APPENDIX B

Statistical and Non-Statistical Sampling and the Use of CAATs

B.1 Introduction

This annex illustrates how an auditor can use CAATs software to select audit samples when statistical or non-statistical sampling is being performed.

There are two basic sampling methods – monetary unit sampling (MUS) for substantive testing, and attribute sampling. The latter is primarily used for compliance testing.

For example, Section B.5 discusses MUS for substantive tests of details. It illustrates how the auditor determines the sample size, selects the sample (either statistically or non-statistically), and then evaluates the sample results. If the auditor were performing a compliance with authority audit on a particular contract or project, the auditor could use the same approach to determine the sample size and select the sample.

The auditor could also use the same error evaluation process, only now the auditor's errors would be amounts not in compliance with a particular authority, as opposed to monetary errors in the recorded amounts. The only real difference in the process is that the auditor would normally not need to deal with "taintings" – the percentage by each a particular sample item is in error – because all errors would normally be 100% errors. (We normally cannot have, for example, an invoice that is 50% approved, or a transaction that is 70% consistent with the nature of the appropriation to which it was charged. The invoice will normally be approved or not approved and the transaction will normally be consistent with the nature of the appropriation to which it was charged or not consistent.)

Sampling is only one source of audit assurance. It is important for the auditor to consider the assurance that is derived from all audit procedures when determining whether he/she has sufficient appropriate audit evidence.

Note that the only difference between statistical sampling and non-statistical sampling is the method of selecting sample items. Planning requirements and the evaluation process remain the same.

This Appendix contains guidelines that the auditor can use to plan a sample. These guidelines should *not* replace the use of professional judgement. No set of guidelines can be expected to be valid on all audits.

In this Appendix, three terms occur often enough that acronyms are used:

MLE	Most likely error
MUS	Monetary unit sampling
UEL	Upper error limit

Use of CAATs

Using a CAATs tool to perform sample size calculations and error evaluations is superior to performing a manual process for a number of reasons:

(a) It will take significantly longer to do the calculations manually than to use CAATs. Error evaluation calculations, in particular, can take a long time to do manually, and can be done in seconds using CAATS.

(b) Manual calculations are complex and prone to error. For example, auditors may fail to properly rank errors from the largest tainting to the smallest tainting when doing the error evaluation.

(c) Should the auditor use planning parameters that are not are normally used or that are theoretically questionable, CAATS will either provide a warning message or will not allow the calculation to proceed. When a calculation is being performed manually, the auditor loses this protection.

(d) Manual calculations result in sample sizes that are larger than those produced by the CAATs software, further adding to the time required to perform the audit.

(e) Statistical tables do not contain every possible confidence level. Using the tables in manual calculations therefore requires rounding up the confidence level, which will further increase the sample size and further increase the time required to perform the audit.

(f) Manual calculations result in very conservative error evaluations, which could cause DAGP to conclude that the results of the work are unacceptable when, in fact, they are acceptable.

Because of the above-noted advantages, auditors are strongly encouraged to use CAATS. However, there may be rare circumstances where manual calculations are required. For this reason, the Standard Audit Working Paper Kit contains standard forms that can be used.

B.2 Basic concepts and definitions

Sampling

Sampling is the selection of a sub-set of a population. The auditor takes a sample to reach a conclusion about the population as a whole. As such, it is important that the sample be representative of the population from which it was selected.

Statistical sampling

Statistical sampling is the selection of a sub-set from a population in such a way that each sampling unit has an equal and known chance of selection.

The main advantages of statistical sampling over non-statistical sampling are:

1. Because each sampling unit has an equal and known chance of selection, there is a better chance that the sample will be representative of the population than is the case with a non-statistical sample. When expressing an opinion on financial statements, having a representative sample is very important.

2. Because there is a better chance that the sample will be representative of the population, the sample results are more objective and defensible, as are the projections of those results to the population as a whole.

3. It provides a direct estimate of the maximum possible error (referred to as the upper error limit (UEL) in some CAATS).

Non-statistical sampling

Non-statistical samples are samples selected by other means which are intended to approximate the representative character of a statistical sample. However, they lack the objectivity of a statistically selected sample.

Given the advantages of a statistical sample, a non-statistical sample should, in theory, always be larger than a statistical sample. When non-statistical sampling is used, it may be appropriate to increase the sample size by 20% to 50%.

Sampling unit

The *sampling unit* is the specific item of which the population is assumed to be composed for sampling purposes.

As an example, consider a population of purchases for the year. Assume that the purchases are recorded by cash disbursement, that each disbursement may relate to several supplier invoices, and that each supplier invoice may relate to several purchases. In this example, the sampling unit could be:

1. Each cash disbursement;

2. Each supplier invoice within each cash disbursement;

3. Each purchase within each supplier invoice; or

4. Each Rupee of value within each purchase.

If the auditor sets an individual cash disbursement as the sampling unit, the sample selection process would be much simpler than if the auditor set an individual purchase within a supplier invoice as the sampling unit. However, by setting each cash disbursement as the sampling unit, the auditor would have to audit all supplier invoices and all purchases within each selected cash disbursement.

Physical unit

The physical unit is the specific document (cash disbursement, individual supplier invoice or individual purchase, for example) to which the sampling unit is assumed to relate.

The physical unit is normally the same as the sampling unit. The primary exception is MUS where the sampling unit is each individual monetary unit (Rupee).

Population size

The population size is the number of sampling units (cash disbursements, supplier invoices, purchases or Rupees) in the population.

