Sampling Guidance

on how to design audit sampling strategies under the EEA and Norwegian Financial Mechanisms for 2014-2021

Version 2
March 2021

TABLE OF CONTENTS

List of Acronyms

AA – Audit Authority

AE – Anticipated Error

BV – Book Value (expenditure in reference period)

EE – Projected Error

EEA – European Economic Area

ER – Error rate

MUS – Monetary Unit Sampling

PPS – Probability Proportional to Size

SE – (Effective accuracy, i.e., after performing audit work)

SRS - Simple random sampling

TE – Tolerable error

UEL – Upper Error Limit

1 Introduction

The present guidance on sampling has been prepared to provide audit authorities an overview of the main methods and options that can be followed in the development of substantive testing when designing audit procedures for the EEA and Norwegian Financial Mechanisms for 2014-2021. This guidance document is provided for information purposes only and is not part of the legal framework of the EEA and Norwegian Financial Mechanisms.

This guidance has been inspired by the EU Guidance on sampling methods for audit authorities Programming periods 2007-2013 and 2014-2020[1] and aims at assisting the audit authorities in preparing their audit strategies and in their audit work for all types of programmes and types of costs. The guidance proposes efficient strategies in line with the applicable regulatory framework and the international auditing standards. In particular, the guide is framed by:

- Regulation on the implementation of the European Economic Area (EEA) Financial Mechanism 2014-2021;
- Regulation on the implementation of the Norwegian Financial Mechanism 2014-2021;
- Financial guidance on how to carry out financial management and reporting under the EEA and Norwegian Financial Mechanisms for 2014-202;
- and recommendations of the *international standard on auditing 530 audit sampling*.

Complementary reading of these additional documents is advised in order to get a complete view of the guidelines.

[1] https://ec.europa.eu/regional_policy/en/information/publications/guidelines/2017/guidance-on-sampling-methods-for-audit-authorities-programming-periods-2007-2013-and-2014-2020. The present Guideline can be read together with the EC Guideline for examples and other detailed information.

2 Concepts related to sampling for audit

2.1 Assurance/confidence level for the audit

Most of the audit work required under the EEA and Norwegian Financial Mechanisms will require a substantial part of substantive testing. This applies to audit of projects, management costs, costs under technical assistance or bilateral funds.

Substantive testing should be performed using sampling, the size of which will depend on the confidence level determined based on the assurance level obtained from the control system testing, mainly:

- not less than 60% if assurance is high;
- between 70% and 80% in case of average assurance;
- not less than 90% if assurance is low.

The recommended relationship between the system's risk and the confidence level used in substantive testing is shown in the table below:

Risk assessment of control system	Assurance level from the system audits	Confidence level for substantive testing
Minimal	Works well, only minor improvements are needed	Not less than 60%
Low	Work, but some improvements are needed	70%
Moderate	Works partially, substantial improvements are needed	80%
High	Essentially does not work	Not below 90%

Table 1. Confidence level for the substantive testing according to the risk assessment of the control system.

The actual confidence level to be obtained from substantive testing can deviate from the examples identified in the tables above, however these should be considered as minimum requirements at the given level of risk assessment of the control system.

2.2 Materiality

Information is material if its omission or misstatement could influence the economic decision of users taken based on the financial statements.

A recommended materiality level of 2% maximum is applicable to the actual incurred expenditure in the reference period. The materiality is used:

- As a threshold to compare the projected error in expenditure;
- To define the tolerable/acceptable error that is used for determining sample size.

In case of pre-financing model, the Audit Authorites should take a population of actual incurred expenditure reported by project promoters to PO in projects financial reports.

2.3 Sampling and selection method

The sampling method encompasses two elements:

1. the sampling design (e.g. equal probability, probability proportional to size) and
2. the projection (estimation) procedure.

These two elements provide the framework to calculate sample size.

The most well-known sampling methods suitable for substantive testing are presented in Section 4. These include statistical and non-statistical sampling.

A statistical sampling method has the following characteristics:

- each item in the population has a known and positive selection probability;
- randomness should be ensured by using proper random number generation;
- sample size is calculated in such a way that allows to achieve a certain level of desirable accuracy.

Due to distributional assumptions, the minimum sample size for statistical sampling should be set at 30 items.

Statistical sampling methods allow the selection of a sample that is "representing" the population that allows the calculation of accuracy and the control of audit risk. The final goal is to extrapolate or estimate to the population, the value of a parameter (the "variable") observed in a sample, allowing to conclude whether a population is materially misstated or not and, if so, by how much (an error amount).

Non-statistical sampling does not allow the calculation of accuracy, consequently there is no control of the audit risk and it is impossible to ensure that the sample is representing the population. Therefore, the error must be assessed empirically.

Statistical sampling is therefore the preferable approach in substantive testing. Non-statistical methods should be used in situations where the effort associated with statistical methods would be disproportionated due to its application to very small population sizes. (cf. sections 4.1 and 4.7).

Even when using non-statistical methods, the selection of the units to the sample should always be based on probabilistic selection (i.e. based on the generation of random numbers). Consequently, non-random selection (e.g. judgment selection, risk-based selection) are not to be used under substantive testing.

2.4 Projection (estimation) and accuracy assessment

The final goal when applying a sampling method is to extrapolate or estimate the level of error observed in the sample to the whole population. This process will allow to conclude whether a population is materially misstated or not and, if so, by how much (an error amount).

The act of projection is called estimation and the value calculated from the sample (projected value) is called **the estimate**. This estimate, only based on a fraction of the population, is affected by an error called the sampling error.

Sampling error represents, in fact, the uncertainty in the projection of results to the population. A measure of this error is usually called **accuracy which** depends mainly on the **sample size**, **population variability** and to a lesser degree the **population size**.

Planned accuracy should be differentiated from effective accuracy. Planned accuracy is the maximum planned sampling error for sample size determination (usually is the difference between maximum tolerable error and the anticipated error and it should be set to a value lower than the materiality level). The effective accuracy is an indication of the difference between the sample projection (estimate) and the true (unknown) population parameter (value of error) and represents the uncertainty in the projection of results to the population obtained from sample data.

The tolerable error is the maximum error that can be accepted in the population for a certain reference period. With a 2% materiality level this maximum tolerable error is therefore 2% of the expenditure declared for that reference period.

The planned accuracy is the maximum sampling error accepted for the projection of errors in a certain reference period, i.e. the maximum deviation between the true population error and the projection produced from sample data. It should be set by the auditor to a value lower than the tolerable error, because otherwise the results of substantive testing will have a high risk of being inconclusive.

The most adequate way to settle the planned accuracy is to calculate it as equal to the difference between the tolerable error and the anticipated error (the projected error that the auditor expects to obtain at the end of the audit). This anticipated error will be based on the auditor professional judgment and supported by the evidence gathered in the auditing work in previous periods for the same of similar population or in preliminary/pilot sample.

The choice of a realistic anticipated error is important, since the sample size is highly dependent on the value chosen for this error. See also Section 2.10.

Section 4 presents detailed formulas to use in the sample size determination process.

2.5 Population

For substantive testing, the AA should ensure that the population to sample is **complete** and includes all items (i.e. expenditure).

The population to sample includes the expenditure declared in the reference period[2]. For projects expenditure this means that only actual incurred expenditure by projects should be included in the sampling population. In case pre-financing model is used for the payments, the AA should take a sample from project actual incurred expenditure reported. Other types of expenditure such as proposed expenditure or the pre-financing payments to projects included in the interim financial reports submitted to the FMO should not be included in the reference population.

Although it is mandatory that all items in the population have a chance to be selected to the sample, once one item (e.g. project) is selected in the sample, it is not mandatory to fully audit all its expenditure. Whenever the selected item (e.g. project) includes a large number of expenditure items (e.g. invoices), **the AA may apply two-stage sampling**, as briefly explained in section 4.8.

The nature of 4 different types of costs leads to the consideration of 4 different populations:

1. projects expenditure
2. programme management costs,
3. funds for bilateral relations,
4. technical Assistance to the Beneficiary State

These 4 types of populations may be grouped together for the projection of errors under the restrictions and strategies presented in Section 3.6.

[2] Although the reference period is not set by regulation, it is recommended to use the period from 1 July of year *t*-1 to 30 June of year *t*.

2.6 Grouping of programmes/populations

The AA may decide to group several programmes or populations into a single group[3]. If this option is followed the group of programmes/populations is treated as one single population. In this case, errors will be projected the whole group and not to each programme individually and audit conclusions will be drawn to the whole group of programmes/populations. Consequently, the audit opinion will be formed over the whole group of programmes.

The audit authority should apply **one** assurance level in the case of grouping of programmes.

In case the systems audit reveal that within the group of programmes there are differences in the conclusions on the functioning of the various programmes, it is recommended to apply the lowest assurance level (associated with the functioning of control systems) obtained at the individual programme level, for the whole group of programmes.

