PRACTICE GUIDE

Sampling Methodology for Performance Audits

CANADIAN AUDIT & ACCOUNTABILITY FOUNDATION

About the Canadian Audit and Accountability Foundation

The Canadian Audit and Accountability Foundation is a premier Canadian research and education foundation. Our mission is to strengthen public sector performance audit, oversight, and accountability in Canada and abroad. We build capacity in legislative audit offices, oversight bodies, and departments and crown corporations by developing and delivering:

- Training workshops and learning opportunities;
- Methodology, guidance, and toolkits;
- Applied and advanced research;
- Information sharing events and community building initiatives.

Visit us at www.caaf-fcar.ca for more information about our products and services.

Sampling Methodology for Performance Audits – A Practice Guide

Published by:

Canadian Audit and Accountability Foundation
1505 Laperriere Avenue
Ottawa, Ontario CANADA
K1Z 7T1

Tel: 613-241-6713

www.caaf-fcar.ca/

ISBN: 978-1-990150-02-9

This publication is available in French under the title:
Méthodes d’échantillonnage pour les audits de performance – Un guide pratique

Acknowledgements

The Canadian Audit and Accountability Foundation's mission is to promote and strengthen public sector performance audit, oversight, and accountability in Canada and abroad, through research, education, and knowledge sharing.

This Practice Guide is part of our performance audit capacity-building program and is the seventh in a series of such guides. It has been made possible by funding provided by the foundation's members and partners.

The consultative process undertaken for the Practice Guide project was extensive and included consultations with many leaders and professionals, both on an individual and collective basis, through networks such as the Canadian Council of Legislative Auditors, the Institute of Internal Auditors Canada, the Government Internal Auditors Council of Canada, and the Australasian Council of Auditors-General.

On behalf of our Board of Directors, we would like to acknowledge the support, thought leadership, and active contributions of the members of the core project team who guided the design and development of this Practice Guide[1]:

- **Jane Bryant**, Director, Performance Audit, Office of the Auditor General of British Columbia
- **Justin Fisher**, Assistant Director, Statistics, U.S. Government Accountability Office
- **Harold John Hilton**, Internal Specialist, Research Methods and Quantitative Analysis, Office of the Auditor General of Canada
- **Rachel Mathieu**, Director, Advisory Services Division, Auditor General of Québec
- **Ronald-Frans Melchers**, Associate Professor (retired), Criminology, Faculty of Social Sciences, University of Ottawa
- **Erika Thomas**, Audit Principal, Office of the Auditor General of Manitoba

We also wish to thank the many Canadian and Australian audit professionals who supported this project and provided comments on draft versions of the guide, including:

- **David Adams**, Director, Audit Practice, Queensland Audit Office
- **Andrew Atherton**, Assistant Auditor General, Office of the Auditor General of Nova Scotia
- **Tracy Cohen**, Manager, Audit and Evaluation Branch, Statistics Canada
- **Evangeline Colman-Sadd**, Auditor General, Halifax Regional Municipality
- **Ted Doane**, Executive Director, Province of Nova Scotia, Internal Audit Centre
- **Michèle Galipeau**, Auditor General, City of Montréal
- **Sheri Griffin**, Audit Director, Office of the Auditor General, Prince Edward Island
- **Roberto Grondin**, Project Leader, Advisory Services Division, Auditor General of Québec
- **Leslie Levita**, Chief Audit Executive, Internal Audit Directorate, Fisheries and Oceans Canada

[1] Titles and organizations of individuals named in this publication are those that were in effect during the project's development.

- **Yen Linh Ly**, Acting /Director, Policy and Professional Practices, Office of the Comptroller General, Treasury Board Secretariat of Canada
- **Tyson Shtykalo**, Auditor General of Manitoba
- **Dale Stoddart**, Senior Director, Audit Technical, Professional Services and Relationships Group, Australian National Audit Office
- **Martine Vadnais**, Chief Audit Executive, Yukon Territory Executive Council Office, Government Internal Audit Services
- And anonymous reviewers from the Institute of Internal Auditors (Canada) and the Provincial Auditor of Saskatchewan.

