1}{r_2}} \right]$$

Example: $M = 1140$, $N_i = 1800$, $\bar{X}_i = 4$, $\gamma = 0.3$

Step 1: $r_2 = \frac{1140}{(0.3)(4)} = 950$

Step 2: $n_{v1_i} = 1800 \left[1 - (0.05)^{\frac{1}{950}} \right]$

$$n_{v1_i} = 5.6 = 6$$

Repeat for n_{v1_2} , n_{v1_3} , etc. See Table 8.

Variable Sample Size for Bias Detection, n_{v2} .

Step 1: Compute the goal for the random error standard deviation of $\hat{D}$ which must be achieved to detect a bias of M in the presence of systematic error. Let $\alpha = 0.05$ and $\beta = 0.05$. First test the ratio of $\frac{M}{\sigma_{Ds}}$ to see if it is less than 3.53. If $\frac{M}{\sigma_{Ds}}$ is less than 3.53 then a β of 0.05 cannot be achieved* for $\alpha = 0.05$. In this case let the random error goal for $\hat{D}$ be $\sigma_{\hat{D}r|H_1} = 1/2 \sigma_{Ds}$. If $\frac{M}{\sigma_{\hat{D}r|H_1}}$ is larger than 3.53 solve the equation below for $\sigma_{\hat{D}r|H_1}$:

$$\frac{M}{1.645} = \sqrt{\sigma_{Ds}^2 + \sigma_{\hat{D}r|H_1}^2} + \sqrt{\sigma_{Ds}^2 + \left(\frac{\sigma_{\hat{D}r|H_1}}{2}\right)^2}$$

Use the graphical method given in Section B or trial and error.

Step 2: Compute for each stratum the random error standard deviation of an individual difference for amount of element. This is σ_{dr_i} . Multiply by N_i the number of items in a stratum to obtain for each stratum, $N_i \sigma_{dr_i}$. Sum these to obtain: $\sum N_i \sigma_{dr_i}$. (See page 92)

Step 3: Compute the sample size for the whole material balance using the minimum variance formula as follows:

$$n = \frac{4 \left(\sum N_i \sigma_{dr_i} \right)^2}{\sigma_{\hat{D}r|H_1}^2}$$

where the value for $\sigma_{\hat{D}r|H_1}^2$ was obtained in Step 1.

*Cannot be achieved without reducing the random error below $1/2 \sigma_{Ds}$, specifically: $3.53 = 1.645 \left(\sqrt{1.25} + \sqrt{1.0625} \right)$.

Step 4: Allocate the total sample size n to each stratum as follows:

$$n_{v2_i} = n \frac{(N_i \sigma_{dr_i})}{\sum \sigma N_i \sigma_{dr_i}}$$

Repeat for each stratum to obtain n_{v2_1} , n_{v2_2} , etc.

Example: $M = 1140$, $\hat{\sigma}_{Ds} = 313.7$ Kgs U, and σ_{dr_1} , σ_{dr_2} ... as given in Table 9, and N_1 , N_2 ... as given in Table 9.

Step 1: Test $\frac{M}{\hat{\sigma}_{Ds}} = \frac{1140}{313.7} = 3.63$

This is larger than 3.53, thus solve equation below for $\hat{\sigma}_{Dr|H_1}$ by trial and error:

$$\frac{1140}{1.645} = \sqrt{(313.7)^2 + \hat{\sigma}_{Dr|H_1}^2} + \sqrt{(313.7)^2 + \left(\frac{\hat{\sigma}_{Dr|H_1}}{2}\right)^2}$$

$$693 = \sqrt{(313.7)^2 + (188)^2} + \sqrt{(313.7)^2 + (94)^2}$$

$$693 = 693$$

Thus $\hat{\sigma}_{Dr|H_1} = 188$ and $\hat{\sigma}_{Dr|H_0} = 94$. See Section B for graphical solution.

Step 2: Compute $N_i \sigma_{dr_i}$ and $\sum N_i \sigma_{dr_i}$. See Table 9.

$$N_1 \sigma_{dr_1} = (1800)(0.111) = 200$$

$$N_2 \sigma_{dr_2} = (12000)(0.022) = 264$$

$$N_3 \sigma_{dr_3} = (47760)(0.0066) = 315$$

$$N_4 \sigma_{dr_4} = (2770)(0.061) = 170$$

$$N_5 \sigma_{dr_s} = (1800)(0.111) = 200$$

$$\sum \sigma N_i \sigma_{dr_i} = 1149$$

Step 3: Compute n for MBA by:

$$n = \frac{4(1149)^2}{(188)^2} = 149$$

Step 4: Allocate n to each stratum by:

$$n_{v2_1} = 149 \frac{200}{1149} = 26$$

$$n_{v2_2} = 149 \frac{264}{1149} = 34$$

$$n_{v2_3} = 149 \frac{315}{1149} = 40$$

$$n_{v2_4} = 149 \frac{170}{1149} = 22$$

$$n_{v2_5} = 149 \frac{200}{1149} = 26$$

Note: σ_{dr_i} may be calculated for the examples using the variance formulas given in Table 7 and substituting in $2(\bar{X}_i)^2$ for the first coefficient and $(\bar{X}_i)^2$ for the second coefficient and letting $N_i = 1$. e.g.,

$$\sigma_{dr_1}^2 = 32 (0.0025)^2 + 16 \left[\frac{(0.033)^2}{5} + \frac{(0.033)^2}{2} \right]$$

$$\sigma_{dr_1}^2 = 0.01236 \text{ (Kgs U)}^2$$

$$\sigma_{dr_1} = 0.111 \text{ Kgs U (See Table 9)}$$

Sample Size for Variance Test, n_{v3}

Step 1: From a table of the Chi-square distribution find for $\alpha = 0.05$ and $\beta = 0.05$, the degrees of freedom which correspond to the ratio value of 4.0, where the ratio of concern is

$$\frac{s_{dr_i}^2}{\sigma_{dr_i}^2} .$$

The degrees of freedom for the ratio of about 4.0 is 12, thus sample size is the degrees of freedom plus one or 13 samples for each stratum.