When grouping programmes or other populations it is recommended to use stratification in order to ensure the representation of all programmes/populations in the final sample. Refer to Section 2.8 for a brief description on stratification.

Whenever the AA decides to group 2 or more programmes/populations, it may calculate sample sizes in 2 different ways:

- Calculate an independent sample size for each programme/population in the group; the final sample size is the sum of all sample sizes obtained for each population;
- To calculate a single global sample size for the group of programmes/populations using a stratified methodology; after the global sample size is calculate is to be allocated between strata according to the rules presented in Chapter 4.

2.7 Sampling units

Different types of main sampling units should be considered according to the different types of populations.

- **Projects expenditure**

The AA may opt to use either projects or payment claims as the main sampling unit.
In case two-stage sampling is used, the subsample (secondary) units will be the expenditure items inside the projects (expenditure documents, invoices, etc).

[3] Different reasons that may suggest the grouping of programmes. For example, they may be related with the need to increase population size in order to avoid the use of non-statistical sampling or the goal of reducing sample size due to the fact that one single sample will be selected in opposition to the selection of several samples (one per programme). Grouping of programmes is possible even if the programmes do not share the same management and control system.

- **Programme management costs**

Sampling units are expenditure items directly selected from the programme operator or operators (in the case of a group of programmes).

- **Bilateral funds (Funds for bilateral relations)**

For sampling purposes bilateral funds costs may be separated into 3 groups:

a) bilateral funds associated with specific programmes and managed by programme operators;
b) bilateral fund available at national level and managed by the national focal point;
c) bilateral funds that correspond to projects (in case national focal point uses part of national fund to support specific projects).

Sampling units are, in any case, expenditure items directly selected from:

a) the programme operator or operators (in the case of a group of programmes);
b) the national focal point;
c) the project beneficiaries.

- **Technical assistance**

For technical assistance the sampling unit are expenditure items selected from the 3 beneficiary entities (Focal Point, Certifying Authority and Audit Authority).

2.8 Stratification

Stratification corresponds to segment the population into sub-populations called strata and to select independent samples from each stratum.

The main goal of stratification is two-folded: (1) it improves the accuracy (for the same sample size) or allows the auditor to reduce the sample size (for the same level of accuracy); (2) it ensures that the subpopulations corresponding to each stratum are represented in the sample.

Whenever the AA expects that the level of error will be different for different groups in the population (e.g. by programme, risk of the project), the stratification is valuable.

Different sampling methods can be applied to different strata. For example, it is common to apply a 100% audit of the high-value items and apply a statistical sampling method to audit a sample of the remaining lower-value items that are included in the additional stratum or strata. This is useful if the population includes a few quite high-value items, as it lowers the variability in each stratum and can therefore improve the accuracy or allow the auditor to reduce the sample size.

Stratification is also useful when grouping programmes or different populations. In this case, specific strata for each population should be foreseen in order to ensure that all programmes/populations are represented in the sample.

Th recommended minimum sample size for one particular stratum, whatever is its nature, is 3 to 5 units.

2.9 Evaluation of results

The confidence interval is the interval that contains the true (unknown) population value (error) with a certain probability (called confidence level). The confidence interval's general formula is as follows:

$$[EE - SE; EE + SE]$$

where

- EE represents the projected or extrapolated error;
- SE represents the accuracy (sampling error);

The projected/extrapolated error (EE) and the Upper Error Limit (EE+SE) are the two most important instruments to conclude whether a population is materially misstated or not. Of course, the UEL can only be calculated when statistical sampling is used; hence, for non-statistical sampling the EE is always the best estimate of the error in the population.

When statistical sampling is used, the following situations can arise:

- If EE is larger than the materiality threshold (hereafter 2%, for simplification), then the AA concludes that there is material error;
- If EE is lower than 2% and the UEL is lower than 2%, the AA concludes that the population is not misstated by more than 2% at the specified level of sampling risk.
- If EE is lower than 2% but the UEL is larger than 2%, the AA concludes that additional work is needed. Accordingly, to the INTOSAI guideline n° 23, the additional work can include:
  - *"requesting the audited entity to investigate the errors/exceptions found and the potential for further errors/exceptions. This may lead to agreed adjustments in the financial statements;*
  - *carrying out further testing with a view to reducing the sampling risk and thus the allowance that has to be built into the evaluation of results;*
  - *using alternative audit procedures to obtain additional assurance."*

In most cases where an UEL is well above 2% this could be prevented or minimized if the AA considers a realistic anticipated error when calculating the original sample size (see section 2.10, for more details).

2.10 Sampling parameters

The most important sampling parameters to consider when calculating sample size are:

- Confidence level;
- Anticipated error; and
- Anticipated standard-deviation

Confidence level

The confidence level is set in line with the assessment of the risk of the control systems as proposed in Section 2.1.

Anticipated error

The anticipated error can be defined as the amount of error the auditor expects to find in the population. Factors relevant to the auditor's consideration of the expected error include the results of the test of controls, the results of audit work from previous periods and the results of other substantive procedures. The more the anticipated error differs from the true error, the higher the risk of reaching inconclusive results after performing the audit is (EE <2% and UEL > 2%).

To set the value of the anticipated error the auditor should take into consideration:

1. Information on the error rates of previous periods. The anticipated error should, in principle, be based on the projected error obtained in the previous period. Nevertheless, if the auditor has received information about changes in the quality of the control systems, this information can be used either to reduce or increase the anticipated error. For example, if for the last period the projected error rate was 0.7% and no further information exists, this value can be imputed to the anticipated error rate. However, if the auditor has received convincing evidence about an improvement of the systems which would lead to a lower error rate in the current period, this information can be used to reduce the anticipated error to a smaller value of, for example, 0.4%.
2. If there is no historical information about error rates, the auditor can use a preliminary/pilot sample to obtain an initial estimate of the error rate of the population. The anticipated error rate is considered to be equal to the projected error from this preliminary sample. Pilot samples are suggested to have a sample size of 30 as this is the minimum threshold for statistical sampling.
3. If there is no historical information to produce an anticipated error and a preliminary sample cannot be used due to uncontrollable restrictions, then the auditor should set a value to the anticipated error based on professional experience and judgment. The value should mostly reflect the auditor expectation regarding the true level of error in the population.

In summary, the auditor should use historical data, auxiliary data, professional judgement or a mix of the above to set a value as realistic as possible for the anticipated error.

An anticipated error based on objective quantitative data is usually more accurate and avoids carrying out additional work when audit results are inconclusive. For example if the auditor sets an anticipated error of 10% of materiality, i.e. 0.2% of expenditure, and at the end of the audit he/she obtains a projected error of 1.5%, results will most probably be inconclusive as the upper limit of error will be higher than the materiality level. To avoid these situations the auditor should use as anticipated error, in future sampling exercises, the most realistic possible measure of the true error in the population.

Also, the Audit Authority should plan its audit work in a way to achieve sufficient accuracy even when the anticipated error is well above materiality (i.e. equal or above 4.0%). In this case it is advisable to compute the sample size formulas with an anticipated error resulting in a maximum planned accuracy of 2.0%, i.e. by imputing the anticipated error to be equal to 4.0%.

Anticipated standard-deviation

The anticipated standard-deviation is obtained with the same approach as proposed for the anticipated error.

3 Sampling strategies

3.1 Introduction

Despite the specific sampling method that is selected, sampling activities for substantive testing always follow a basic common structure:

1. **Define the objectives of the substantive tests**: usually the determination of the level of error in the expenditure declared for a given reference period based on a projection from a sample;
2. **Define the population**: actual expenditure incurred for a given period, and the **sampling unit**, which is the item to be selected to the sample;
3. **Define population parameters**: this includes defining the tolerable error (2% of the expenditure declared), the anticipated error expected by the auditor, the confidence level (considering the audit risk model) and (usually) a measure of population variability;
4. **Determine the sample size**, according to the sampling method used. It is important to note that the final sample size is always rounded up to the nearest integer[4];
5. **Select the sample and perform the audit**; and
6. **Project results calculate accuracy and draw conclusion**: this step covers the computation of the accuracy and projected error and comparing these results with the materiality threshold.

Each sampling method (cf. Section 4) provides specific formulas to compute the sample size, to project errors and to assess accuracy.

The consideration of the 4 types of populations presented in Section 2.5, will lead to different possible strategies as explained in the following sections.

3.2 Projects

For the expenditure declared by projects, the reference population is the programme or group of programmes if applied, i.e. errors are project to the respective programme or group at programmes level.

The main sample is based on projects or payments claims. If two-stage sampling is chosen, the secondary sampling units will correspond to expenditure items inside the projects. Samples may be stratified both at the main stage or at the level of the secondary items based on any criteria considered relevant by the AA.

[4] In case the sample size is calculated for different strata and periods, it is acceptable that the sample sizes for some strata/periods are not rounded up provided that the general sample size is rounded up.