This Practice Guide project would not have been possible without the important contributions of its three co-authors:

- **Pierre Fréchette**, Director of Research, Product and Services, CAAF
- **Yves Genest**,[2] Vice-president, Products and Services, CAAF (Project Lead)
- **Paul Pilon**, statistical expert and consultant

Finally, we would like to extend our thanks to the teams and individuals who contributed their skills to producing and publishing this Practice Guide:

- **Lynne Casiple**, IT and Web Officer, CAAF
- **Véronique Coch**, Director of Communications and Stakeholder Relations, CAAF
- **Cyan Solutions**, Graphic design
- **Laurel Hyatt**, Editor
- **Francine Watkins Translation Services Inc.**

We hope this Practice Guide will be a useful and practical reference tool for audit professionals in Canada and abroad.

Carol Bellringer, FCPA, FCA, President and CEO, Canadian Audit and Accountability Foundation

[2] Comments, suggestions, and ideas can be provided to Yves Genest at the Canadian Audit and Accountability Foundation (ygenest@caaf-fcar.ca).

Table of Contents

Purpose of the Practice Guide

The purpose of this Practice Guide is to provide guidance for public sector auditors, both internal and external, on how to use sampling techniques in **performance audits**.[3] Specifically, this Practice Guide aims to provide guidance on how to:

- select appropriate sampling approaches and methods
- implement sampling plans
- report results obtained through sampling

While extensive material already exists on the appropriate use of **generalizable sampling** (also known as probabilistic, representative, or statistical sampling), guidance on **purposeful sampling** is more limited. This Practice Guide reviews basic concepts and techniques of generalizable sampling useful to performance auditors while exploring various theoretical and practical aspects of purposeful sampling. It takes recent developments in the standardization of purposeful sampling methods from the discipline of program evaluation and applies this knowledge to performance auditing.

Scope of the Practice Guide

This Practice Guide covers the basic concepts, approaches, and methods that auditors should be familiar with to select and apply appropriate sampling methodology in performance audits. For example, it includes discussions of:

- the main types of sampling approaches and methods that can be used in performance audits
- the key factors that affect sampling
- the steps that need to be taken in the planning and examination phases of performance audits to select and apply appropriate sampling approaches and methods
- the limitations and risks of using sampling as an audit procedure
- what to consider when reporting the results obtained from a **sample**

This Practice Guide is not meant to be a comprehensive textbook on sampling in the context of audits. It does not cover all the intricacies of statistical theory or fully describe all the sampling techniques available, although it has tried to cover the ones useful for performance auditors. Furthermore, it does not cover sampling methods for financial audits in any detail.

Using the Practice Guide

The Practice Guide is a flexible tool to be used within each audit office's existing processes and procedures, in accordance with auditing and assurance standards. It is therefore a complement to current audit methodology.

Back to Table of Contents

[3] Terms that are defined in the Glossary at the end of this document appear in bold the first time that they are used in the text.

Practice Guide to Sampling Methodology for Performance Audits

Part 1 – Concepts and Context

Sampling in a Performance Audit Context

Sampling is a methodology used for a variety of purposes in different professions, from ecological studies to surveys of political opinions. Some general concepts are relevant for all users of sampling methodology, but each discipline also uses sampling in its own specific way, and performance auditing is no exception. This section presents a short introduction to sampling in a performance audit context.

What Is Sampling and Why Is It Important When Conducting an Audit?

Essentially, sampling is the selection of a subset of the totality of elements that could be analyzed, called the **population**. It is a useful tool whenever assessing the whole population (which is called performing a **census**) is not efficient or insights regarding the nature or cause of a problem are required.