The population size will vary depending on the sampling unit being used. For example, our population of purchases for the year may be composed of 16,000 cash disbursements, 30,000 supplier invoices, 70,000 purchases, and 100,000,000 individual Rupees. Depending on which sampling unit has been selected, any of these amounts could constitute the population size.

Except for small population sizes, the size of the population does not influence the size of the sample selected. For other than very small populations, the sample size is dependent on the assumed variability (error rate) of the population, on the accuracy required from the sample (determined by consideration of materiality) and the confidence level (determined by consideration of risk). Accordingly, the auditor should not think in terms of selecting a percentage of the population. Taking a fixed percentage will tend to under-sample a small population and over-sample a large population.

The following table illustrates the relationship between sample size and population size. Note how, after a certain size, the population size does not influence the sample needed to achieved the desired level of confidence.

A simple approach when using automated tools to calculate sample size, is, if the calculated sample is as large or larger than the population, then take the whole population into the sample.

Population	**Sample Size for Precision Percentage of Plus or Minus**			
Size	**1.00**	**2.00**	**3.00**	**4.00**
50	48	45	40	34
100	94	82	66	53
150	138	112	86	64
200	180	139	100	72
250	219	161	111	78
500	392	238	144	92
1000	645	313	168	102
2000	954	371	184	107
5000	1336	418	194	111
10000	1543	436	198	112
20000	1672	446	200	113
50000	1760	452	201	113
100000	1792	454	202	113

Table: Sample sizes for attributes sampling, expected error rate not over 5%, confidence level 95%

Population value

The population value is the monetary amount of the population being sampled. In the above example, it would be Rs. 100,000,000.

There may be individually significant transactions that the auditor wants to examine. These could be very large transactions or transactions with high risk. Auditors often audit 100% of these transactions, and take a sample of the remaining transactions.

To arrive at the population value for sampling purposes, the auditor needs to subtract the total value of the individually significant transactions from the total population value. For example, if the auditor decides to audit all transactions greater than Rs. 500,000 and to take a sample of the remaining transactions, the total value of the items greater than Rs. 500,000 would be removed from the population value when determining the required sample size.

Sometimes the auditor does not know the population value when determining the sample size. For example, the auditor may wish to select a sample of supplier invoices for the year, and may start auditing the transactions before the end of the year. In this case, the auditor will make an estimate of the population value at the planning stage.

Sampling risk

Sampling risk is the chance that a sample is not representative of the population from which it was selected.

If the sample is not representative the auditor could reach an incorrect conclusion about the population from which the sample was selected. The auditor could incorrectly conclude that:

1. The population is not materially misstated when, in fact, it is materially misstated; or

2. The population is materially misstated when, in fact, it is not materially misstated.

When planning an audit, auditors normally try to control the first risk and ignore the second. This is because, should an auditor conclude that a population is materially misstated, entity officials will normally conduct an investigation to determine if the auditor is correct. This follow-up work would normally lead the auditor to the correct conclusion.

Confidence level

The *confidence level* is the degree of assurance that the auditor has that the sample is representative of the population from which it was selected. This is the converse of the sampling risk.

If the auditor uses a 90% confidence level, this means that there is a 90% chance that the sample will be representative of the population from which it was selected, and that the audit results will be correct. Put another way, there is a 10% chance that the sample is not representative of the population, and therefore the auditor may not reach a correct conclusion from the results of the work. For example, the auditor may conclude that the population does not contain a material error when, in fact, it does.

Precision gap widening and basic precision

Planned precision is the materiality amount less the expected aggregate error for the financial statements as a whole.

To illustrate using our example:

Materiality	Rs. 3,000,000
Expected aggregate error in financial statements	816,500
Planned precision	Rs. 2,183,500

When planning a statistical sample, though, there is one other factor that needs to be taken into account – precision gap widening.

The reason why we need to consider precision gap widening is because, for each additional Rs. 1 in the MLE, the UEL increases by more than Rs. 1. Simply subtracting the expected aggregate error from materiality does not deal with this effect. Therefore, planned precision needs to be reduced by a further amount. This further amount is referred to as precision gap widening.

Planned precision less precision gap widening is referred to as "basic precision". It is equal to the error that could exist in the population even if no errors were found in the sample. It therefore represents the UEL when the MLE is nil.

Basic precision and precision gap widening are calculated automatically by some CAATS.

B.3 Factors affecting sample size

Possible factors and impact

Various factors will affect the sample size, as illustrated in the following table:

Factor	Impact on Sample Size if Factor Increases	Comments
Population value	Increase	If population value increases with all other factors remaining the same, materiality and planned precision become smaller percentages of the population value. Hence, the auditor would need a more precise estimate of the error in the population. This would require a larger sample size.
Population size	Nil, except for very small populations	See discussion in Section B.2
Variability of sampling units	Nil for the types of sampling discussed in this annex	Variability is only a factor for those types of sampling plans based on a standard deviation. These types of sampling plans are rarely used in practice.

Materiality	Decrease	If materiality increases while all other factors remain the same, materiality and planned precision become larger percentages of the population value. Hence the auditor would not need to have as precise an estimate of the error in the population. The auditor could then decrease the required sample size.
Planned precision	Decrease	Same discussion as materiality.
Expected aggregate error	Increase	The expected aggregate error is subtracted from the materiality amount to arrive at planned precision. Increasing the expected aggregate error decreases planned precision, which increases the required sample size.
Confidence level	Increase	Increasing the confidence level means that the auditor wants to be more certain about the results of his/her procedure. The auditor will need to take a larger sample to achieve this.
Sampling risk	Decrease	Increasing the sampling risk is the same as decreasing the confidence level. The auditor is willing to be less certain about the results of his/her procedure, and can therefore take a smaller sample.