3.3 Management costs

For management costs of a single programme, the strategy should be based on selecting a sample of expenditure items directly from the single programme operator.

Errors will be projected to the category of management costs of the specific programme.

The projected error will be added to the projected error resulting from auditing the declared expenditure of projects of the same programme and will contribute the global error/opinion of the respective programme.

For management costs of a group of programmes, the strategy should be based on selecting a sample of expenditure items:

- either directly selected from the whole group of programme operators; or
- stratified by operator.

Errors will be projected to the category of management costs of the group of programmes.

The projected error will be added to the projected error resulting from auditing the declared expenditure of projects in the same group of programmes and will contribute the global error/opinion of the respective group of programmes.

3.4 Funds for bilateral relations

Bilateral funds are treated as a separate programme for which a separate error/opinion is required.

The sampling strategy should be based on the selection of expenditure items:

- either selected from the whole set of expenditure items;
- or stratified by the 3 potential beneficiaries (Focal Point, Programme operators and Project beneficiaries).

Further stratification can also be implemented:

- for bilateral funds associated with programme operators, additional stratification by programme operator can be implemented;
- for the bilateral funds associated with projects, additional stratification by project can be implemented.

The first level of stratification (by Focal Point, Programme operators and Project promoters) is highly recommended.

In any case errors will be projected to the category of bilateral funds to produce one single audit opinion (cf. Section 3.6).

3.5 Technical assistance

Technical assistance is treated as a separate programme for which a separate error/opinion is produced.

The sampling strategy should be based on the selection of expenditure items:

- either selected from the whole set of expenditure items;
- or stratified by the 3 beneficiary entities (Focal Point, Certifying Authority and Audit Authority).

3.6 Total error

The total projected error for a specific programme is obtained by adding the projected error of declared expenditure and the projected error of the respective management costs:

- **Total projected error of programme = projected error of project declared expenditure + projected error of management costs**

As explained in Section 2.6 the size of the sample used to assess the total error of a programme, may be obtained in two different ways:

- A bottom-up approach where sizes for the sample of project declared expenditure and for management costs are calculated separately (considering these 2 components as two different populations); in this case the global sample size is just the sum of the two sample sizes;
- A top-down approach where a global sample size is calculated for the whole population composed by project declared expenditure and management costs, using a stratified approach; after the global sample size is calculated it has to be allocated between the two strata using any of the methods proposed in Section 4[5]. A simplified methodology to implement the top-down approach is to consider the management costs expenditure as if it were an additional project that is mandatorily included in the sample and audited.

Also, as explained in Section 2.1.2 programmes may be grouped together for the sake of projecting errors. Should this option be followed, the projected errors of the declared expenditure and of management costs for all programmes within the group will be added in order to find the total projected error of the group of programmes. Considering a group of *k* programmes:

- **Total projected error of group of programmes =** $\sum_{i=1}^{k}\left(\textbf{projected error of project declared expenditure of programme } \boldsymbol{i} + \textbf{projected error of management costs of programme } \boldsymbol{i}\right)$

[5] Other allocation methods are available and may be used.

Again, both a bottom-up or top-down approach can be followed when grouping the samples of project expenditure and management costs.

Technical assistance and bilateral relations fund are to be treated as 2 separate programmes for which specific errors are projected. Nevertheless, the AA is free to group together these 2 populations in order to produce one single projected error:

- **Total projected error for BF + TA = projected error of bilateral funds (BF) + projected error of technical assistance (TA)**

If the AA decides to group bilateral relations and technical assistance costs the sample size may be calculated either using the bottom-up or top-down approaches presented earlier in this section.

In summary the most desegregated way errors can be projected is:

- At programme level (including project expenditure and management costs);
- Technical assistance;
- Bilateral relations fund.

The more aggregated way error can be projected is:

- Group of all programmes (including project expenditure and management costs);
- Group of technical assistance and bilateral relations funds.

3.7 Audit opinion

Independently on the number of independent projected errors that the AA is able to produce, the audit authority will have to issue one global (consolidated) audit opinion including the whole expenditure of the reference period, despite its origin (type of costs)[6].

To support the global audit opinion the AA will have to aggregate the projected errors and precisions produced at programme or group of programmes level, in one single error rate and one single precision (and upper error limit).

Technical details how to calculate global error rate and precision are presented in Section 4.10.

[6] In line with art. 5.5, 1, e) of the Regulation on the implementation of the European Economic Area (EEA) Financial Mechanism 2014-2021.

4 Sampling methods

4.1 Conditions of applicability

The choice of sampling method is based on multiple criteria. From a statistical point of view, the choice is mainly based on the expectation regarding the variability of errors and their relationship with the expenditure.

The table below gives some indications on the recommended methods[7] and the relevant criteria.

Sampling Method	**Favourable conditions**
Standard MUS	Errors have high variability[8] and are approximately proportional to the level of expenditure (i.e. error rates are of low variability) The values of expenditure per item show high variability
Simple random sampling	General proposed method that can be applied when the previous conditions do not hold Can be applied using mean-per-unit estimation or ratio estimation
Non-statistical methods	If the application of statistical method is impossible or leads to disproportionate audit effort due to small population size
Stratification	Can be used in combination with any of the above methods It is particularly useful whenever the level of error is expected to vary significantly among population groups (subpopulations)

Table 2. Favourable conditions for the choice of sampling methods

Stratification can be used in combination with any sampling method. Stratification allows the partition of the population in more homogeneous groups (strata) with less variability than the whole population. Instead of having a population with high variability it is possible to have two or more subpopulations with lower variability. Stratification should be used to either **minimise variability or isolate error-generating subsets of the population**. In both cases stratification will reduce the sample size.

Statistical sampling should be used to draw conclusions on the amount of error in a population. However, there are special justified cases where a non- statistical sampling

[7] Other sampling methods and particularly the ones presented in the EU Guidance on sampling methods (e.g. conservative MUS, difference estimation) may be used and the AA is recommended to follow internationally accepted guidelines.

[8] High variability means the errors across items are not similar, that is, there are small and large errors in contrast with the case where all the errors are more or less of similar values.

method may be used on the professional judgement of the audit authority, in accordance with internationally accepted audit standards.

In practice, the specific situations that may justify the use of non-statistical sampling are related to the population size. For example, in case of a very small population, the size of the population might not allow the use of statistical methods (the population is smaller or very close to the recommended sample size, cf. Section 4.7).

The audit authority must use all possible means to achieve a sufficiently large population: by grouping programs, and/or by using as the unit the beneficiaries' payment claims. The AA should also consider that even if statistical approach is not possible in the beginning of the programming period, it should be applied as soon as it is feasible.

4.2 Notation

Before presenting the main sampling methods for substantive testing it is useful to define a set of concepts related to sampling that are common to all the methods.

Thus:

- z is a parameter from the normal distribution related to the confidence level determined from system audits. The possible values of z are presented in the following table

Confidence level	**60%**	**70%**	**80%**	**90%**	**95%**
System assurance level	High	Moderate	Moderate	Low	No assurance
z	0.842	1.036	1.282	1.645	1.960

Table 3. Values of *z* by confidence level

- N is the population size (e.g. number of items in a programme or group of programmes); if the population is stratified, an index h is used to denote the respective stratum, $N_h, h = 1,2, \dots, H$ and H is the number of strata;
- n is the sample size; if the population is stratified, an index h is used to denote the respective stratum, $n_h, h = 1,2, \dots, H$ and H is the number of strata;
- TE be the maximum tolerable error, that is, 2% of the total expenditure declared (the Book Value, BV);
- $BV_i, i = 1,2, \dots, N$ is the book value (the expenditure declared) of an item;
- $E_i, i = 1,2, \dots, N$, is the amount of error of an item; if the population is stratified an index h is used to denote the respective stratum, $E_{hi}, i = 1,2, \dots, N_h, h = 1,2, \dots, H$ and H is the number of strata;
- AE is the anticipated error defined by the auditor based on the expected level of error (e.g. an anticipated error rate times the Total expenditure at the level of the population). AE can be obtained from historical data (projected error in past

period) or from a preliminary/pilot sample of low sample size (the same used to determine the standard deviation).
- σ represents the standard deviation (depending on the used sampling methods the standard deviation may be calculated over the errors or the error rates)

The above-mentioned parameters are often accompanied in the guidance by specific subscripts which could relate to the character of the parameter or a stratum that the parameter refers to. In particular:
- r is used with standard deviation when it refers to standard deviation of error rates;
- e refers to exhaustive stratum/high value stratum; if used with standard deviation this notation could also refer to standard deviation of errors (as opposed to standard deviation of error rates);
- w is used with standard deviation when a weighted value is used;
- s refers to a non-exhaustive stratum;
- h refers to a stratum.

If a parameter is accompanied by several subscripts, they could be used in different order without changing the meaning of the notation.