Although performing a census is more and more facilitated by computer-assisted audit techniques, there are still plenty of situations in which auditors must use sampling to obtain information that will allow them to conclude on their audit objective. The sample is therefore of critical importance to auditors because it provides the basis for their analysis and findings. For example, in the Office of the Auditor General of Canada's 2018 reports, 7 of the 14 performance audits conducted that year used sampling to draw **audit observations**.

When files are mostly paper-based or when electronic files and data are not amenable to data analytics, sampling uses resources more efficiently and takes less time than performing a census (which would be unrealistic in most circumstances). Depending on the type of sampling approach used, it may also be possible to extrapolate the findings of the analysis to the whole population.

Knowing how to design an appropriate sampling methodology is a useful skill for performance auditors. It is not uncommon for larger audit institutions to have one or more sampling experts on staff. Regardless, it is advantageous for all performance auditors to have at least a basic understanding of sampling approaches and methods and how to apply them. At a minimum, having this knowledge can enable auditors to implement less complex sampling methods and collaborate with external sampling experts who may be consulted on occasional projects.

What Can Be Sampled?

Numerous aspects of government operations can be the subject of performance audits. These audits often focus on the management of specific programs and services or the implementation of policies and regulations. Because of the varied topics covered by performance audits, there is a wide range of opportunities for sampling. Commonly sampled items (called **sampling units**) in performance audits include, for example:

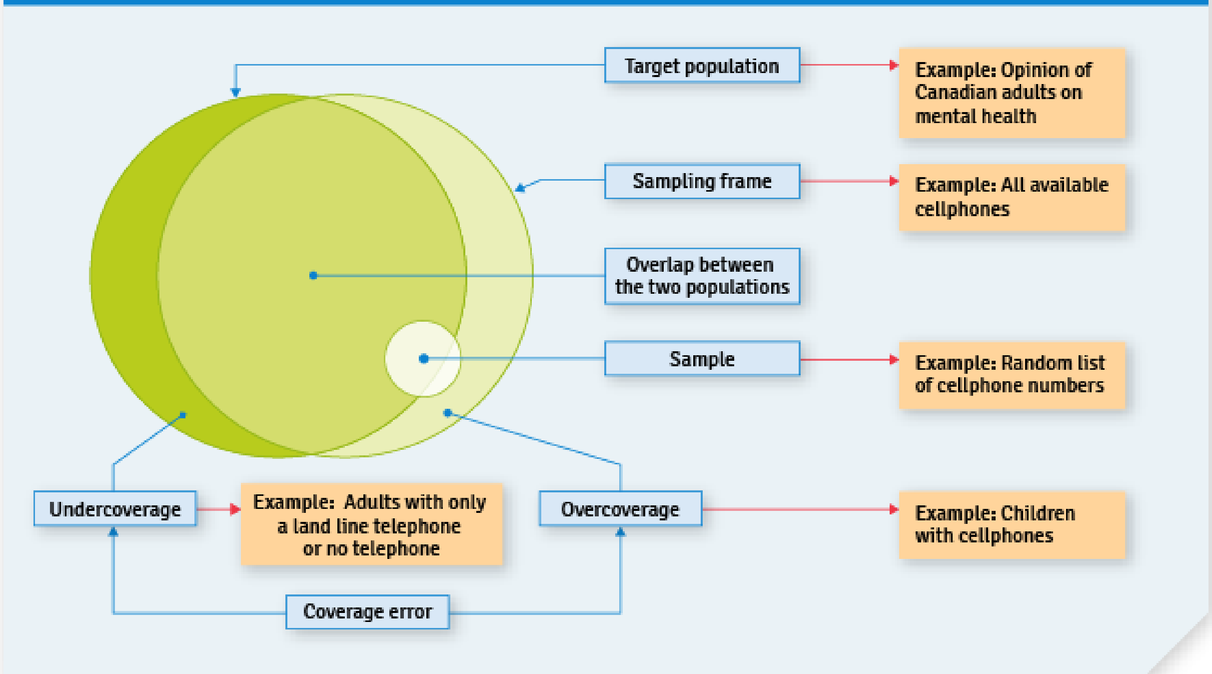
- program files, project files, personal files (personal files can include health care files, education files, immigration files, taxpayer files at the revenue agency, etc.)
- people (employees, managers, clients, contractors, stakeholders, etc.)
- transactions (financial transactions, contracts, sales, purchases, etc.)
- locations (service-delivery locations, schools, health care institutions, embassies, etc.)
- equipment (airplanes, trucks, ships, computers, medical devices, etc.)