B.4 Sample selection

This section considers the basic sample selection rules and methods. It also illustrates how these are applied in the case of (i) MUS for substantive tests of details, (ii) MUS for compliance tests (tests of internal control), and (iii) attribute sampling for compliance tests (tests of internal control).

The difference between statistical sampling and non-statistical sampling is the method of selecting the sample items. All of the planning requirements remain the same, and the evaluation process remains the same.

There are two basic sample selection rules:

1. The sample conclusion only applies to the population from which it is selected; and

2. The sample should be representative of the population from which it is selected.

The rule in (1.) applies equally to statistical and non-statistical sampling.

The auditor has a better chance of achieving (2.) with a statistical sample than with a non-statistical sample. When using a non-statistical sample, though, the auditor should still strive to ensure the sample is as representative of the population as is possible.

There are several sample selection methods that are very good at ensuring that the sample is representative of the population from which it is selected, as follows:

1. Random;
2. Fixed interval (systematic);
3. Cell (random selection); and
4. Stratified random.

These are discussed below. The example used assumes that the auditor wishes to select 200 supplier invoices from a population of 30,000 supplier invoices.

For non-statistical sampling, the objective is to try to approximate one of these methods.

For both statistical and non-statistical sampling, there normally needs to be a complete listing of the valid transactions that adds up to the total amount reported on the financial statements.

To satisfy the validity objective, there needs to be a way in which the auditor can locate the individual items that have been selected from the listing.

To satisfy the completeness objective, there needs to be a way that the auditor can go from the individual items contained in boxes and filing cabinets to the listings that make up the total amount reported on the financial statements.

In some cases the listings used may be totals of other listings. In these cases, the auditor will first make a selection from among the totals, and will then make a second selection of individual transactions from the listing supporting each selected total.

Population from which to select

The first basic rule states that the sample conclusion only applies to the population from which it is selected. Therefore, if the auditor wants to rely on the internal control structure for the entire year, it would be best if the auditor sampled from the entire year. A less effective (but still generally accepted) method is to select a sample of transactions up to some interim date, reach a conclusion on that period, and then conduct inquiries, observations and walk-through procedures to reach a conclusion on the internal control structure for the rest of the year.

As the second approach is not as good as the first, it would normally not be acceptable when a high level of reliance is being placed on the internal control structure.

As another example, if the auditor selects a sample of transactions from only one month, the auditor's conclusion only applies to that one month – the auditor really does not have any assurance with respect to the other 11 months.

Another unacceptable approach is block or cluster sampling. Here, for example, the auditor may select all revenue receipts in a month, audit them, and attempt to conclude as to the revenue receipts for the entire year based on this "sample". In reality, the auditor has not done any sampling at all, and cannot reach a conclusion on the periods of time not covered by the test.

Random selection

Random selection involves numbering all of the items in the population and then using a random number table or software programme to select 200 random numbers. The auditor then identifies the sampling unit that corresponds to each number.

This method is difficult to use unless the sampling units are already pre-numbered (pre-numbered sales invoices, for example) or can easily be numbered (30 supplier invoices per page and the pages are numbered, for example).

Fixed interval (systematic) selection

This method involves selecting a random start and then every *n*th item. In our example, the auditor could select every 150th supplier invoice – 30,000 divided by 200. The random start would be a number between 1 and 150. If, say, the auditor picked a random start of 50, he/she would select the 50th item, the 200th item, the 350th item, etc.

Cell (random interval) selection

This method essentially combines the previous two methods. The auditor divides the population into cells and then picks a random item from within each cell. In our example, the first cell would contain the first 150 items, the second cell items 151 to 300, the third cell items 301 to 450, etc.

Stratified random selection

CAATS tools may offer a fourth method – stratified random sampling. Using this approach, the population is first stratified based on monetary ranges, type of expenditure, etc., and then a random sample is drawn from each range. This could be used, for example, to weight an attribute sample to the larger dollar items or specific expenditure types, or to ensure that at least one sample item is drawn from each expenditure type.

Non-statistical selection

Here the auditor uses judgement to approximate one of the methods described above. If done with care, this can be an acceptable way to select a sample. However, it would be prudent to increase the sample size by 20 to 50 percent to compensate for the fact that the sample may not be truly representative. The size of the increase would depend on how close the auditors believed they had come to approximately a statistical sample.

B.5 Monetary unit sampling (MUS) for substantive tests of details

Introduction

The key feature of MUS is that the population is considered to be composed of individual monetary units, as opposed to physical transaction vouchers like individual supplier invoices, cash disbursements, etc. The auditor selects an individual Rupee from the population and uses it as a hook to catch the voucher in which it occurs, so it can be audited.

To illustrate, let's return to our population of 30,000 supplier invoices that had a population value of Rs. 100,000,000. When using MUS, the auditor would consider the population to be composed of 100 million individual Rupees, as opposed to 30,000 invoices.

Under MUS, all sampling units (individual Rupees) will have the same chance of being selected. This means that, the larger the supplier invoice, the greater the chance of it being selected. This is why MUS is sometimes referred to as *sampling proportionate to size*.

If, for example, fixed interval selection is being used and the sampling interval is Rs. 1,000,000, then:

(g) A Rs. 100,000 invoice would have a 10% chance of selection;

(h) A Rs. 500,000 invoice would have a 50% chance of selection; and

(i) A Rs. 1,000,000 invoice (or larger) would have a 100% chance of selection.