4.3 Simple random sampling

4.3.1 Introduction

Simple random sampling (SRS) is a statistical sampling method. It is the most well-known among the equal probability selection methods. This method aims at projecting the level of error observed in the sample to the whole population.

Items in the sample are selected randomly with equal probabilities. Simple random sampling is a generic method that fits different types of populations. Although, as it does not use auxiliary information, it usually requires larger sample sizes than MUS (whenever the level of expenditure varies significantly among items and there is positive association between expenditure and errors). The projection of errors can be based on two sub-methods: mean-per-unit estimation or ratio estimation.

As all other methods, this method can be combined with stratification.

4.3.2 Sample size

The sample size n for simple random sampling is based on the following information:
- Population size N;
- Confidence level determined from systems audit and the related coefficient z from a normal distribution (see Section 4.2);
- Maximum tolerable error TE (usually 2% of the total expenditure);

- Anticipated error AE chosen by the auditor according to professional judgment and previous information; and
- The standard deviation σ_e of the errors.

The sample size is computed as follows[9]:

$$n = \left(\frac{N \times z \times \sigma_e}{TE - AE}\right)^2$$

where σ_e is the standard-deviation of errors in the population. The standard-deviation of errors for the total population is assumed to be known in the above calculation. In practice, this will almost never be the case and audit authorities will have to rely either on historical data (standard-deviation of the errors for the population in the past period(s)) or on a preliminary/pilot sample of low sample size (sample size is recommended to be not smaller than 30 units).

If the sampling parameters and the audit assumptions are confirmed by the error evaluation of the pilot sample, the pilot sample can become the actual sample tested (minimum 30 items). Note that, in any case, the pilot sample can subsequently be used as a part of the sample chosen for audit (Cf. Section 4.9).

4.3.3 Projected error

There are two possible ways to project the sampling error to the population. The first is based on mean-per-unit estimation (absolute errors) and the second on ratio estimation (error rates).

Mean-per-unit estimation (absolute errors)

Multiply the average error per item observed in the sample by the number of items in the population, yielding the projected error:

$$EE_1 = N \times \frac{\sum_{i=1}^{n} E_i}{n}.$$

Ratio estimation (error rates)

[9] When dealing with a small population size, i.e. if the final sample size represents a large proportion of the population (as a rule of thumb more than 10% of the population) a more exact formula can be used leading to $n = \left(\frac{N \times z \times \sigma_e}{TE - AE}\right)^2 \Big/ \left(1 + \left(\frac{\sqrt{N} \times z \times \sigma_e}{TE - AE}\right)^2\right)$. This correction is valid for simple random sampling and for difference estimation. It can also be introduced in two steps by calculating the sample size n with the usual formula and sequentially correct it using $n' = \frac{n \times N}{n + N - 1}$.

Multiply the average error rate observed in the sample by the book value at the level of the population:

$$EE_2 = BV \times \frac{\sum_{i=1}^{n} E_i}{\sum_{i=1}^{n} BV_i}$$

The sample error rate in the above formula is just the division of the total amount of error in the sample by the total amount of expenditure of units in the sample (expenditure audited).

It is not possible to know *a priori* which is the best extrapolation method as their relative merits depend on the level of association between errors and expenditure.

As a basic rule of thumb, the second method should just be used when there is the expectation of high association between errors and expenditure (higher value items tend to exhibit higher errors) and the first method (mean-per-unit) when there is an expectation that errors are relatively independent from the level of expenditure (higher errors can be found either in units of high or low level of expenditure). In practice this assessment can be made using sample data as the decision about the extrapolation method can be taken after the sample is selected and audited. To select the most adequate extrapolation method one should use the sample data to calculate the variance of the book values of the sample units (VAR_{BV}) and the covariance between the errors and the book values over the same units ($\text{COV}_{\text{E,BV}}$). Formally, the ratio estimation should be selected whenever $\frac{\text{COV}_{\text{E,BV}}}{\text{VAR}_{\text{BV}}} > \text{E}R/2$, where ER represents the sample error rate, i.e. the ratio between the sum of errors in the sample and the audited expenditure. Whenever the previous condition is not verified the mean-per-unit estimation should be used to project the errors to the population[10].

4.3.4 Accuracy

Accuracy (sampling error) is a measure of the uncertainty associated with the projection (extrapolation). It is calculated differently according to the method used for extrapolation.

Mean-per-unit estimation (absolute errors)

The accuracy is given by the following formula

$$SE_1 = N \times z \times \frac{s_e}{\sqrt{n}}$$

where s_e is the standard-deviation of errors in the sample (now calculated from the same sample used to project the errors to the population)

[10] An alternative to this methodology is to calculate accuracy for both projection methods and choose the one with the most favourable accuracy.

$$s_e^2 = \frac{1}{n-1}\sum_{i=1}^{n}(E_i - \bar{E})^2$$

Ratio estimation (error rates)

The accuracy is given by the following formula

$$SE_2 = N \times z \times \frac{s_q}{\sqrt{n}}$$

where s_q is the sample standard deviation of the variable q:

$$q_i = E_i - \frac{\sum_{i=1}^{n} E_i}{\sum_{i=1}^{n} BV_i} \times BV_i.$$

This variable is for each unit in the sample computed as the difference between its error and the product between its book value and the error rate in the sample.

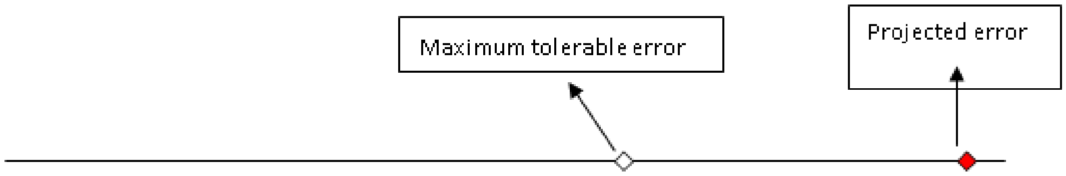
4.3.5 *Evaluation*

To draw a conclusion about the materiality of the errors, the upper limit of error (UEL) should be calculated. This upper limit is equal to the sum of the projected error EE itself and the accuracy of the extrapolation

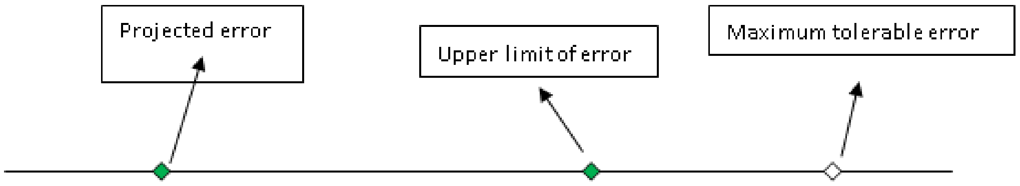
$$ULE = EE + SE$$

Then the projected error and the upper limit should both be compared to the maximum tolerable error to draw audit conclusions:

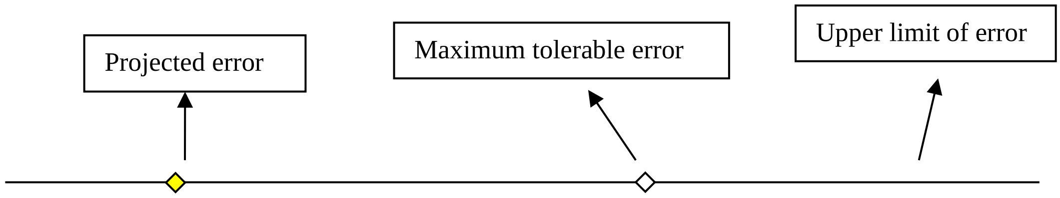
- If projected error is larger than the maximum tolerable error, it means that the auditor would conclude that there is enough evidence to support that errors in the population are larger than the materiality threshold:



- If the upper limit of error is lower than the maximum tolerable error, then the auditor should conclude that the errors in the population are lower than materiality threshold.



- If the projected error is lower than the maximum tolerable error but the upper limit of error is larger than the maximum tolerable error, this means that the sampling results may be inconclusive.



4.4 Stratified simple random sampling

4.4.1 Introduction

In stratified simple random sampling, the population is divided in sub-populations called strata and independent samples are drawn from each stratum, using the standard simple random sampling approach.

In stratification we aim to find groups (strata) with less variability than the whole population. With simple random sampling, the stratification by level of expenditure per item is usually a good approach, whenever it is expected that the level of error is associated with the level of expenditure. Other variables that we expect to explain the level of error in the items are also good candidates for stratification. Some possible choices are programmes, regions, classes based on the risk of the item, etc.