Over the course of a year, audit institutions will often need to assess samples in many different categories. For example, in the 2018 reports of the Office of the Auditor General of Canada, a variety of samples were used to draw conclusions. The sampling units in these samples included many personal files (detainee cases, offender parole cases, and tax files), locations (foreign affairs missions), program files (military justice cases), and project files (asset disposal decisions).

A sample is always taken from a target population. This target population can be extracted from a project database, a human resource database, a client list, or any similarly organized record. In some instances, a sample can also be extracted while conducting an inspection or a site visit (e.g., in an audit of public transit, stopping every fifth passenger to ask them questions about their user experience). A full extract can be performed or, alternatively, a selective extract can be obtained based on clearly stated parameters depending on the audit scope or audit objective. For example, parameters can be defined to select only projects with values between $1 million and $10 million, or service beneficiaries between 18 and 30 years old, or transactions conducted in the first quarter of the 2019–20 fiscal year. Once a target population is defined, a **sampling frame** can be developed.

The sampling frame is a list of the items in the population. Although ideally a sampling frame is the same as the target population, there may be a difference between them due to inaccuracies in the database, the absence of unique identifiers, out-of-date information, or other factors. This difference is called a coverage error. **Figure 1** illustrates these concepts for a study on the opinion of Canadian adults about mental health issues. A coverage error arises because a part of the population does not own a cellphone and is not included in a sampling frame based on a random list of cellphone numbers.

Figure 1 – Population, Sampling Frame, and Coverage Error



Source: Based on Buskirk (2016), Target Population and Sampling Frame in Survey Sampling

Back to Table of Contents

Population Characteristics

Before selecting any sample, it is necessary to have a good understanding of the population as a whole. This includes knowing what types of variables can be studied in the population, as well as determining the population's level of **homogeneity**. Other issues, such as data availability and quality, are also important and are addressed later in this guide. (See the section on Analyzing the Population.)

Continuous and Categorical Variables

The variables of interest in a population can be of two general types: continuous and categorical. Continuous variables are variables that are expressed by a number along a continuum. Examples of continuous variables include:

- dollar value (e.g., the value of a transaction, purchase, or contract)
- distance (e.g., the distance from the nearest fire department)
- time (e.g., waiting time in an emergency room or the number of days taken to process a transaction)
- pollution levels (e.g., milligrams of mercury in fish)

Occasionally, it is necessary to categorize units of the population based on a continuous variable, such as creating categories for low, medium, and high dollar-value contracts. These categories can then be used to tailor a sampling approach.[1]

Categorical variables are non-continuous. Some examples include:

- marital status (single, married, separated, divorced, widowed)
- citizenship and immigration status (Canadian citizen, permanent resident, new immigrant, refugee)
- employment status (full-time, part-time, unemployed)
- educational level attained (primary school, high school, college, master's degree, Ph.D.)
- geographic location (continent, country, province, region, city, neighbourhood)
- public administration (entity type, sector of activity, governance model)