Determining the sample size

The material in this section applies equally to statistical and non-statistical sampling. In this respect, it should be noted that the auditing standards in several developed countries now require the auditor to consider the same factors regardless of whether statistical sampling or non-statistical sampling is being used. As noted previously, though, when non-statistical sampling is being used, it may be appropriate to increase the sample size by 20% to 50%.

Assume the auditor wishes to take a sample of supplier invoices from a population of 30,000 supplier invoices with a total value of Rs. 100,000,000. The auditor needs to know:

1. The population value;
2. The materiality amount for the financial statements as a whole;
3. The expected aggregate error in the financial statements; and
4. The confidence level to be used for the test.

Assuming there are no individually significant transactions that the auditor wants to examine (very large transactions or high risk transactions), the population value in this case is Rs. 100,000,000. The materiality amount and the expected aggregate error are for the financial statements as a whole and therefore would be determined at the start of the planning process. For this example, we will use a Rs. 3,000,000 materiality amount and a Rs. 816,500 expected aggregate error.

The confidence level represents the amount of assurance that the auditor wishes to derive from the substantive test of details. The auditor normally determines this by using the audit risk model. Using that model, the confidence level is the converse of the substantive test of details risk (STDR) that results from the use of the audit risk model. For example, if STDR is determined to be 15%, the auditor will use an 85% confidence level for his/her sampling procedures.

While guidelines should not replace the use of professional judgement, the following may be useful:

Sources of Assurance	Possible Confidence Level
Very high level of assurance desired from the substantive test of details (because, for example, it is the only source of assurance)	95%
High level of assurance desired from the substantive test of details (because, for example, only limited reliance is being placed on the internal control structure and on analytical procedures)	90%
Moderate level of assurance desired from the substantive test of details (because, for example, moderate reliance is being placed on the internal control structure and/or on analytical procedures)	80%
Low level of assurance desired from the substantive test of details (because, for example, high reliance is being placed on the internal control structure and/or on analytical procedures)	70%

Let's assume that the auditor wants a very high level of assurance from the substantive test of details and selects a 95% confidence level. Using CAATS, the auditor would key in:

1. Confidence level (95);
2. Population value (100,000,000);
3. Materiality (3,000,000); and
4. Expected aggregate errors (816,500).

Assume the CAATS software will then produce a sample size of 181.

CAATS also offers the auditor several additional options – splitting the expected aggregate error between sampling applications and other audit tests, assuming that all errors found in the sample will be less than 60%, and changing the basic precision pricing. All of these will reduce the sample size. However, they need to be used with caution. A description of their use is beyond the scope of this annex.

Selecting the sample – statistical sampling

The "standard" sample selection technique with MUS is fixed interval (systematic) selection. Cell (random interval) selection can also be used if the population has been downloaded into a computer and a CAATS software tool is being used to do the selection.

Random selection is also possible, but contains all of the difficulties of cell selection. In addition, it has a further disadvantage – should fixed interval or cell selection be used, the sample size will automatically be adjusted for any over or underestimations of the population value. With random selection, this will not occur. Hence random selection is rarely used with MUS.

For both fixed interval selection and cell selection, the auditor needs to know the sampling interval. This is simply the population value (Rs. 100,000,000 in our example) divided by the sample size (181 in our example). The resulting interval is Rs. 553,350.

Note: some CAATS do not round the sample size to the nearest whole number when calculating the sampling interval. Thus the sampling interval may be slightly different than you get if you do the arithmetic manually.

Using CAATS for fixed interval selection, the auditor will key in:

1. The interval (Rs. 553,350 in our example); and

2. A random starting point (in our example a number between 1 and 553,350).

The CAATS will then select the 181 items.

For cell selection using CAATS, the auditor keys in:

1. The interval (Rs. 553,350 in our example); and

2. A "seed" number (a number to be used for the random number generator).

CAATS will then select the sample of 181 items. This sample will most likely *not* contain the same 181 items as were selected using fixed interval selection.

For both fixed interval selection and cell selection, the auditor does not need to key in the population value or the sample size – only the sampling interval. This is good because the auditor may not know the population value at the time he/she determines the sample size.

In our illustration, for example, the auditor is selecting a sample of supplier invoices for the year. The auditor may have estimated the Rs. 100,000,000 population value at the planning stage, which may occur well before the end of the year. The auditor may also wish to take a sample of, say, the first 8 months and then another sample of the last 4 months. By having the CAATS tool select every 553,350th Rupee (or create cells of that width), the sample size will automatically be adjusted for any overestimations or underestimations of the population value.

Selecting the sample – non-statistical sampling

To select a non-statistical sample that approximates a monetary unit sample, the auditor needs to find a way to bias the sample towards the larger monetary items. Some form of value-oriented selection is therefore required.

With a monetary unit sample, any item in the population greater than the sampling interval will automatically be selected, any item with a value of 80% of the aver0000age sampling interval will have an 80% chance of being selected, any item with a value of 50% of the average sampling interval will have a 50% chance of being selected, and so on.

Therefore, to approximate a monetary unit sample, the auditors could divide the population to be sampled into ranges, and then select an appropriate percentage of items from each range. For example, let's assume that there are approximately 10 items in the population with a value greater than the sampling interval, 40 items in the population with a value of 80% to 100% of the sampling interval, 110 items in the population with a value of 60% to 80% of the average sampling interval, and so on. The auditors could then randomly select all of the items in the first group, 90% (36) of the items in the second group, 70% (77) of the items in the next group, and so on.