If stratification by level of expenditure is implemented, the AA should consider to identify a high-value stratum[11], apply a 100% audit of these items, and apply simple random sampling to audit samples of the remaining lower-value items that are included in the additional stratum or strata. This is useful when the population included a few high-value items. In this case, the items belonging to the 100% stratum should be taken out of the population and all the steps considered in the remaining sections will apply only to the population of the low-value items. It is not mandatory to audit 100% of the high-value stratum units. The AA may develop a strategy based on several strata, corresponding to different levels of expenditure, and have all the strata audited through sampling. If a 100% audited stratum exists, the planned accuracy for sample size determination should be however based on the total book value of the population. Indeed, as the only source of error is the low-value items stratum, but the planned accuracy refers to the population level, the tolerable error and the anticipated error should be calculated at population level, as well.

4.4.2 Sample size

The sample size is computed as follows

$$n = \left(\frac{N \times z \times \sigma_w}{TE - AE}\right)^2$$

where σ_w^2 is the weighted mean of the variances of the errors for the whole set of strata:

[11] There is not a general rule to identify the cut-off value for the high value stratum. A rule of thumb would be to include all items whose expenditure is larger than the materiality (2%) times the total population expenditure. More conservative approaches use a smaller cut-off usually dividing the materiality by 2 or 3, but the cut-off value depends on the characteristics of the population and should be based on professional judgment.

$$\sigma_w^2 = \sum_{i=1}^{H} \frac{N_h}{N} \sigma_{eh}^2, h = 1,2, \dots, H;$$

and σ_{eh}^2 is the variance of errors in each stratum. The variance of the errors is computed for each stratum as an independent population as

$$\sigma_{eh}^2 = \frac{1}{n_h^p - 1} \sum_{i=1}^{n_h^p} (E_{hi} - \bar{E}_h)^2, h = 1,2, \dots, H$$

where E_{hi} represent the individual errors for units in the sample of stratum h and $\bar{E}_h$ represent the mean error of the sample in stratum h.

These values can be based on historical knowledge or on a preliminary/pilot sample of low sample size as previously presented for the standard simple random sampling method. In this later case the pilot sample can as usual subsequently be used as a part of the sample chosen for audit. If no historical information is available in the beginning of a programming period and it is not possible to access a pilot sample, the sample size may be calculated with the standard approach.

Once the total sample size, n, is computed the allocation of the sample by stratum is as follows:

$$n_h = \frac{N_h}{N} \times n.$$

This is a general allocation method, usually known as proportional allocation. Many other allocation methods are available. A more tailored allocation may in some cases bring additional accuracy gains or reduction of sample size. The adequacy of other allocation methods to each specific population requires some technical knowledge in sampling theory. Sometimes, it may happen that the allocation method produces a very small sample size for one or more strata. In practice it is advisable to use a minimum sample size of 3 to 5 units for every stratum in the population in order to allow the calculation of the standard-deviations that are necessary to calculate accuracy.

4.4.3 Projected error

Based on H randomly selected samples of items, where the size of each one has been computed according to the above formula, the projected error at the level of the population can be computed through the two usual methods: mean-per-unit estimation and ratio estimation.

Mean-per-unit estimation

In each group of the population (stratum) multiply the average error per item observed in the sample by the number of items in the stratum (N_h); then sum all the results obtained for each stratum, yielding the projected error:

$$EE_1 = \sum_{h=1}^{H} N_h \times \frac{\sum_{i=1}^{n_h} E_i}{n_h}.$$

Ratio estimation

In each group of the population (stratum) multiply the average error rate observed in the sample by the population book value at the level of the stratum (BV_h):

$$EE_2 = \sum_{h=1}^{H} BV_h \times \frac{\sum_{i=1}^{n_h} E_i}{\sum_{i=1}^{n_h} BV_i}$$

The sample error rate in each stratum is just the division of the total amount of error in the sample of stratum by the total amount of expenditure in the same sample.

The choice between the two methods should be based upon the considerations presented for the standard simple random sampling method.

If a 100% stratum has been considered and previously taken from the population then the total amount of error observed in that exhaustive stratum should be added to the above estimate (EE_1 or EE_2) in order to produce the final projection of the amount of error in the whole population.

4.4.4 Accuracy

As for the standard method, accuracy (sampling error) is a measure of the uncertainty associated with the projection (extrapolation). It is calculated differently according to the method that has been used for extrapolation.

Mean-per-unit estimation (absolute errors)
The accuracy is given by the following formula

$$SE_1 = N \times z \times \frac{s_w}{\sqrt{n}},$$

where s_w^2 is the weighted mean of the variance of errors for the whole set of strata (now calculated from the same sample used to project the errors to the population):

$$s_w^2 = \sum_{i=1}^{H} \frac{N_h}{N} s_{eh}^2, h = 1,2, \dots, H;$$

and s_{eh}^2 is the estimated variance of errors for the sample of stratum h

$$s_{eh}^2 = \frac{1}{n_h - 1}\sum_{i=1}^{n_h}(E_{hi} - \bar{E}_h)^2\, , h = 1,2, \dots , H$$

Ratio estimation (error rates)

The accuracy is given by the following formula

$$SE_2 = N \times z \times \frac{s_{qw}}{\sqrt{n}}$$

where

$$s_{qw}^2 = \sum_{h=1}^{H}\frac{N_h}{N}s_{qh}^2$$

is a weighted mean of the sample variances of the variable q_h, with

$$q_{ih} = E_{ih} - \frac{\sum_{i=1}^{n_h} E_{ih}}{\sum_{i=1}^{n_h} BV_{ih}} \times BV_{ih}.$$

This variable is for each unit in the sample computed as the difference between its error and the product between its book value and the error rate in the sample.

4.4.5 Evaluation

To draw a conclusion about the materiality of the errors the upper limit of error (UEL) should be calculated. This upper limit is equal to the summation of the projected error EE itself and the accuracy of the extrapolation

$$ULE = EE + SE$$

Then the projected error and the upper limit should both be compared to the maximum tolerable error to draw audit conclusions using exactly the same approach presented in Section 4.3.

4.5 Monetary unit sampling

4.5.1 Introduction

Monetary unit sampling is the statistical sampling method that uses the monetary unit as an auxiliary variable for sampling. This approach is usually based on systematic sampling with probability proportional to size (PPS), i.e. proportional to the monetary value of the sampling unit (higher value items have higher probability of selection).

This method is particularly useful if book values have high variability and there is positive correlation between errors and book values. In other words, whenever it is expected that items with higher values tend to exhibit higher errors, situation that frequently holds in the audit framework.

Whenever the above conditions hold, i.e. book values have high variability and error are positively correlated (associated) with book values, then MUS tends to produce smaller sample sizes than equal probability-based methods, for the same level of accuracy.

Samples produced by this method will typically have an over representation of high value items and an under representation of low value items. This is not a problem by itself as the method accommodates this fact in the extrapolation process but makes sample results (e.g. sample error rate) as non-interpretable (only extrapolated results can be interpreted).

Similarly to probability-based methods, this method can be combined with stratification (favourable conditions for stratification are discussed in sections 2.7 and 4.1).

4.5.2 Sample size

The sample size *n* for monetary unit sampling is based on the following information:

- Population book value (total declared expenditure) *BV;*
- Confidence level determined from systems audit and the related coefficient z from a normal distribution (see Section 4.2);
- Maximum tolerable error *TE* (usually 2% of the total expenditure);
- Anticipated error *AE* chosen by the auditor according to professional judgment and previous information and
- The standard deviation σ_r of the error rates (produced from a MUS sample).

The sample size is computed as follows:

$$n = \left(\frac{z \quad \times BV \quad \times \sigma_r}{TE - AE} \right)^2$$

where σ_r is the standard-deviation[12] of error rates produced from a MUS sample. To obtain an approximation to this standard-deviation before performing the audit the AA will have to rely either on historical knowledge (variance of the error rates in a sample of past period) or on a preliminary/pilot sample of low sample size, n^p (sample size for the preliminary sample is recommended to be not less than 30 items.

If the sampling parameters and the audit assumptions are confirmed by the error evaluation of the pilot sample, the pilot sample can become the actual sample tested (minimum 30 items). In any case, and as usual, this sample can be subsequently used as a part of the full sample chosen for audit.

4.5.3 Sample selection

After determining the sample size, it is necessary to identify the high value population units (if any) that will belong to a high value stratum to be audited at 100%. The cut-off value for determining this top stratum corresponds to the ratio between the book value (BV) and the planed sample size (n). All items whose book value is higher than this cut-off (if $BV_i > BV/n$) will be placed in the 100% audit stratum.

The sampling size to be allocated to the non-exhaustive stratum, n_s , is computed as the difference between n and the number of sampling units (e.g. projects) in the exhaustive stratum (n_e).