Variable Sample Size for Implementation, n_v

Step 1: From the values obtained on the same stratum for n_{v1} , n_{v2} , and n_{v3} choose the largest value. That value is n_v , the sample size for the variables method i which is used in implementation in the i th stratum. Repeat for each stratum.

- * From a mathematical viewpoint, it is possible to reduce n_a by an amount n_v since the variables tester will surely detect a large removal. However, from an operational viewpoint, this is not advisable because of the time element. The attributes instrument supplies an immediate indication of a defect, whereas it is assumed that indication of a defect by the variables instrument does not come until some time after completion of the inspection. It is the immediacy of inspection results that make it advisable to use the complete sample size for the attributes tester. The fact that the variables tester also detects a complete removal is factored into the evaluation stage.

Probability of Detection

The final stage in the design stage is to evaluate the over-all power of detection provided by the sampling plans and MUF evaluation. As was noted earlier, there are four probabilities of interest. These are the probability of detecting:

- Gross defects equal to the nominal item value,
- medium-sized defects which cannot be detected by the attribute measurement,
- small biases which by accumulation could disguise diversion, and
- the diversion or loss of material which would only appear as MUF.

In Section B it was shown that the power to detect either gross defects or medium-sized defects by measurements made in the attribute mode can be represented by a single probability of detection. This reduces the problem to three probabilities. These are more easily illustrated by considering the probability of non-detection, Q , where $Q = 1 - P$ and P = the probability of detection. The individual probabilities of nondetection are:

- Q_1 = The probability of not including at least one medium-sized or grossly defective item in the random sample if an amount G was accumulated by such sized defects from a single stratum or partitioned over all strata.
- Q_2 = The probability of not detecting an accumulation of small biases over the entire MBA which could disguise the diversion of an amount D .
- Q_3 = The probability of not detecting the loss or the direct diversion into MUF of an amount MUF (true) where: $G + D + \text{MUF}(\text{true}) = M$.

When the sampling plans for single defects are set at $\beta = 0.05$ for a quantity M , then the probability of not including in the random sample at least one each defect when only an amount G is taken is:

There are several ways in which the Agency may express its power of detection (or nondetection). Firstly, it may wish to express the power of the verification process itself. Secondly, it may wish to consider the power of detecting diversion when an amount M is partitioned between falsifications to disguise diversion and diversion which

Overall Power of Detection

where: MUF_T is the true MUF.

$$Q_3 = \Phi \left[1.645 - \frac{\sigma_{MUF}}{MUF_T} \right]$$

$\alpha = 0.05$ is:

and $\Phi(x)$ is the 100 $\Phi(x)$ percentile value of the value x in the $N(0,1)$ distribution.
Lastly the probability of not detecting the loss or diversion into MUF of a quantity MUF_T rather than M when

$$Q_2 = \Phi \left[\frac{\frac{\sigma_{D|H_1}}{\sqrt{\frac{\sigma_{D_S}^2}{2} + \frac{\sigma_{D_T}^2}{2}}}}{1.645 - \frac{\sigma_{D|H_0}}{\sqrt{\frac{\sigma_{D_S}^2}{2} + \frac{\sigma_{D_T}^2}{2}}}} \right] \quad \text{where } \sigma_{D|H_0} = \sqrt{\frac{\sigma_{D_S}^2}{2} + \frac{\sigma_{D_T}^2}{2}}$$

Similarly the probability of not detecting small biases which accumulate to an amount D rather than M is for $\alpha = 0.05$:

$$Q_1 = (0.05) \frac{M}{G}$$

would appear as MUF. This second approach may in turn be considered from two standpoints - MUF and bias detection separately or MUF + bias detection jointly.

The power of nondetection for the verification process itself is given by the following when $\alpha = 0.05$:

$$C_{\text{verification}} = (0.05)^{\frac{G}{M}} \Phi \left[\frac{1.645 \sigma_{\hat{D}|H_0} - D}{\sigma_{\hat{D}|H_1}} \right]$$

where $G + D = M$.

Since the operator can partition M between gross defects and biases, the Agency wants to express its power of detection for the operator's best strategy or at Q_{max} . The values for C_{max} may be obtained from Stewart's solution given in Section B. Here the value of G for which Q is a maximum is given by:

$$G(\text{for } C_{\text{max}}) = M + \theta \sigma_{\hat{D}|H_1} - Z_{1-\alpha} \sigma_{\hat{D}|H_0}.$$

Where $\theta = R^{-1} \left[\frac{-(\ln 0.05) (\sigma_{\hat{D}|H_1})}{M} \right]$

and R^{-1} is that value of the term in the brackets which corresponds to the θ value for Q_{max} .

Values for $R(\theta)$ and θ are given in Table 11. Values for C_{max} for verification for the three values of M considered here are given in Table 12.