To further approximate a monetary unit sample, the auditor should also select the sample from throughout the year. For example, when selecting the 36 items in the second group, the auditor could select three from each month.

If the auditor has a good knowledge of the population, the above approach would most likely result in a good approximation of a monetary unit sample, and increasing the statistical sample size by only 20% would most likely be sufficient. If, on the other hand, the auditor only had a limited knowledge of the population and had to guess at the monetary distribution, then it would most likely be prudent to increase the statistical sample size by 50%.

Sample evaluation – most likely error

The material in this section applies equally to statistical and non-statistical sampling. In this respect, it should be noted that the auditing standards in several developed countries now require the auditor to project the sample results over the population and to consider further possible errors, regardless of whether statistical sampling or non-statistical sampling is being used.

Let’s assume that the auditor tests the 181 items and finds 5 overstatement errors, as follows:

	Book Value	Audit Value	Error
1	600	nil	600
2	1,000	300	700
3	3,000	1,200	1,800
4	1,800	1,200	600
5	2,000	1,700	300
	8,400	4,400	4,000

The auditor has a *known error* of Rs. 4,000. However, he/she has only selected a sample of 181 supplier invoices out of the 30,000 supplier invoices in the population. The auditor needs to determine the *most likely error* (MLE) in the population.

To arrive at the MLE, MUS uses a concept called *tainting*. Tainting is the percentage by which each physical unit is in error. In our example, we have the following:

	Book Value	Audit Value	Error	Error % (Tainting)
1	600	nil	600	100.00
2	1,000	300	700	70.00
3	3,000	1,200	1,800	60.00
4	1,800	1,200	600	33.33
5	2,000	1,700	300	15.00
	8,400	4,400	4,000	278.33

What we have, then, are 5 items in the sample having errors totalling 278.33% and 176 items in the sample having no errors. The average percentage error in the sample is therefore 1.53775% (278.33% ÷ 181).

Assuming the sample is representative of the population, then the average percentage error in the population would be 1.53775%, and the MLE for the population as a whole would be Rs. 1,537,750 (Rs. 100,000,000 x .0153775).

Many CAATS packages use the same theory but rearrange the formula somewhat. In the example above we took the sum of the taintings, divided by the sample size, and multiplied by the population value, as follows:

$$\text{MLE} = \frac{\text{Sum of taintings}}{\text{Sample Size}} \times \text{Population Value}$$

$$= \frac{2.7833}{181} \times \text{Rs. } 100{,}000{,}000$$

$$= \text{Rs. } 1{,}537{,}750.$$

Rearranging this formula:

$$\text{MLE} = \text{Sum of Taintings} \times \frac{\text{Population Value}}{\text{Sample Size}}$$

$$= \text{Sum of Taintings} \times \text{Sampling Interval.}$$

To use a CAATS, the auditor creates an Excel (or other acceptable) spread sheet containing the book values and audited values for each item in error. The auditor then imports the file into the CAATS and keys in:

1. The confidence level (95);
2. The sampling interval (553,350); and
3. The sample size (181).

The CAATS then calculates the MLE of Rs. 1,540,150.

This approach produces a MLE of Rs. 1,540,150 (2.7833 x Rs. 553,350). The slight difference between the manually calculated solution and the CAATs result (Rs. 1,537,750 vs. Rs. 1,540,150) is due to the fact that the CAATS does not round the sample size to the nearest whole number when calculating the sampling interval.

Sample evaluation – upper error limit – overstatements

The MLE of Rs. 1,540,150 is the auditor's best estimate of the error in the population. However, because the auditor has only taken a sample, the actual error in the population could be larger or smaller than Rs. 1,540,150.

MUS is designed to produce an UEL as opposed to a range. To get a range, the auditor does two error evaluations – one for overstatement errors and one for understatement errors. The auditor then combines the results of the two evaluations.

The information imported or keyed in by the auditor is all the information the CAATS needs to calculate the UEL. In our example, all five errors were overstatement errors. In addition to calculating the MLE of Rs. 1,540,150, CAATS calculates an UEL of Rs. 3,584,850.

The error evaluation indicates that the auditor is 95% confident that the population is most likely overstated by Rs. 1,540,150, and that the population is not overstated by more than Rs. 3,584,850. As the materiality amount was Rs. 3,000,000, this is an unacceptable result.

The reason for the unacceptable result is that the MLE of Rs. 1,540,1500 is larger than the Rs. 816,500 expected aggregate error that was allowed for when planning the audit. As a general rule, if the MLE found in the sample exceeds the expected aggregate error allowed for when determining the sample size, then the results will be unacceptable.

Section 9.7 describes how the auditor deals with unacceptable results.

The error evaluation produces two additional amounts – basic precision of Rs. 1,644,040 and precision gap widening of Rs. 400,660. These amounts make up the Rs. 2,044,700 difference between the MLE of Rs. 1,540,150 and the UEL of Rs. 3,584,850.

Error evaluation – upper error limit – understatements

As noted above, the evaluation of understatements is done separately from evaluation of overstatements. The approach used is identical, and the data keyed in are identical.

The example includes no understatement errors so the MLE for understatement errors is nil. The UEL will be Rs. 1,644,040, being the basic precision amount.