Finally, the selection of the sample in the non-exhaustive stratum will be made using probability proportional to size, i.e. proportional to the item book values BV_i. A popular way to implement the selection is through systematic selection, using a sampling interval equal to the total expenditure in the non-exhaustive stratum (BV_s) divided by the sample size (n_s), i.e.

$$SI = \frac{BV_s}{n_s}$$

In practice, it may happen that after the calculation of the sampling interval based on the expenditure and sample size of the sampling stratum, some population units will still exhibit an expenditure larger than this sampling interval BV_s/n_s (although they have not previously exhibit an expenditure larger than the cut-off (BV/n). In fact, all items whose book value is still higher than this interval ($BV_i > BV_s/n_s$) have also to be added to the high-value stratum. If this happens, and after moving the new items to the high value stratum, the sampling interval has to be recalculated for the sampling stratum taking into consideration the new values for the ratio BV_s/n_s. This iterative process may have to be performed several times until no further units present expenditure larger than the sampling interval.

[12] Specific formulas for standard-deviations and other parameters like the mean or variance are not presented in this guidance. In case of interest th reader can refer to the EC guidance on sampling for audit.

4.5.4 *Projected error*

The projection of the errors to the population should be made differently for the units in the exhaustive stratum and for the items in the non-exhaustive stratum.

For the exhaustive stratum, that is, for the stratum containing the sampling units with book value larger than the cut-off, $BV_i > \frac{BV}{n}$, the projected error is just the summation of the errors found in the items belonging to the stratum:

$$EE_e = \sum_{i=1}^{n_e} E_i$$

For the non-exhaustive stratum, i.e. the stratum containing the sampling units with book value smaller or equal to the cut-off value, $BV_i \leq \frac{BV}{n}$ the projected error is

$$EE_s = SI \sum_{i=1}^{n_s} \frac{E_i}{BV_i}$$

To calculate this projected error:

1) for each unit in the sample calculate the error rate, i.e. the ration between the error and the respective expenditure $\frac{E_i}{BV_i}$
2) sum these error rates over all units in the sample
3) multiply the previous result by the sampling interval (SI)

The projected error at the level of population is the sum of these two components:

$$EE = EE_e + EE_s$$

4.5.5 *Accuracy*

The accuracy is given by the formula:

$$SE = z \times \frac{BV_s}{\sqrt{n_s}} \times s_r$$

where s_r is the standard-deviation of error rates in the sample of the non-exhaustive stratum (calculated from the same sample used to extrapolate the errors to the population)

The sampling error is only computed for the non-exhaustive stratum, since there is no sampling error to account for in the exhaustive stratum.

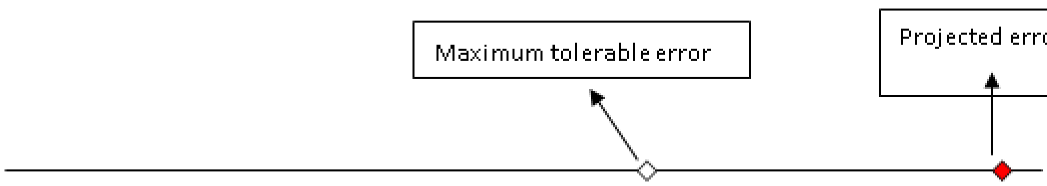
4.5.6 *Evaluation*

To draw a conclusion about the materiality of the errors, the upper limit of error (UEL) should be calculated. This upper limit is equal to the sum of the projected error EE itself and the accuracy of the extrapolation

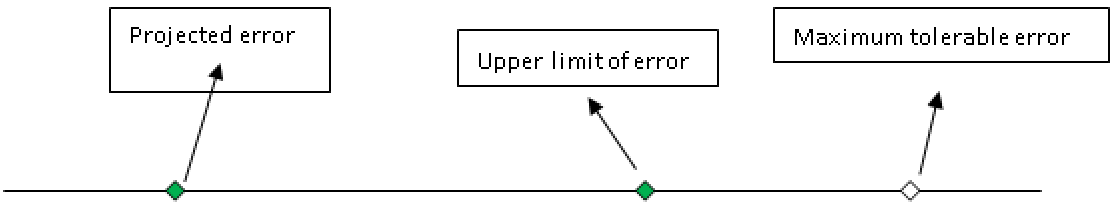
$$UEL = EE + SE$$

Then the projected error and the upper limit should both be compared to the maximum tolerable error to draw audit conclusions.

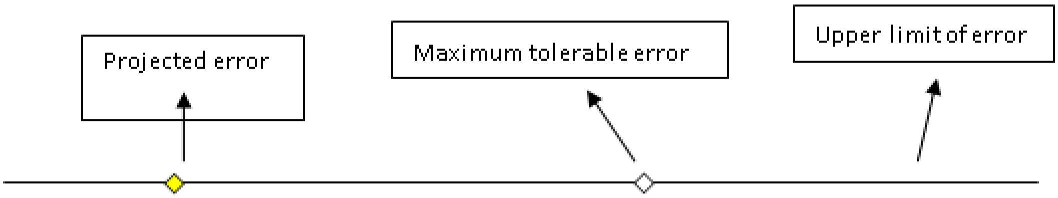
- If projected error is larger than maximum tolerable error, it means that the auditor would conclude that there is enough evidence to support that errors in the population are larger than materiality threshold:



- If the upper limit of error is lower than maximum tolerable error, then the auditor should conclude that errors in the population are lower than materiality threshold.



If the projected error is lower than maximum tolerable error but the upper limit of error is larger, results are considered inconclusive.



4.6 Stratified monetary unit sampling

4.6.1 Introduction

In stratified monetary unit sampling, the population is divided in sub-populations called strata and independent samples are drawn from each stratum, using the standard monetary unit sampling approach.

In stratified MUS, the stratification by level of expenditure is not relevant, as MUS already takes into account the level of expenditure in the selection of sampling units.

4.6.2 Sample size

The sample size is computed as follows:

$$n = \left(\frac{z \times BV \times \sigma_{rw}}{TE - AE}\right)^2$$

where σ_{rw}^2 is a weighted mean of the variances of the error rates for the whole set of strata, with the weight for each stratum equal to the ratio between the stratum book value (BV_h) and the book value for the whole population (BV).

$$\sigma_{rw}^2 = \sum_{i=1}^{H} \frac{BV_h}{BV} \sigma_{rh}^2, h = 1,2,\ldots,H;$$

and σ_{rh}^2 is the variance of error rates in each stratum[13].

As previously presented for the standard MUS method, these values can be based on historical knowledge or on a preliminary/pilot sample of low sample size. In this later case the pilot sample can subsequently be used as a part of the sample chosen for audit.

[13] For this calculation, whenever the book value of unit *i* (BV_i) is larger than the cut-off BV_h/n_h the ratio $\frac{E_i}{BV_i}$ should be substituted by the ratios $\frac{E_i}{BV_h/n_h}$.

When starting using the stratified MUS method for the first time, it may happen that historical stratified data is unavailable. In this case, sample size can be determined using the formulas for the standard MUS method (see Section 4.5). If no historical knowledge is available for the first period to audit, the sample size will be larger than in fact would be needed if that information were available. Nevertheless, the information collected in the first period of application of the stratified MUS method can be applied in future periods for sample size determination.

Once the total sample size, n, is computed the allocation of the sample by stratum is as follows:

$$n_h = \frac{BV_h}{BV} n.$$

This is a general allocation method, where the sample is allocated to strata proportionally to the expenditure (book value) of the strata. Other allocation methods are available. A more tailored allocation may in some cases bring additional accuracy gains or reduction of sample size.

4.6.3 Sample selection

In each stratum h, there will be two components: the exhaustive group inside stratum h (that is, the group containing the sampling units with book value larger than the cut-off value, $BV_{hi} > \frac{BV_h}{n_h}$); and the sampling group inside stratum h (that is, the group containing the sampling units with book value smaller or equal than the cut-off value, $BV_{hi} \leq \frac{BV_h}{n_h}$)

After determining sample size, it is necessary to identify in each of the original stratum (h) the high value population units (if any) that will belong to a high value group to be audited at 100%. The cut-off value for determining this top group is equal to the ratio between the book value of the stratum (BV_h) and the planed sample size (n_h). All items whose book value is higher than this cut-off (if $BV_{hi} > \frac{BV_h}{n_h}$) will be placed in the 100% audit group.

The sampling size to be allocated to the non-exhaustive group, n_{hs} , is computed as the difference between n_h and the number of sampling units (e.g. projects) in the exhaustive group of the stratum (n_{he}).

Finally, the selection of the samples is done in the non-exhaustive group of each stratum using probability proportional to size, i.e. proportional to the item book values BV_i. A common way to implement the selection is through systematic selection, using a selection interval equal to the total expenditure in the non-exhaustive group of the stratum (BV_{hs}) divided by the sample size (n_{hs}) [14], i.e.

$$SI_h = \frac{BV_{hs}}{n_{hs}}$$

Note that several independent samples will be selected, one for each original strata.

4.6.4 Projected error

The projection of errors to the population is made differently for units belonging to the exhaustive groups and for items in the non-exhaustive groups.