As shown by the data in Table 12, the probability of nondetection for verification alone is larger than the stated inspection goal of 0.05 except for the large value of $M = 2370$ kilograms of uranium. Before considering how the Agency might increase its sample sizes to reach the desired

TABLE 11 Values of $R(\theta)$

R(θ) for Positive θ Values										
	0.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
.90	.3261	.3271	.3187	.3142	.3104	.3065	.3026	.2988	.2951	.2913
.80	.3676	.3633	.3590	.3548	.3506	.3465	.3423	.3382	.3342	.3301
.70	.4119	.4074	.4028	.3983	.3938	.3894	.3850	.3806	.3762	.3719
.60	.4591	.4543	.4495	.4447	.4399	.4352	.4305	.4258	.4211	.4165
.50	.5092	.5040	.4989	.4939	.4888	.4838	.4788	.4739	.4689	.4640
.40	.5619	.5565	.5511	.5458	.5405	.5352	.5299	.5247	.5195	.5143
.30	.6172	.6116	.6059	.6004	.5948	.5892	.5837	.5782	.5727	.5673
.20	.6751	.6692	.6633	.6575	.6516	.6458	.6401	.6343	.6286	.6229
.10	.7353	.7292	.7231	.7170	.7109	.7049	.6989	.6929	.6869	.6810
0.00	.7979	.7915	.7852	.7789	.7726	.7663	.7601	.7539	.7477	.7415

R(θ) for Negative θ Values										
	-0.00	-.01	-.02	-.03	-.04	-.05	-.06	-.07	-.08	-.09
-0.00	.7979	.8043	.8107	.8171	.8235	.8300	.8365	.8430	.8495	.8560
-.10	.8626	.8692	.8758	.8824	.8891	.8958	.9025	.9092	.9159	.9226
-.20	.9294	.9362	.9430	.9498	.9567	.9636	.9704	.9773	.9843	.9912
-.30	.9982	1.0051	1.0121	1.0192	1.0262	1.0332	1.0403	1.0474	1.0545	1.0616
-.40	1.0688	1.0759	1.0831	1.0903	1.0975	1.1047	1.1119	1.1192	1.1265	1.1338
-.50	1.1411	1.1484	1.1557	1.1631	1.1705	1.1779	1.1853	1.1927	1.2001	1.2076
-.60	1.2150	1.2225	1.2300	1.2375	1.2450	1.2526	1.2601	1.2677	1.2753	1.2829
-.70	1.2905	1.2981	1.3058	1.3134	1.3211	1.3288	1.3365	1.3442	1.3519	1.3596
-.80	1.3674	1.3752	1.3829	1.3907	1.3985	1.4064	1.4142	1.4220	1.4299	1.4378
-.90	1.4456	1.4535	1.4614	1.4694	1.4773	1.4852	1.4932	1.5012	1.5091	1.5171
-1.00	1.5251	1.5331	1.5412	1.5492	1.5573	1.5653	1.5734	1.5815	1.5896	1.5977
-1.10	1.6058	1.6139	1.6221	1.6302	1.6384	1.6465	1.6547	1.6629	1.6711	1.6793
-1.20	1.6876	1.6958	1.7040	1.7123	1.7205	1.7288	1.7371	1.7454	1.7537	1.7620
-1.30	1.7703	1.7787	1.7870	1.7954	1.8037	1.8121	1.8205	1.8288	1.8372	1.8456
-1.40	1.8541	1.8625	1.8709	1.8794	1.8878	1.8963	1.9047	1.9132	1.9217	1.9302
-1.50	1.9387	1.9472	1.9557	1.9642	1.9728	1.9813	1.9899	1.9984	2.0070	2.0156
-1.60	2.0241	2.0327	2.0413	2.0499	2.0585	2.0672	2.0758	2.0844	2.0931	2.1017
-1.70	2.1104	2.1190	2.1277	2.1364	2.1451	2.1537	2.1624	2.1711	2.1799	2.1886
-1.80	2.1973	2.2060	2.2148	2.2235	2.2323	2.2410	2.2498	2.2586	2.2674	2.2761
-1.90	2.2849	2.2937	2.3025	2.3114	2.3202	2.3290	2.3378	2.3467	2.3555	2.3643
-2.00	2.3732	2.3821	2.3909	2.3998	2.4087	2.4176	2.4264	2.4353	2.4442	2.4532
-2.10	2.4621	2.4710	2.4799	2.4888	2.4978	2.5067	2.5156	2.5246	2.5336	2.5425
-2.20	2.5515	2.5605	2.5694	2.5784	2.5874	2.5964	2.6054	2.6144	2.6234	2.6324
-2.30	2.6414	2.6504	2.6595	2.6685	2.6775	2.6866	2.6956	2.7047	2.7137	2.7228
-2.40	2.7319	2.7409	2.7500	2.7591	2.7682	2.7773	2.7864	2.7955	2.8046	2.8137
-2.50	2.8228	2.8319	2.8410	2.8501	2.8593	2.8684	2.8775	2.8867	2.8958	2.9050
-2.60	2.9141	2.9233	2.9324	2.9416	2.9508	2.9600	2.9691	2.9783	2.9875	2.9967
-2.70	3.0059	3.0151	3.0243	3.0335	3.0427	3.0519	3.0611	3.0703	3.0796	3.0888
-2.80	3.0980	3.1073	3.1165	3.1257	3.1350	3.1442	3.1535	3.1627	3.1720	3.1812
-2.90	3.1905	3.1998	3.2090	3.2183	3.2276	3.2369	3.2462	3.2554	3.2647	3.2740