Error evaluation – upper error limit – combining overstatements and understatements from the same test

To arrive at a net MLE and a net UEL, the results of the two separate evaluations are combined, as follows:

Net MLE	=	MLE (over) - MLE (under)
Net UEL (over)	=	UEL (over) - MLE (under)

Net UEL (under)	=	UEL (under) - MLE (over)

In our example we have:

Net MLE	=	MLE (over) - MLE (under)
	=	Rs. 1,540,150 - nil
	=	Rs. 1,540,150.
Net UEL (over)	=	UEL (over) - MLE (under)
	=	Rs. 3,584,850 - nil
	=	Rs. 3,584,850.
Net UEL (under)	=	UEL (under) - MLE (over)
	=	Rs. 1,644,040 – Rs. 1,540,150
	=	Rs. 103,890.

CAATS does this netting automatically.

The result is that the auditor can conclude with 95% confidence that:

1. The population is most likely overstated by Rs. 1,540,150;
2. The population is not overstated by more than Rs. 3,584,850; and
3. The population is not understated by more than Rs. 103,890.

As materiality is Rs. 3,000,000, this is again an unacceptable result.

Section 9.7 describes how the auditor deals with unacceptable results.

B.6 MUS for compliance tests

Introduction

"Compliance test" is another term for a test of internal control.

MUS for compliance testing is not used as frequently as attribute sampling for compliance tests. However, it has a significant advantage over attribute sampling – unlike attribute sampling, larger transactions have a greater chance of selection using MUS. In effect, the auditor is taking the position that internal control deviations (unapproved supplier invoices, for example) in large monetary items are more significant than internal control deviations in small monetary items.

Determining the sample size

The material in this section applies equally to statistical and non-statistical sampling. Note that the auditing standards in several developed countries now require the auditor to consider the same factors regardless of whether statistical sampling or non-statistical sampling is being used. When non-statistical sampling is used, it may be appropriate to increase the sample size by 20% to 50%.

For MUS for compliance testing, the information that the auditor needs to key into CAATS is somewhat different than for substantive testing, as follows:

MUS for Substantive Test of Details	MUS for Compliance Test
Population value	Population value
Materiality	Materiality
Expected aggregate error	Tolerable number of internal control deviations (in the sample)
Confidence level	Confidence level
- N/A-	Multiplier

An *internal control deviation* is the result of the failure to properly perform a particular internal control. For example, if a DDO is supposed to approve a particular document, then failure to properly approve the document would be an internal control deviation.

The *multiplier* used for test of compliance reflects the fact that not all internal control deviations result in a monetary error. Just because a supplier invoice is not properly approved, for example, does not mean that it is incorrect. The higher the number of internal control deviations that can occur before a monetary error occurs, the higher the number of internal control deviations that the auditor can tolerate. The multiplier is used to take this fact into account.

The reason why the multiplier is not applicable for substantive testing is because every substantive error is a monetary error, so the multiplier would always be 1.0.

What is a reasonable estimate of the multiplier for compliance tests (tests of internal control)? We know that the multiplier will be greater than 1 because, if every internal control deviation resulted in a monetary error, the auditor would really be performing a substantive test as opposed to a compliance test. Similarly, if the multiplier was really high (say around 20 or more), we would be saying that the failure of the internal control rarely results in a monetary error. In that case, there is really no need for the auditor to test the internal control at all. Accordingly, auditors normally set the multiplier at a conservative low number – 3.

Assume that the internal control deviation in which we are interested is the improper approval of the supplier invoices. Using this example and the multiplier of 3, we are saying that, if Rs. 9,000,000 worth of supplier invoices are not properly approved, the internal control will have failed often enough to permit errors aggregating to more than Rs. 3,000,000 to occur in the recorded amount.

For the *tolerable number of internal control deviations*, a low number (1 or 2) is often used. This is because there is no point testing an internal control that we know is not working well simply to prove that it cannot be relied on. Auditors therefore normally only test internal controls that are expected to be working well, and these are the controls that have a low internal control deviation rate. Auditors rarely use "zero" as the tolerable number of internal control deviations.

The *confidence level* relates to how much assurance the auditor wishes to obtain from his/her specific compliance test.

As noted previously, the confidence level for a substantive sample is the converse of substantive test of details risk in the audit risk model. For compliance testing, this is *not* the case – the confidence level is *not* the converse of control risk in the audit risk model. This is because:

1. usually several key internal controls need to work together to prevent or detect material error; and

2. The auditor is also usually performing non-sampling procedures as well, such as inquiries, observations and walk-through procedures.

While guidelines should not replace the use of professional judgement, the following may be useful:

Level of Reliance	**Possible Confidence Level**
High level of reliance on the specific internal control (control risk set at low – 20%)	95%
Moderate level of reliance on the specific internal control (control risk set at moderate – 50%)	90%
Low level of reliance on the specific internal control (control risk set at high – 80%)	80%

Assume the auditor decides to place moderate reliance on a particular internal control (approval of the supplier invoices in our example), wishes to use the "standard" multiplier of 3, and allows for one internal control deviation. Using the CAATS software, the auditor would key in:

1. Confidence level (90);

2. Population value (100,000,000);

3. Materiality (3,000,000);

4. Multiplier (3); and

5. Tolerable # of internal control deviations (1).

This results in a sample size of 43.

Selecting the sample

The sample selection for MUS for compliance testing is identical to the sample selection for MUS for substantive testing illustrated previously.

Sample evaluation

To evaluate a MUS for compliance test, the auditor uses the same evaluation software as for attribute sampling. This is illustrated below.

B.7 Attribute sampling for compliance tests

Introduction

As noted earlier, a compliance test is the same as a test of internal control.