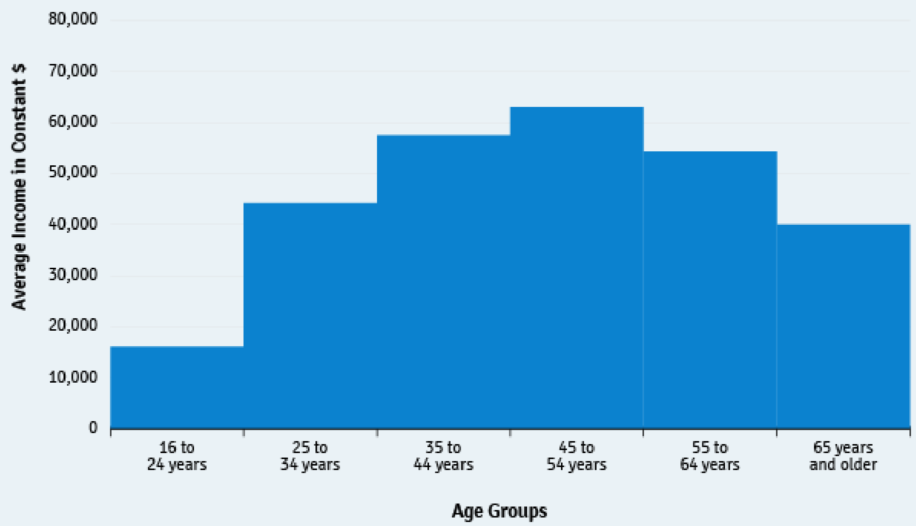
While some categorical variables have units that can be rank ordered, such as educational levels, most represent distinct, non-hierarchical categories.

Regardless of the type of variable, categorical or continuous, the distribution of the variable can be assessed using diagrams (histograms for continuous variables and bar charts for categorical variables). As shown in **Figure 2** and **Figure 3**, charts are an effective tool to assess how a population is distributed. (More information on how to analyze populations is in **Part 2** of this guide.)

[1] It is also possible to use a probability proportional to size sampling, a method in which a size measure is available for each population unit before sampling and where the probability of selecting a unit is proportional to its size. This is similar to a dollar unit sampling in a financial audit.

Figure 2 – Example of a Histogram

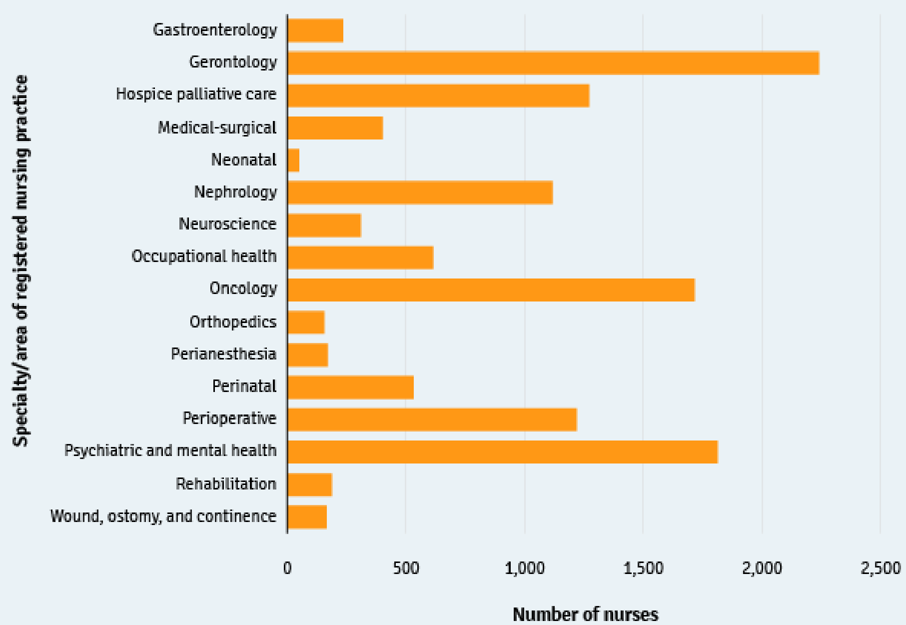
Average Income of Individuals, by Age Group, Canada, 2017



Source: Data from Statistics Canada, Income of individuals by age group, sex and income source, Canada, provinces and selected census metropolitan areas (accessed on 10 December 2020).