Examples: $R(0.20) = 0.6751$, $R(0.24) = 0.6516$, $R(0.246) = 0.6481$,

$R(-1.73) = 2.1364$, $R(-1.732) = 2.1381$, $R(e) = \phi'(e)/\phi(e)$

TABLE 12 Probability of Nondetection for Verification

Parameter Values												
M, kgs U	(1)											
	$\sigma_{\hat{D}_r H_0}$	$\sigma_{\hat{D}_r H_1}$	$\sigma_{\hat{D}_s}$	$\sigma_{\hat{D} H_0}$	$\sigma_{\hat{D} H_1}$	$R(\theta)$	θ	G	D	Q_1	Q_2	$Q_{\max}$
833	83.7	167.4	313.7	325	356	1.78	-0.69	53	780	0.83	0.245	0.20
1140	91.1	182.2	313.7	327	363	0.95	-0.23	519	621	0.255	0.41	0.10
2370	111.1	222.2	313.7	333	384	0.49	0.54	2030	340	0.077	0.70	0.054

(1) Random error σ 's based on final sample sizes given in Table 10.

where: $Q_1 = (0.05)^{\frac{G}{M}}$ $Q_{\max} = Q_1 Q_2$

$$Q_2 = \Phi\left(\frac{1.645 \sigma_{\hat{D}|H_0} - D}{\sigma_{\hat{D}|H_1}}\right)$$

$$G + D = M \text{ and}$$

$$R(\theta) = - \frac{\ln(0.05) \sigma_{\hat{D}|H_1}}{M}$$

$$G = M + \theta \sigma_{\hat{D}|H_1} - 1.645 \sigma_{\hat{D}|H_0}$$

θ, R : From Table 11

power of detection, it is best to consider the overall power of detection first.

If the operator's diversion of M is assumed to be partitioned between single defects, small biases, and direct diversions which would appear as MUF, then the probability of nondetection is given by:

$$Q_{\text{tot}} = Q_1 Q_2 Q_3.$$

Where Q_1 , Q_2 and Q_3 are defined as before.

For this situation, the inspector evaluates the three probabilities separately. That is, defects are tested by measurements made in the attribute mode, biases are tested versus $\sigma_{\hat{D}}|_{H_0}$ and MUF is evaluated separately. For purposes of illustration only, values for C_{max} are given in Table 13. This is not the recommended approach. However, the values in Table 13 do clearly illustrate the decrease in the power of detection when M can be partitioned between three paths rather than two paths. This effect is best seen by comparing the values of Table 12 with those of Table 13. As the data show, the loss in detection power, $(1-Q)$ is about 10-15 percent.

Because of the correlation between $\hat{D}$ and MUF, the above approach is not recommended. A better approach for the over-all power of detection is to use the $(\hat{D} + \text{MUF})$ approach. As was shown earlier, $(\hat{D} + \text{MUF})$ contains both the path of falsification by small biases as well as direct diversion. In addition the variance of $(\hat{D} + \text{MUF})$ is smaller than that of $\hat{D}$. Thus, the over-all power of nondetection reduces to:

$$Q_{\text{overall}} = Q_1 Q_{2,3} \text{ or } Q_1 Q(\hat{D} + \text{MUF})$$

The variance of $(\hat{D} + \text{MUF})$ is given by

TABLE 13 Probability of Nondetection when Single Defects, $\hat{D}$, and MUF are Treated Separately

Parameter Values										
M, kgs U	(1)		$\sigma_{\hat{D} H_1}$	σ_{MUF}	G	D	MUF _T	•		
	$\sigma_{\hat{D} H_0}$							Q_1	Q_2	Q_3
833	325	356	230.5	0	555	278	1.0	0.477	0.669	0.32
1140	327	363	230.5	210	625	305	0.575	0.405	0.627	0.146
2370	333	384	230.5	1882	337	151	0.093	0.71	0.838	0.055

(1) Random error component of $\sigma_{\hat{D}|H_0}$ based on final sample sizes of Table 10.

$$\begin{aligned}
 Q_1 &= (0.05)^{\frac{G}{M}} \\
 Q_2 &= \Phi \left[\frac{1.645 \sigma_{\hat{D}|H_0} - D}{\sigma_{\hat{D}|H_1}} \right] \\
 Q_3 &= \Phi \left[1.645 - \frac{MUF_T}{\sigma_{MUF}} \right]
 \end{aligned}$$

$$Q_{\max} = Q_1 Q_2 Q_3$$

Solutions to $Q_{\max}$ are given in Table 13 (cont.) on the following page.

TABLE 13 (cont.)

M, kgs U	R(θ_D)	(θ_D)	R(θ_{MUF})	(θ_{MUF})	G	D	MUF _T
1140	0.954	-0.24	0.605	0.32	210	625	305
2370	0.485	0.55	0.29	0.99	1882	337	151

When G not = 0							
$D = 1.645 \sigma_D _{H_O} - \theta_D \sigma_D _{H_1}$							
$MUF = 1.645 \sigma_{MUF} - \theta_{MUF} \sigma_{MUF}$							
$- \ln(0.05) (\sigma_D _{H_1})$							
$R(\theta_D) = \frac{M}{M}$							
$R(\theta_{MUF}) = \frac{- \ln(0.05) (\sigma_{MUF})}{M}$							
$R \left[\frac{1.645 \sigma_D _{H_O} - D}{\sigma_D _{H_1}} \right] = \frac{\sigma_D _{H_1}}{\sigma_{MUF}} R \left[\frac{1.645 - MUF_T}{\sigma_{MUF}} \right]$							