Attribute sampling involves the selection of a sample of physical units (supplier invoices, for example), as opposed to individual monetary units (Rupees).

Because the auditor is selecting individual supplier invoices as opposed to individual Rupees, all physical units, regardless of their size, will have an equal chance of selection. In effect, the auditor is taking the position that internal control deviations (unapproved supplier invoices, for example) in small monetary items are just as significant as internal control deviations in large monetary items. If this not the case, the auditor should stratify the population and have more than one sample, or should use MUS for compliance tests.

Determining the sample size

The material in this section applies equally to statistical and non-statistical sampling. The auditor should consider the same factors regardless of whether statistical sampling or non-statistical sampling is being used. However, when non-statistical sampling is used, the auditor should increase the sample size by 20% to 50%.

The main consideration in selecting a sample is controlling the risk of incorrectly accepting the population (called *beta risk*). There is a complementary risk of incorrectly rejecting the population (called *alpha risk*).

In this Appendix we only consider beta risk. Auditors normally do not concern themselves with the risk of incorrectly rejecting the population.

The information that the auditor needs to key into a CAATS for attribute sampling is somewhat different from what the auditor keys into CAATS for MUS, as follows:

MUS for Compliance Testing	Attribute Sampling for Compliance Testing
Population value	Population size
Materiality	Tolerable deviation rate
Tolerable number of internal control deviations	Expected deviation rate

Confidence level	Confidence level
Multiplier	- N/A -

Auditors normally do not know the *population size* – the number of physical units in the population. However, for the purposes of determining the sample size, this is not normally a problem because the population size only affects the sample size when the population size is less than 10,000 units.

While knowing the population size may not be required to *determine* the sample size, it is required to *select* the sample using either fixed interval selection or cell selection. Both these methods require the auditor to determine the sampling interval which, in the case of attribute sampling, requires the auditor to know the population size.

The *tolerable deviation rate* represents the maximum percentage of internal control deviations that can occur in a population before errors aggregating to more than materiality occur. For example, if the auditor used 9%, he/she would be saying that, if more than 2,700 supplier invoices (9% of the 30,000 supplier invoices) are not properly approved, the internal control will have failed often enough to permit errors aggregating to more than Rs. 3,000,000 to occur in the recorded amount.

The tolerable deviation rate has a relationship to the population value, the materiality amount, and the multiplier. To illustrate:

1. If the population value doubled then, on average, each of the 30,000 supplier invoices would have twice the value. One would logically expect that it would then take fewer than 2,700 improperly approved supplier invoices to allow errors aggregating to Rs. 3,000,000 to occur. The tolerable deviation rate would therefore decrease.

2. If materiality was increased from Rs. 3,000,000 to Rs. 6,000,000, one would logically expect that it would then take more than 2,700 improperly approved supplier invoices to allow errors aggregating to materiality to occur. The tolerable deviation rate would therefore increase.

3. If it took 6 improperly approved supplier invoices to produce a monetary error instead of 3 improperly approved invoices, one would logically expect that it would then take more than 2,700 improperly approved invoices to allow errors aggregating to Rs. 3,000,000 to occur. The tolerable deviation rate would therefore increase.

Generally accepted auditing standards in Pakistan and most other countries do not require the auditor to relate the tolerable deviation rate to the population value, the materiality amount or the multiplier. However, if the auditor wishes to do so, the following formula could be used:

$$\text{Tolerable Deviation Rate} = \frac{\text{Materiality x Multiplier}}{\text{Population Value}}$$

Using the data used earlier, we arrive at a Tolerable Deviation Rate of 9%, determined as follows:

$$\text{Tolerable Deviation Rate} = \frac{\text{Materiality x Multiplier}}{\text{Population Value}}$$

$$= \frac{3{,}000{,}000 \text{ x } 3}{100{,}000{,}000}$$

$$= 9\%.$$

For the *expected deviation rate*, a low rate, such as 1%, is often used. For the *confidence level*, the same considerations apply as for MUS, discussed earlier.

Assuming the auditor decides to place moderate reliance on a particular internal control, has set the tolerable deviation rate at 9% (as above) and the expected deviation rate at 1%. Using CAATS, the auditor would key in:

(j) Population size (30,000);

(k) Tolerable deviation rate (9);

(l) Expected deviation rate (1); and

(m) Confidence level (90).

This then produces a sample size of 42 – virtually the same as the 43 calculated in the MUS for compliance testing example. This is because, by using a 9% tolerable deviation rate, we are effectively using the same parameters.

It should be noted that, when using CAATS, any expected deviation rate between 0 and 2.38% would result in the same sample size. This is because internal control deviations are virtually always whole numbers – most internal controls are either performed or they are not performed. (You cannot have an invoice that is partly properly approved, for example.) Any error rate up to 2.38 percent allows for one internal control deviation in a sample of 42.

Selecting the sample – statistical sampling

The sample selection technique most often used for attribute sampling is *random selection*. *Fixed interval selection* is sometimes used. A third option, *cell selection*, exists but is rarely used.

Both fixed interval selection and cell selection require the auditor to know the sampling interval, which, in turn, requires an estimate of the population size. As such, they are not used that often. Another weakness of fixed interval selection is that there may be a pattern in the population of which the auditor is not aware. For example, there may be 19 employees for each supervisor, and the payroll register may list the supervisor followed by the employees for each group. If the auditor were to select every 20^{th} employee, he/she will have a sample that contained either no supervisors or nothing but supervisors.