For the exhaustive groups, that is, for the groups containing the sampling units with book value larger than the cut-off value, $BV_{hi} > \frac{BV_h}{n_h}$, the projected error is the summation of the errors found in the items belonging to those groups:

$$EE_e = \sum_{h=1}^{H} \sum_{i=1}^{n_h} E_{hi}$$

In practice:

1) For each stratum h, identify the units belonging to the exhaustive group and sum their errors; and
2) Sum the previous results over the all set of *H* strata.

For the non-exhaustive groups, i.e. the groups containing the sampling units with book value lower or equal to the cut-off value, $BV_{hi} \leq \frac{BV_h}{n_h}$, the projected error is

$$EE_s = \sum_{h=1}^{H} \frac{BV_{hs}}{n_{hs}} \sum_{i=1}^{n_{hs}} \frac{E_{hi}}{BV_{hi}}$$

To calculate this projected error:

1) in each stratum h, for each unit in the sample calculate the error rate, i.e. the ratio between the error and the respective expenditure $\frac{E_{hi}}{BV_{hi}}$
2) in each stratum h, sum these error rates over all units in the sample
3) in each stratum h, multiply the previous result by the total expenditure in the population of the non-exhaustive group (BV_{hs}); this expenditure will also be equal to the total

[14] If some population units will still exhibit an expenditure larger than this sampling interval, then the procedure explained in section 6.3.1.3 shall be applied.

expenditure in the stratum minus the expenditure of items belonging to the exhaustive group
4) in each stratum h, divide the previous result by the sample size in the non-exhaustive group (n_{hs})
5) sum the previous results over the whole set of H strata

The projected error at the level of population is just the sum of these two components:

$$EE = EE_e + EE_s$$

4.6.5 Accuracy

As for the standard MUS method, accuracy is a measure of the uncertainty associated with the extrapolation. It represents the sampling error and should be calculated in order to subsequently produce a confidence interval.

The accuracy is given by the formula:

$$SE = z \times \sqrt{\sum_{h=1}^{H} \frac{BV_{hs}^2}{n_{hs}} . s_{rhs}^2}$$

where s_{rhs} is the standard-deviation of error rates in the sample of the non-exhaustive group of stratum h (calculated from the same sample used to extrapolate the errors to the population).

The sampling error is only computed for the non-exhaustive groups, since there is no sampling error arising from the exhaustive groups.

4.6.6 Evaluation

To draw a conclusion about the materiality of the errors the upper limit of error (UEL) should be calculated. This upper limit is equal to the summation of the projected error EE itself and the accuracy of the extrapolation

$$UEL = EE + SE$$

Then the projected error and the upper limit should both be compared to the maximum tolerable error to draw audit conclusions using exactly the same approach presented in Section 4.5.

4.7 Non statistical sampling

4.7.1 Introduction

A non- statistical sampling method may be used on the professional judgement of the AA, in duly justified cases, in accordance with internationally accepted audit standards and in any case, where the number of operations is insufficient to allow the use of a statistical method (Art 5.5, Point 4 of the Regulation).

Non-statistical sampling does not allow the calculation of accuracy, consequently there is no control of the audit risk and it is impossible to ensure that the sample represents the population (non-statistical sampling does not allow the calculation of accuracy, and consequently there is no control of the audit risk.). Therefore, the error has to be assessed empirically. As explained above, statistical sampling should be used, as a general rule, for substantive testing. The specific situations that may justify the use of non-statistical sampling are related to the population size. In fact, it may happen to work with small population, and for which the use of statistical methods would generate disproportionate audit efforts (the population is smaller or very close to the recommended sample size).

This means that it is not possible to state the exact population size below which non-statistical sampling is needed as it depends on several population characteristics, but one should consider the threshold to be between 50 and 100 population units. This means that for populations bellow 50 units the AA may freely decide to apply non-statistical sampling and that for populations above 100 units, statistical sampling should be applied. For populations with size between 50 and 100 units, the decision should be taken based on a case-by-case analysis. This decision should take into consideration the balance between the cost and benefit associated with each of the methods.

Before taking the decision to apply non-statistical sampling the AA should consider increasing population size either through the grouping of programmes or the change of sampling units to one with a lower level of granularity. For example, if there are 2 programmes, one with 44 projects and other with 47 projects, they tend to be too small to support the use of statistical sampling. Nevertheless, if the AA decides to group the 2 programmes in one single population, its size will be 91 projects, most probably allowing the use of statistical sampling.

Even in the situations where the AA applied a non-statistical sampling method, the sample shall be selected using a probabilistic method. The size of the sample must be determined taking into account the level of assurance provided by the system and must be sufficient to enable the AA to draw a valid audit opinion. **The AA should be able to extrapolate the results to the population from which the sample was drawn.**

When using a probabilistic method to select the sampling units, the only differences between a statistical and a non-statistical sample are due to:

1) The fact that in non-statistical sampling the sample size is determined empirically and without controlling the global audit risk;
2) Accuracy of the projections is not calculated, which means that conclusions will be drawn by comparing the projected error with the materiality threshold.

When using non-statistical sampling, the AA should consider stratifying the population by dividing it into sub-populations, each one being a group of sampling units with similar characteristics, in particular in terms of risk or expected error rate. Stratification is a very efficient tool to improve the quality of the projections and it is strongly recommended to use some kind of stratification in the framework of non-statistical sampling.

4.7.2 Sample size

In non-statistical sampling, the sample size is calculated using professional judgment and taking account the level of assurance provided by the control systems.

A non-statistical sample should cover a minimum of 10% of population items and also 15 % of the expenditure (Art. 5.5, Point 4 of the Regulation). Since the regulation refers to a minimum coverage, these thresholds correspond therefore to the 'best case scenario' of high assurance from the system. In line with annex 3 of the ISA 530, the higher the auditor's assessment of the risk of material misstatement, the larger the sample size needs to be. Also note that the requirement of 15% of expenditure refers to the expenditure effectively audited in the sample.

There is no fixed rule to select the sample size based on the assurance level from the system audits, but as a reference, the AA, when defining the sample size under non-statistical sampling, may consider the following indicative thresholds[15].

Risk assessment of control system	**Recommended coverage**	
	on number of items	**on expenditure declared**
Minimal	10%	15%
Low	12.5%	17.5%
Moderate	15%	20%

[15] These reference values may of course be changed according to the AA's professional judgment and any additional information it may have about the risk of material misstatement.

Risk assessment of control system	Recommended coverage	
	on number of items	**on expenditure declared**
High	20%	25%

Table 4. Recommended coverage for non-statistical sampling

4.7.3 Sample selection

The sample shall be selected using a probabilistic method. In particular, the selection can be made using either:

- equal probability selection (where each sampling unit has equal chance of being selected regardless of the amount of expenditure declared in the sampling unit), as in simple random sampling (cf. sections 4.3 and 4.4 for the reference to simple random sampling and stratified simple random sampling); or
- probability proportional to size (expenditure) using the monetary unit as an auxiliary variable for sampling, as done for the MUS case (cf. sections 4.5 and 4.6 for the reference to monetary unit sampling and stratified monetary unit sampling).

4.7.4 Projection

Please note that the use of non-statistical sampling does not avoid the need to project the errors observed in the sample to the population. The projection has to take into account the sampling design, i.e. the existence of stratification or not, the type of selection (equal probability or probability proportional to size), and any other relevant characteristics of the design. Therefore, the only significant difference between statistical and non-statistical sampling is that for the last the level of accuracy and consequently the upper error limit are not calculated.

The projection formulas to apply are the same as those used in the corresponding statistical methods (simple random sampling and monetary unit sampling) as described in Table 5.

Non-statistical method	Corresponding statistical method
EP non-stratified	SRS
EP stratified	Stratified SRS
PPS non-stratified	MUS
PPS stratified	Stratified MUS

Table 5. Correspondence between sampling and non-sampling methods

4.7.5 Evaluation

In any of the previously mentioned strategies the projected error is finally compared to the maximum tolerable error (materiality times the population expenditure):

- If below the tolerable error, then we conclude that the population does not contain material error;
- If above the tolerable error, then we conclude that the population contains material error.

Even if it is not possible to calculate the upper limit of error and consequently there is no control of the audit risk, the projected error rate is the best estimation of the error in the population and can thus be compared with the materiality threshold in order to conclude that the population is (or not) materially misstated.

4.8 Introduction to two-stage sampling

4.8.1 Introduction

Whenever the selected items (usually projects) include a large number of expenditure items[16], the AA can apply two-stage sampling, selecting the expenditure items by using the same sampling principles used to select the projects. This offers the possibility to significantly reduce the audit work, allowing to still control the reliability of the conclusions.

In this case, appropriate sample sizes have to be calculated within each item (project). A very simple approach to the determination of sub-sample sizes is to use the same sample size determination formulas that are proposed to the main sample under the several sampling designs and based on parameters compatible with expected subpopulation characteristics. For ex. when selecting projects in the first stage, one should acknowledge that the reference population for the second sampling stage is now the project, inside which the subsample is selected and that the population parameters used for the determination the sub-sample size should, whenever possible, reflect the characteristics of the corresponding project.