R(θ), θ values are given in Table 11

$$\sigma^2_{(\hat{D} + \text{MUF})} = \sigma^2_{\hat{D}} - \sigma^2_{\text{MUF}}$$

To be conservative in estimating the power of detection, the alternative and null hypotheses for $\hat{D}$ are used in computing an alternative and null hypothesis for $(\hat{D} + \text{MUF})$. The over-all power of detection is given for an α of 0.05 by:

$$C_{\text{overall}} = (0.05)^{\frac{G}{M}} \Phi \left[\frac{1.645 \sigma_{(\hat{D} + \text{MUF}) | H_0} - (D + \text{MUF}_T)}{\sigma_{(\hat{D} + \text{MUF}) | H_1}} \right]$$

$$\text{where } G + (D + \text{MUF}_T) = M$$

Again, the Agency is interested in evaluating its power of detection under the operator's best strategy or at C_{max} . These may be obtained as before with Stewart's method using $R(\theta)$ and θ substituting in the values for $\sigma_{(\hat{D} + \text{MUF})}$ for $\sigma_{\hat{D}}$ in the previous expression.

For example the C_{max} , for $M = 1140$ may be found as follows:

$$\begin{aligned} \theta &= R^{-1} \left[\frac{(-\ln 0.05) \sigma_{(\hat{D} + \text{MUF}) | H_1}}{1140} \right] \\ &= R^{-1} \left[\frac{(2.9957)(280)}{1140} \right] \\ &= R^{-1}(0.736) \end{aligned}$$

From Table 11, $\theta = 0.10$ for $R(\theta) = 0.73$.

Therefore:

$$\begin{aligned} G &= 1140 + 0.10(280) - 1.645(232) \\ G &= 786 \\ M-G &= 354 \end{aligned}$$

and

$$C_{\max} = (0.05)^{\frac{786}{1140}} \Phi \left[\frac{1.645 (232) - 354}{280} \right]$$

$$C_{\max} = (0.127)(0.54) = 0.0686$$

The values for $C_{\max}$ are given in Table 14 for the $(\hat{D} + \text{MUF})$ approach. Those values correspond to the sample sizes given in Table 10 for the final sample sizes for implementation.

As shown by the data in Table 14 for the $C_{\max}$ values, the $(\hat{D} + \text{MUF})$ approach has a greater over-all power of detection than either verification alone, or the separate treatment of verification and MUF evaluation.

The values for the standard deviations shown for $(\hat{D} + \text{MUF})$ in Table 14 need explanation. The computational formula for $\sigma_{(\hat{D} + \text{MUF})}$ or $\sigma_{(\hat{D} + \text{MUF})}^2$ are based on the following:

$$\sigma_{(\hat{D} + \text{MUF})}^2 = \sigma_{\hat{D}}^2 - \sigma_{\text{MUF}}^2$$

$$\sigma_{(\hat{D} + \text{MUF}) | H_0}^2 = \sigma_{\hat{D} | H_0}^2 - \sigma_{\text{MUF}}^2$$

$$\sigma_{(\hat{D} + \text{MUF}) | H_1}^2 = \sigma_{\hat{D} | H_1}^2 - \sigma_{\text{MUF}}^2$$

For $M = 833$, then:

$$\sigma_{(\hat{D} + \text{MUF}) | H_0}^2 = (325)^2 - (230.5)^2$$

$$\sigma_{(\hat{D} + \text{MUF}) | H_0} = 229$$

TABLE 14 Over-all Probability of Nondetection Using the $\hat{D} + \text{MUF}$ Approach

M, kgs U	$\sigma \hat{D} H_0$	$\sigma \hat{D} H_1$	$\sigma (\hat{D} + \text{MUF}) H_0$	$\sigma (\hat{D} + \text{MUF}) H_1$	R(θ)	θ	G	D+MUF _T	Q ₁	Q ₂	Q _{Max}
833	325	356	229	271	0.97	0.27	383	450	0.25	0.39	0.098
1140	327	363	232	280	0.74	0.10	786	354	0.127	0.54	0.069
2370	333	384	240	307	0.39	0.75	2205	165	0.062	0.78	0.048

(1) Random error σ 's based on sample sizes given in Table 10.

$$\text{Where } Q_1 = (0.05) \frac{G}{M} \quad : \quad Q_{\max} = Q_1 Q_2$$

$$Q_2 = \Phi \left(\frac{1.645 \sigma (\hat{D} + \text{MUF}) | H_0 - (D + \text{MUF}_T)}{\sigma (\hat{D} + \text{MUF}) | H_1} \right)$$

$$M = G + D + \text{MUF}_T$$

$$G = M + \theta \sigma (\hat{D} + \text{MUF}) | H_1 - 1.645 \sigma (\hat{D} + \text{MUF}) | H_0$$

$$R(\theta) = - \frac{\ln(0.05) \sigma (\hat{D} + \text{MUF}) | H_1}{M} ; \quad R, \theta \text{ given in Table 11.}$$

At first glance it would seem that the value of $\sigma_{(\hat{D} + \text{MUF})|H_0}$ is in error since it is smaller than σ_{MUF} . For example, 229 for $\sigma_{(\hat{D} + \text{MUF})|H_0}$ versus 230.5 for σ_{MUF} . The actual value of 229 is correct based on the assumptions used in the example. However, in general, one would expect $\sigma_{(\hat{D} + \text{MUF})}$ to be larger than σ_{MUF} because of the random error inflation arising from the use of a fractional sampling plan. In the actual example, this effect is not dominant* because of the systematic error term and because of the use of average uranium factors by the operator.