With MUS, the risk of this occurring is remote. The total payroll of each group would have to be same, and would have to be equal to the sampling interval for this to happen.

To select a random sample using CAATS, the auditor would key in:

1. The number of records to select (the sample size);

2. A random number "seed" (a number to be used for the random number generator); and

3. The starting record number and the ending record number (to produce a range for the random numbers). The difference between the two would be the population size.

If the auditor does not have an exact estimate of the population size, he/she could key in a conservative estimate and discard unused sample items. The auditor, though, would have to increase the sample size to ensure that the resulting sample was adequate.

For both fixed interval selection and cell selection, the data that the auditor is required to key in is the same as for MUS for substantive tests of details.

Selecting the sample – non-statistical sampling

As noted above, the "standard" sample selection technique for attribute sampling is random selection. Fixed interval selection is sometimes used; cell selection is rarely used.

The non-statistical approximation of a random sample is often referred to as "haphazard selection".

The various ways in which the auditor could approximate a statistical sample are discussed below. The illustrations assume that the auditor increases his/her statistical sample size from 42 to 60 to compensate for not taking a statistical sample.

Periodic selection. The auditor could divide the sample size by 52 to determine how many sample items should be selected from each week of the year, and then select the appropriate number of items from the transactions processed in each of those weeks. If there are fewer than 52 sample items, the auditor could use half-months or months.

In our illustration (a sample size of 60), the auditor could select 1 sample item from most weeks, and 2 sample items from the other weeks.

DDO selection. The auditor could assign numbers to the Drawing and Disbursing Officers (DDOs) responsible for the transactions being sampled, haphazardly select the required number of DDOs, and then select a transaction from each of the selected DDOs. To better approximate a statistical sample, the auditor could then select the transactions from throughout the year using the periodic approach.

For example, if there are 900 DDOs and the auditor wishes to select a sample of 60 transactions, the auditor could haphazardly select 60 of those DDOs and then select 1 transaction from each of the selected DDOs. The sample selection could be done is such a way that 1 sample item is selected from most weeks, and 2 sample items are selected from the other weeks for a total of 60 transactions.

Page selection. The auditor could divide the number of pages on which the transactions are listed by the sample size, and then select the appropriate number of items from each group of pages.

For example, if there are 300 pages of transactions and the auditor wishes to select a sample of 60 transactions, the auditor could select one sample item from each fifth page, or from each group of five pages.

Selecting from boxes, filing cabinets, etc. The auditor could divide the sample size by the number of boxes or filing cabinets of transactions in which the documents are stored, and the then select the appropriate number of items from each of the boxes or filing cabinets. If there are fewer sample items than boxes or filing cabinets, the auditor could group the boxes or filing cabinets into groups, and then select a transaction from each group of boxes or filing cabinets.

For example, if there are 600 boxes, the auditor could group the boxes into sets of 10, and then select 1 transaction from each set of boxes. To better approximate a statistical sample, the auditor could select the transactions from throughout the year using the periodic approach.

Sample evaluation

The material in this section applies equally to statistical and non-statistical sampling. Auditing standards in several developed countries now require the auditor to project the sample results over the population and consider further possible errors, regardless of whether statistical sampling or non-statistical sampling is being used.

Some CAATS offers two options when evaluating the sample results: a one-sided evaluation in which an upper limit is determined, and a two-sided evaluation in which both an upper limit and a lower limit are determined.

In this Appendix, we will only deal with the former option. Auditors normally do not concern themselves with the lower limit.

When evaluating a monetary unit sample, the auditor determines if the upper error limit in the population exceeds the materiality amount. When evaluating an attribute sample, the auditor determines if the maximum possible deviation rate in the population exceeds the tolerable deviation rate.

Assume the auditor finds 2 deviations in the sample of 42. Using CAATS, the auditor would key in:

1. Population size (30,000);
2. Sample size (42);
3. Number of deviations in sample (2); and
4. Confidence level (90).

The CAATS then calculates a maximum possible deviation rate of 12.18%.

This indicates that the auditor has 90% confidence that the internal control will function often enough to ensure that, at most, 12.18% of the supplier invoices contain internal control deviations. However, the auditor had concluded that a deviation rate above 9% could produce errors aggregating to more than materiality. Therefore, the sample results are unacceptable and the auditor must *normally* reduce reliance on the internal control.

By how much must the auditor reduce reliance on internal control? Some CAATs tools provide the auditor with an indication of this. The sample size calculation shows the achieved confidence level that will be associated with several different numbers of internal control deviations. For example, in our illustration the CAATS informs the auditor that, should he/she find 2 internal control deviations in a sample of 42, the achieved confidence level would be 74.17%. Therefore, an alternative conclusion is that the auditor has 74.17% confidence that the internal control will function often enough to ensure that, at most, 9% of the supplier invoices contain internal control deviations. However, the auditor wanted to have 90% confidence. Therefore, the auditor must *normally* reduce his/her reliance on the internal control.

As a general rule, if the most likely error found in a monetary unit sample exceeds the expected aggregate error allowed for when determining the sample size, then the results will be unacceptable. For attribute sampling, we have an equivalent rule – as a general rule, if the actual number of deviations found in the sample exceeds the number of deviations allowed for when determining the sample size, then the results will be unacceptable.

In our illustration, we allowed for 1 deviation and found 2. Therefore, we knew that the results would be unacceptable.

This is why auditors usually allow for at least one internal control deviation in their samples. While auditors will get smaller sample sizes if they do not allow for any internal control deviations, the discovery of a single deviation will produce unacceptable results.