Figure 3 – Example of a Bar Chart

Number of valid Canadian Nurses Association certifications by specialty/area of registered nursing practice, Canada, 2018



Source: Canadian Institute for Health Information (2019), Health Workforce, 2018: Indicators (based on information provided by the Canadian Nurses Association, Certification and Professional Development).

Homogeneity and Heterogeneity

Homogeneity is the level of uniformity among sampling units within a population. Homogeneity is commonly interpreted as meaning that all the items in the sample are chosen because they have similar or identical traits (for example, people in a homogeneous sample might share the same age, location, or employment). However, the mathematical meaning of homogeneous is that a data set can be analyzed mathematically and is operating under the same rules and constraints.

The more homogenous a population, the more valid the conclusions drawn from a small sample. Lack of homogeneity, known as **heterogeneity**, within a population can have a major negative impact on the interpretability and validity of results obtained from a sample. When a population is heterogeneous, there is a higher likelihood that a single sample will not reflect the complexity of the population—that is, important characteristics may be misrepresented or ignored.

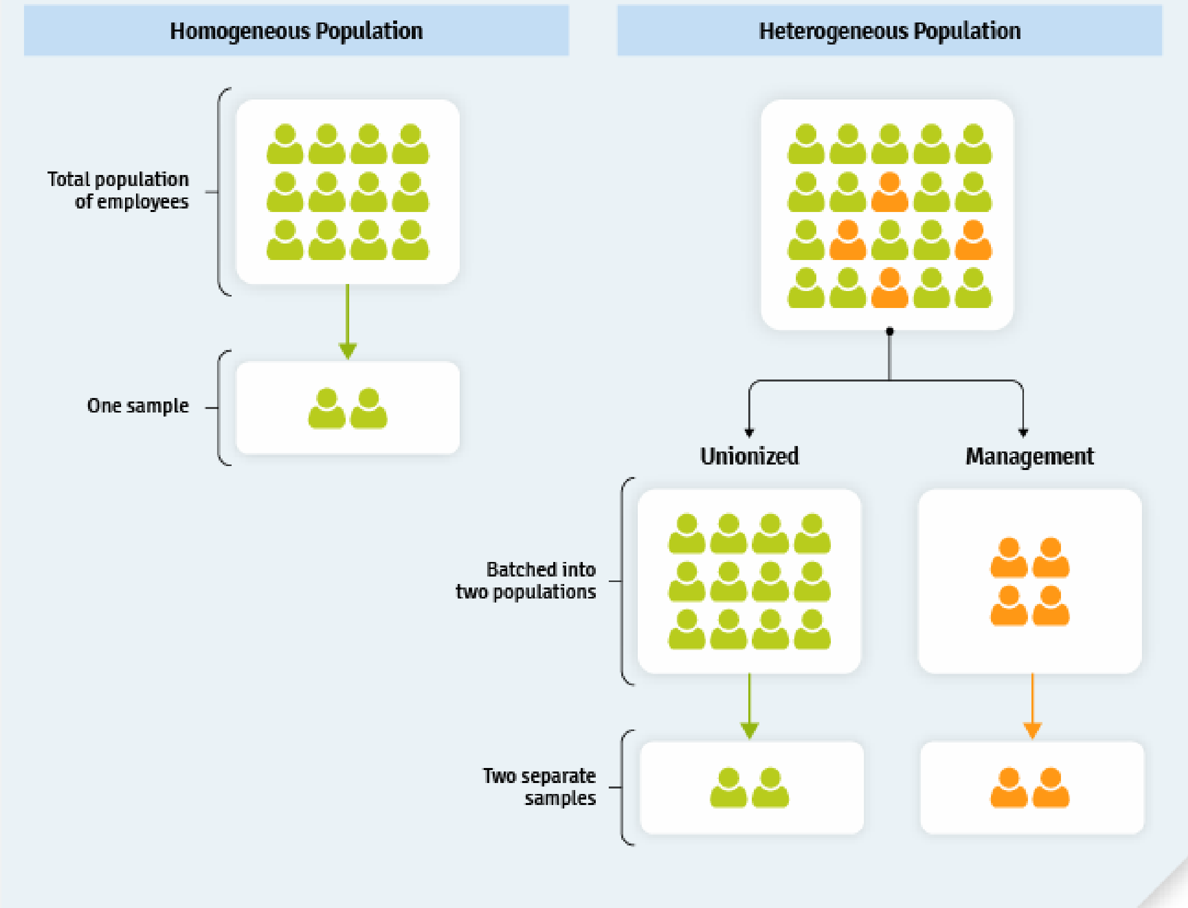
For this reason, assessing the heterogeneity level in a population is a key step in the sampling process, for both generalizable sampling and purposeful sampling. As a general rule, when dealing with a heterogeneous population, the population should be divided into as many groups as necessary to ensure that each subgroup is sufficiently homogeneous for the sampling purpose as defined by the audit objective and scope.