Final Evaluation of Sampling Plans

The final step in the planning stage is the evaluation of the designed sampling plans in light of the Agency's desired detection goals. The previous Q_{max} values given in Table 14 for the $(\hat{D} + \text{MUF})$ show the probability of non-detection for the three values of M and the sample sizes shown in Table 10.

Since in all cases the sample size design was approached from a conservative sense of always taking at least enough samples, the probabilities of nondetection are at least as low as those shown in the table. Also, from the standpoint of the values themselves, which range in terms of probabilities of detection from 90.- 95 percent, the plans appear adequate. However, if the goal is to detect the loss or diversion of M with $\beta = 0.05$ and $\alpha = 0.05$, that goal is not met for all values of M.

If a higher power of detection is desired, the sample sizes can be recomputed using a lower β value for the attribute plans and a lower random error goal for $\sigma_{\hat{D}}$ or $\sigma_{(\hat{D} + \text{MUF})}$. However, it should be noted that this approach is fruitful only when the limits of measurement uncertainty do not preclude the detection of a quantity M. A case in point is the goal quantity of M = 833 kgs U which was used in the example. That quantity closely approaches the detection limits imposed by the measurement uncertainty.

* See page 59.

Basis for the Values of M Used in the Example

The values of M which were used in the example were chosen to illustrate several verification goals: These are:

- M = 833 kgs U corresponds to 25 kg of contained ^{235}U .
- M = 1140 kgs U corresponds to the practical detection goal for the difference statistic $\hat{D}$, for bias detection under the measurement uncertainties assumed for the example.
- M = 2370 kgs U corresponds to the 95 percent detection goal for an international standard of accountability of $\sigma_{\text{ISOA}} = 720$ kgs U, or 0.3 percent of feed in the example. The 95 percent detection goal is taken as $2(1.645)\sigma_{\text{ISOA}}$ or $3.29 \times 720 = 2370$.

The value of 1140 kgs U for M was chosen to illustrate a practical detection goal for bias detection using the difference statistic. That this choice is approximately true is best illustrated by the example in Step 1 on page 91. Here the ratio of $\frac{M}{\sigma_{\hat{D}_S}}$ is 3.63 and ratio of $\sigma_{\hat{D}_R|H_1}/\sigma_{\hat{D}_S}$ is about 0.6. A ratio of $\sigma_{\hat{D}_R|H_1}/\sigma_{\hat{D}_S}$ of 0.5 is considered a practical limit.

E. Glossary of Terms

Following is a list of symbols which are frequently used in this document. Sometimes it is necessary within the test to use additional subscripting but this will be easily understood in the context of use. It is generally desirable to cap a statistical estimate with a caret such as in $\hat{D}$ and to let the uncapped symbol denote the true amount to be estimated. The unadorned and traditional symbol MUF was retained as an estimate, however, and the symbol MUF_T was introduced to indicate the true value.

<u>Symbol</u>	<u>Meaning</u>
M	The goal quantity under the alternative hypothesis. The quantity used by the inspector in designing attribute and variable sample sizes so that a diversion of the amount M will be detected at a preassigned probability level.
MUF	The operator's estimate of the amount of material missing in a material balance area during an accountability period. The material balance area may be a facility or a complex of facilities as well as an area within a facility. Since the symbol MUF, rather than $\hat{MUF}$, has been designated as the appropriate symbol for this statistic for so long its meaning is retained in this document.
N	The number of items which go into an operator's MUF calculation.
N_i	The number of items in the <i>i</i> th class.
Class	The terms class and stratum are used interchangeably.

<u>Symbol</u>	<u>Meaning</u>
$\bar{X}_i$	The nominal amount per item in the <i>i</i> th stratum where the amounts per item over the different items are relatively constant. For design purposes it is better to choose a value larger than the average when the amounts per item are somewhat heterogeneous. Another possibility when there is marked heterogeneity is to stratify the population in the particular stratum by amount.
r_1	The number of items needed by the operator to attain the goal quantity <i>M</i> by the removal of the gross quantities of $\bar{X}$ per item.
γ	γ (gamma) is the proportion of the total removal of $\bar{X}$ from an item such that removals below $\gamma\bar{X}$ would not be detected by the attribute instrument. $\gamma\bar{X}$ is treated as a point on a continuum for design purposes but in fact there is a range of values where the attribute instrument does not give an unequivocal signal. The crossover point $\gamma\bar{X}$ is chosen, however, so that the sample size indicated will be in the direction of requiring at least enough samples.
r_2	The number of items needed by the operator to attain the goal quantity <i>M</i> by the removal of the partial amounts $\gamma\bar{X}$ per item.