The AA may choose to use any sampling method for selecting the expenditure items within the projects. In fact, the sampling method used at the sub-sample level does not need to be equal to the one used for the main sample. For example, it is possible to have

[16] In theory, the project can be subject to subsampling regardless the number of items. Whenever the subsample size calculation produces a value close to the population size, the subsampling strategy won't produce any significant reduction in the audit effort. Therefore, the threshold that suggests the use of subsampling is just the result of the AA evaluation (based on subsample size calculation) of the gain (reduction of audit effort) that can be brought by this strategy.

a sample selection of projects based on MUS and a subsample of expenditure items within one project based on simple random sampling. Nevertheless, the subsampling strategy (sampling within the primary unit) should always be statistical (unless the sampling of primary units is not itself statistical). The choice between the possible methods is made under the same conditions of applicability that have been proposed in Section 4.1. For example, if within a project it is expected to have a large variability of the expenditure of the sub-sampled expenditure items and it is expected to have positive correlation between errors and expenditure, then a selection of expenditure items based on MUS may be advisable. Also, when using simple random sampling (SRS) it may happen there are a few units within the project that stand out due to high level of expenditure. In this case, it is highly advisable to use stratified SRS creating a stratum for the high value items (typically exhaustively observed).

Once the sub-sample is selected and audited, the observed errors have to be projected to the respective project (or other main sampling unit) using a projection method compatible with the selected sampling design. For example, if the expenditure items have been chosen with equal probabilities than the error may be projected to the project using the usual mean-per-unit estimation or ratio estimation.

Finally, once the errors have been projected for every project (or other main sampling unit) in the sample that has been sub-sampled, the projection to the population follows the usual procedure (as if one had observed the whole expenditure of the project).

4.8.2 Sample size

In case of statistical sampling, the easiest way to calculate sample sizes is the following:

- First stage: calculate sample size using the usual appropriate formulas and parameters (should always be larger or equal 30).
- Second stage: for each sampling unit subject to subsampling calculate sample size using again the usual formulas (appropriate to the type of selection used at the second stage). Parameters should be compatible with the ones used at the first stage, although some may be adapted to reflect the reality of the reference main sampling unit (often the project). For example if there is historical data about the level of variance of the errors within the project, one should use this variance instead of the variance of the errors used for the sample size calculation at the first stage. At this stage sample size should also be larger or equal to 30.

In case of non-statistical sampling the planned coverage in number of units should be applied to both stages. For ex. when working with a population with minimum risk in the main stage of sampling at least 10% of the main sampling units (usually the projects) are to be selected. In the second stage of sampling, also the minimum of 10% of the items (e.g. expenditure documents, invoices) within each project have to be selected to the subsample. Also, one should notice that the coverage in terms of expenditure is to be calculated based on the effectively audited expenditure. For ex. when targeting for a coverage of 15% of the expenditure the AA has to ensure that the ratio between the expenditure effectively audited and the declared expenditure reaches this threshold. In practice it may be an efficient strategy to increase the sample size of the first stage of sampling in order to avoid the need to audit the full expenditure of the selected projects.

4.8.3 Projection

As for the sample size calculation, also the projection is made under two-stages. Firstly, the subsamples within the main sampling units (projects) are used to project the error for those projects. Once the error of the main sampling units are projected (estimated) they are treated as if they were the "true" errors and will became part of the usual extrapolation process based on the main sample.

In summary:

- For each project main sampling unit (project) subject to subsampling, estimate its error (or error rate) using the sample of expenditure items;
- Once the errors for all main sampling unit (project) have been estimated, use the main sample to project the total error of the population;
- In both cases the projection should be based on the formulas that correspond to the sample designs that have been used to select the units.

4.8.4 Accuracy

For the cases where statistical sampling is applied, the accuracy is calculated as usual, i.e. using the formulas in accordance with the sampling design used for the first stage of sampling and ignoring the existence of subsampling. Errors of projects are filled in accuracy formulas despite their nature (either the true ones when subject to full audit or estimated ones when subject to subsampling).

4.9 Pilot samples and additional samples

Whenever the AA uses a pilot sample, and in case an additional sample is selected, subsequently it is possible to combine the two samples in order to form the global sample.

Typically, the projected error produced from the original pilot sample is substituted in formulas for sample size determination in the place of the anticipated error (in fact the projected error is at that moment the best estimate of the error in the population).

Doing this, a new sample size can be calculated based on the new information arising from the original sample. The size of the additional sample needed can be obtained by subtracting the original sample size from the new sample size.

Finally, a new sample can be selected (using the same method as for the original sample), the two samples are grouped together and results (projected error and precision) should be calculated using data from the final grouped sample.

The projection of the errors from the grouped sample is performed with the standard formulas. For example, under SRS the average error (mean-per-unit estimation) or error rate (ratio estimation) is calculated with the full grouped sample and its global sample size. In the case of MUS, the projection is performed, as usual, summing the error rates of all units in the sample and multiplying by the ratio between the total expenditure of the population and the size of the grouped sample.

Imagine that the pilot sample with sample size equal to 30 projects produced a projected error rate of 1.3%. Subsequently, he projected error rate of 1.3% should be imputed in the formula for sample size determination in the place of the anticipated error, leading to a recalculation of the sample size, which would produce in our example a new sample size of n=48. As the original sample had a size of 30 units, this value should be subtracted from the new sample size resulting in 48-30=18 new observations. Therefore, an additional sample of 18 projects should be now selected from the population using the same method as for the original sample. After this selection, the two samples are grouped together forming a new whole sample of 30+18=48 units. This global sample will finally be used to calculate the projected error and the precision of the projection using the usual formulas. In case of MUS selection, the projected error is obtained summing the 48 error rates in the sample and multiplying by $BV/48$.

4.10 Calculation of global error rate and precision

The AA may be required to produce an aggregation of error rates, precisions and upper error limits calculated at programme or group of programmes level. This may happen, among other situations, when aggregating the projected errors of different programmes in one single projected error to support a global audit opinion.

Consider that the AA has projected errors for K different programmes or groups of programmmes.

Error rate

The global error rate is just a weighted average of the error rates projected at programme (or group of programmes) level, using as weights the expenditure of the programmes/groups of programmes.

$$ER = \frac{1}{BV}\sum_{k=1}^{K} BV_k . ER_k$$

where BV is the global population expenditure; BV_k is the expenditure of programme/group of programmes *k*; ER_k is the projected error rate of programme/group of programmes *k*; and *K* is the number of different programmes/groups of programmes.

Precision

The global precision may be calculated from the precisions previously calculated at programme/group of programmes level, using

$$SE = z_{max} . \sqrt{\sum_{k=1}^{K} \left(\frac{SE_k^2}{z_k^2}\right)}$$

where SE_k^2 is the square of the projected error rate precision for programme/group of programmes *k*; z_k^2 is the square of the *z*-value associated with the confidence level used for projecting the errors of programme/group of programmes *k*; $z_{max} = \max(z_1, z_2, \dots, z_K)$ is the maximum of the *z*-values applied at programme/group of progranmmes level.

Upper error limit

The global upper error limit is calculated, as usual, by adding the global projected error and the global precision

$$UEL = ER + SE$$

5 Conclusion

The guidance provides the different sampling techniques that can be used for substantive testing:

- Simple random sampling (Section 4.3);
- Stratified simple random sampling (Section 4.4);
- Monetary unit sampling (Section 4.5);
- Stratified monetary unit sampling (Section 4.6);
- Non statistical sampling (Section 4.7).

A brief description supporting the implementation of two-stages sampling strategies is presented in Section 4.8. This technique can be used in combination with any of the above-mentioned sampling methods to reduce the audit effort with minimum impact on the reliability of the audit results.

The Audit Authorities should select the sampling methods based on the nature and characteristics of the population and amount of auxiliary information available. Section 4.1 provides information of the favorable conditions for the choice and application of each method. Audit Authorities should carefully consider these conditions along with the population characteristics to choose the most appropriate sampling strategies. The price to pay, for not choosing the most adequate methods, will often be an increase in sample size and therefore of the audit effort or even the generation of non-conclusive audit results.

The sampling method(s) selected by the AA should be described in the Audit Strategy (article 5.5.d of the Regulations and further details provided in in the Annual audit report.

The design of the sampling strategy must also take into consideration the 4 types of populations presented in the Section 2.5 of this document, namely:

- projects expenditure;
- management costs;
- bilateral funds;
- technical assistance.

The nature of the population frames the choice of the sampling units (cf. Section 2.6) and the specific strategies that can be used (cf. sections 3.2 to 3.6).

In any case the description of the sampling approach, including the options taken, should be explaining in the audit strategy.