Two examples in **Figure 4** illustrate the importance of understanding whether a population is homogenous or not. In the first example, the context is a hypothetical survey of employees of a government department about the adequacy of water and air quality in their work environment. Because the employees all breathe the same air and drink the same water, they form a homogenous population and, in that case, a single sample would therefore be sufficient. In the second example, the survey is about management style in the department. Because some employees are part of management and others are unionized, there are two distinct groups of people (two populations) that may have very different opinions about management style in the department. In this instance, it would therefore make more sense to take two separate samples, one from each population.

Assessing a population's level of heterogeneity is a difficult initial step to take and is often conducted with little firm data. In some instances, auditors may not even have the information to assess the degree of heterogeneity in a population. For example, the analysis in **Figure 4** would not be possible if personnel data broken down between unionized staff and management was not available.

Heterogeneity always increases both the cost and complexity of any audit, because more samples and sampling approaches could then be required to complete the audit work. Also, because it is unlikely that any population or subpopulation will be perfectly homogenous, audit teams have to judge the amount of acceptable heterogeneity based on their audit objective and scope.

Figure 4 – Graphical Representation of Different Sampling Approaches for Homogenous and Heterogeneous Populations



Ultimately, auditors have to be comfortable that each subgroup they create is made up of reasonably similar units (in terms of **materiality**, **risk**, population characteristics, or other parameters relevant to the audit objective) that can be analyzed in the same manner.

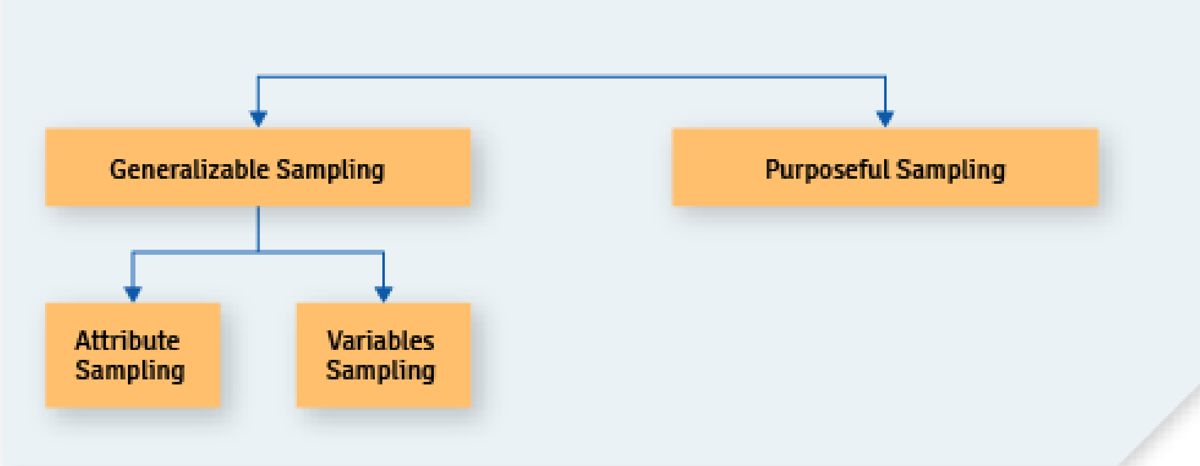
More information on how to assess and optimize the homogeneity of populations is in **Part 2** of this guide.