SymbolMeaning n_a

The number of samples to be taken in a stratum by the attribute instrument in order to detect with probability $100(1-\beta)$ percent or higher the removal of at least one of $r_1 = M/\bar{X}$ items of material. The attribute measurement can detect with 100 percent probability any item which has had 100γ percent to 100 percent of the material removed.

 n_i

The number of items in the i th class used for inspector variable measurement and operator/inspector item amount value comparisons. More precisely

$$n_i = (n_v)_i = (\max (n_{v1}, n_{v2}, n_{v3}))_i$$

where the symbols to the right of n_i are defined elsewhere in the Glossary.

 n_{v1}

The number of samples required by the variable measurement in order to be $100(1-\beta)$ percent sure of detecting the removal of at least one defect if $r_2 = M/(\gamma\bar{X})$ removals of the amount $\gamma\bar{X}$ are taken.

 n_{v2}

The number of items needed to detect a significant bias, where Δ is the bias and $N\Delta = M$, where it is also assumed that an induced process variance by the operator exists.

 n_{v3}

The number of samples needed in order to be $100(1-\beta)$ percent sure of detecting an understatement of 75

<u>Symbol</u>	<u>Meaning</u>
n_{v3} (cont.)	percent in the operator's measurement variance in the item amount values. Another way of saying this is that the test on the operator's stated variance will detect a true variance four times as large as the stated design variance value $100(1-\beta)$ percent of the time. β is usually chosen as 0.05.
n_v	The number of samples to be taken in a stratum by the variable instrument. n_v is the largest of the three values n_{v1} , n_{v2} , and n_{v3} , i.e., $n_v = \max (n_{v1}, n_{v2}, n_{v3})$.
$\bar{d}_i$	The average difference between the inspector/operator amount values in a random sample taken from the i th stratum.
d_{ij}	The difference between the inspector's amount value and the operator's stated amount for the j th item within the i th class.
$\hat{D}$	$\hat{D} = \sum \pm \hat{D}_i$, where the positive or negative sign corresponds to the sign the class has in the MUF statistic. It is essentially an estimate of the negative of the operator's MUF if the inspector's measurements are unbiased; it is an estimate of the inspector's MUF minus the operator's MUF, where the inspector's estimate is generally based on fractional sampling.

Symbol
($\hat{D} + \text{MUF}$)

Meaning

The statistic formed by adding the operator's MUF statistic to the $\hat{D}$ statistic. It is essentially a MUF statistic adjusted for operator bias since $-\hat{D}$ estimates the bias in MUF when the inspector's bias is zero or negligible. If the inspector's bias is not negligible ($\hat{D} + \text{MUF}$) replaces operator bias with inspector bias.

$\sigma_{Is}^2, \sigma_{Ir}^2$

These symbols denote, respectively, the inspector measurement variance on the amount of an individual item due to systematic and random error.

$\sigma_{Os}^2, \sigma_{Or}^2$

These symbols are respectively, the operator counterparts of σ_{Is}^2 and σ_{Ir}^2 . δ denotes a coefficient of variation.

$\delta, \delta_{Or}, \delta_{Os},$
 δ_{Ir}, δ_{Is}

It is a relative standard deviation. Thus, $\delta = \sigma/\mu$ where σ is the standard deviation, μ is the true value. δ_{Or} , for example means $\delta_{Or} = \sigma_{Or}/\mu$.

$\sigma_{\text{MUF } r}^2$

The variance in the operator's MUF due to the random error variance component.

$\sigma_{\text{MUF } s}^2$

The variance in the operator's MUF due to the systematic error variance component.

σ_{MUF}^2

The variance in the operator's MUF including both the systematic and random error variance components.

σ_{dr}

The standard deviation of the difference in the operator's and inspector's amount values for an individual item due to random errors of measurement.

<u>Symbol</u>	<u>Meaning</u>
$\sigma_{\hat{D}_r}^2 H_0$	The random variance component in the statistic $\hat{D}$ under the null hypothesis that the material is in control.
$\sigma_{\hat{D}_r}^2 H_1$	The random variance component in $\hat{D}$ under the alternative hypothesis.
$\sigma_{\hat{D}}^2 H_0$	The variance of $\hat{D}$ under the null hypothesis of material control. Written in this form the variance in $\hat{D}$ is the sum of the random and systematic variance components.
$\sigma_{\hat{D}}^2 H_1$	The variance of $\hat{D}$ under the alternative hypothesis of material diversion of the goal quantity M. The random variance component in this situation may include the induced process variance in addition to the natural sources of variation. The variance in $\hat{D}$ under H_1 is the sum of the random and systematic variance components.
s_{dr}^2, s_d^2	The observed variance in the d values calculated from $s_d^2 = \frac{n \sum d_i^2 - (\sum d_i)^2}{n(n-1)}$ where $n = n_v$. The formula is applied to items in the same MUF stratum. Both symbols are used to denote the same variance estimate.
$\sigma_{(\hat{D}+MUF)}^2$ Var($\hat{D}+MUF$) Variance($\hat{D}+MUF$)	These symbols all denote the same parameter, the variance in the statistic ($\hat{D}+MUF$) including both the systematic and random error variance components.

<u>Symbol</u>	<u>Meaning</u>
$\sigma^2_{(\hat{D}+MUF) H_0}$	The variance of the statistic $\hat{D}+MUF$ when the null hypothesis is true that the special nuclear material is in control.
$\sigma^2_{(\hat{D}+MUF) H_1}$	The variance in the statistic $\hat{D}+MUF$ when the alternative hypothesis that $G + D + MUF_T = M$ is true.
σ^2_p	The induced process variance introduced by the operator in order to increase the inherent variability in the $\hat{D}$ statistic and thus to increase the probability that the alternative hypothesis will not be rejected when in fact it is true.
F_C	The finite population correction factor $F = 1 - \frac{n-1}{N-1} = \frac{N-n}{N-1}$ which appears in the induced process variance formulae for mean values. F_C is neglected by the inspector in design procedures. The effect of this leads to larger and thus more conservative n_{v2} sample size estimates.
D	The falsification amount obtained by biasing.
G	The amount of material obtained by gross defects (total and partial) removals.
MUF_T	The true amount of material diverted into MUF. This is not necessarily the total amount of material diverted from the system but represents the true discrepancy in the mass balance accounting system which is not concealed by falsification.

Symbol

Meaning

η

In maximizing Q , the probability of nondetection in the two-path gross defects, bias case, η is the proportion of M which should be obtained by biasing, $1-\eta$ the proportion by gross defectives. This assumes that $\sigma_{\hat{D}} = \sigma_{\hat{D}|H_0} = \sigma_{\hat{D}|H_1}$ and the attribute sample sizes are obtained by using the binomial distribution to solve for n_a and n_{v1} .

$Q_{\max}$

The maximum probability of non-detection under a certain situation. For example the vector (G, D, MUF_T) which maximizes

$$Q = \beta^{G/M} \Phi \left(\frac{Z_{1-\alpha} \sigma_{\hat{D}|H_0} - \frac{D}{\sigma_{\hat{D}|H_1}}}{\sigma_{\hat{D}|H_1}} \right) \Phi \left(Z_{1-\alpha} - \frac{MUF_T}{\sigma_{MUF}} \right)$$

subject to the constraints $G > 0$, $D + G + MUF_T = M$, yields the value $Q_{\max}$.

$R(\theta)$

$R(\theta) = \Phi'(X) / \Phi(X)$ is a ratio used in determining $Q_{\max}$. It is related to $S(\theta)$, the extensively tabled Mill's function, by the relationship $R(\theta) = 1/S(-\theta)$.

$R^{-1}(\theta)$

The inverse function of $R(\theta)$, the argument of the function which is θ .

$S^{-1}(\theta)$

The inverse function of Mill's ratio.

<u>Symbol</u>	<u>Meaning</u>
$z_{1-\alpha}, z_{1-\beta}$	z_γ is the 100 γ percentile point of the $N(0,1)$ distribution. Thus $z_{1-\alpha}, z_{1-\beta}$ are, respectively, the 100(1- α) and 100(1- β) percentile points of the $N(0,1)$ distribution.
$N(0,1)$	This is a conventional symbol to designate a zero mean, unit variance normal (Gaussian) distribution. The distribution function of this distribution is denoted by $\Phi(X)$, the density by $\Phi'(X)$.
$\Phi(X)$	Several different but equivalent definitions are given. $\Phi(X)$ is the distribution function of the $N(0,1)$ distribution. $\Phi'(X)$ is the corresponding density function. Thus
	$\Phi(X) = \int_{-\infty}^X \Phi'(t) dt, \text{ where}$ $\Phi'(t) = \frac{e^{-t^2/2}}{\sqrt{2\pi}}$
$\Phi'(X)$	The density function of the $N(0,1)$ distribution. (See $\Phi(X)$).
$S(\theta)$ (Mills' Ratio)	$S(\theta) = e^{\theta^2/2} \int_{\theta}^{\infty} e^{-t^2/2} dt$ $= \Phi(-\theta) / \Phi'(\theta)$ $= 1/R(-\theta)$

IV. REFERENCES

1. IAEA Panel on, "Systems of Accounting for and Control of Nuclear Material," Tokyo, Japan, November 1973.
2. J. L. Jaech, Statistical Methods in Nuclear Material Control, USAEC, TID-26298, 1973.
3. K. B. Stewart, "A Source Book for Safeguards Inventory Verification," Statistical Procedures and Concepts, BNWL-B-149, January, 1972.
4. L. F. Wirfs, "A Detailed Analysis of the Safeguards Capabilities of Material Balance Accounting Systems in the Low Enriched UO₂ Conversion and Fabrication Sectors of the Commercial Fuel Cycle," BNWL-1336, January, 1970.
5. K. B. Stewart and R. A. Schneider, "Verification Sampling Techniques in a Safeguards Situation," Nuclear Materials Management, Vol. III, No. 1, Spring, 1974.
6. F. Brown, P. T. Good, J. B. Parker, A. E. Ross and D. R. Terry (UKAEA), A. Lumetti, E. Kerr, T. Otomo, and M. Sultan (IAEA), "Progress in the Development of a System for the Safeguards Inspection of the Fuel Store of a Zero-Energy Fast Reactor." A joint report by the International Atomic Energy Agency and the United Kingdom Atomic Energy Authority, COS 20, October, 1970.
7. Hahn and Shapiro, Statistical Models in Engineering, John Wiley and Sons, 1968.
8. J. L. Jaech, "Pitfalls in the Analysis of Paired Data," Nuclear Materials Management, Vol. II, No. 11, Summer, 1973.

DISTRIBUTION

No. of
Copies

OFFSITE

1	<u>AEC Chicago Patent Group</u> U. S. Atomic Energy Commission Argonne, IL 60439 A. A. Churm
321	<u>AEC Technical Information Center</u>
10	<u>Division of Safeguards and Security</u> U. S. Atomic Energy Commission Mail Stop, A2-1303 Washington, D. C. 20545 J. F. Mahy
10	<u>U. S. Arms Control and Disarmament Agency</u> State Department Building Washington, D. C. 20451 F. S. Houck

ONSITE

1	<u>AEC Richland Operations Office</u> B. J. Melton/P. F. X. Dunigan, Jr.
89	<u>Battelle-Northwest</u> R. A. Schneider (85) Technical Information (3) Technical Publications (1)