Back to Table of Contents

Sampling Approaches

There are two main types of sampling approaches for performance auditors who need to select a sample in order to draw audit observations: generalizable sampling and purposeful sampling. **Figure 5** provides a schematic overview of the options outlined in this section.

Figure 5 – Sampling Approach Options for Performance Audits



Generalizable Sampling

Generalizable sampling is often called probabilistic sampling, representative sampling, or statistical sampling. It is defined by the Office of the Auditor General of Canada (OAGC, 2019) as "the application of auditing procedures to a representative group of less than 100% of the items within a population of audit relevance such that all sampling units have a chance of selection in order to provide the auditor with a reasonable basis on which to draw conclusions about the entire population." In statistical guides, this type of sampling is called "probabilistic sampling" due to the random nature of the method of sample selection.

The term "generalizable" is apt given that it reflects the intent of this type of sampling. The estimates derived from a generalizable sample (amounts, rates, **means**, **variances**, etc.) are meant to reflect or represent the parameters of the population from which the sample was drawn. In other words, it is possible to extrapolate the results of the analysis conducted on the sample to the whole population.

Generalizable sampling requires a strict adherence to the principle of random selection: samples must be selected in an unbiased manner. Equally important, the population from which the sample is drawn must be relatively homogeneous and the **sample size** must be sufficiently large—otherwise, it will not be appropriate to extrapolate the results to the whole population.

Generalizable sampling also requires calculations of sample size, which is often done using specialized software applications. More information on how to calculate sample size for a generalizable sample is in Appendix 1 of this guide.

Generalizable sampling is ideal for **substantive tests** (such as assessing compliance with service standards for social or health programs) or tests of non-automated **controls** (such as verifying that proper approvals are obtained when required).

An example of a generalizable sampling approach in an audit is in **Text Box 1**.

Text Box 1 – Example of a Generalizable Sampling Approach

Audit: Office of the Auditor General of Canada – Status Report on Evaluating the Effectiveness of Programs, published Spring 2013

Sampling approach:

"Use of random samples. This audit used two random samples across the three audited departments:

- A sample of 54 programs from the three departments designed to represent all 120 departmental programs subject to the evaluation requirements.
- A sample of 32 Treasury Board submissions from the three departments designed to represent all 71 departmental submissions during the audit period. The sample was limited to submissions for program funding where an evaluation would be expected. We excluded submissions for other purposes, such as to seek funding for a capital acquisition or to settle a lawsuit, from the sample.

Both samples were sufficient in size to conclude on the sampled populations with a confidence level of 90 percent and a margin of error of +10 percent."

In an audit context, two types of generalizable sampling are prevalent: attribute sampling and a type of **variable sampling** called monetary-unit sampling.

Attribute sampling

An "attribute" is a characteristic of a sampling unit that can be determined by a binary choice, such as yes/no, error/no error, or on time/late. For example, an attribute could be whether a transaction (the sampling unit) was signed off by an official with the proper authority (yes/no) or whether a project has been completed on time and on budget (yes/no).

Attribute sampling is used to assess the proportion of a specified attribute in a sample and to extrapolate this proportion to the entire population being sampled. An example of attribute sampling could be determining, through a sample, the proportion of benefits applications accepted by a social program that are provided to ineligible recipients (a yes/no situation).

Attribute sampling is often used in performance audits to determine whether controls are functioning reliably. Calculating a sample size requires determining a few parameters: the population size, the **confidence interval**, the **confidence level**, and the expected proportions for the attributes being tested (the **expected error rate**).

Monetary-unit sampling (MUS): A type of variables sampling

In monetary-unit sampling (also called dollar-unit sampling or DUS), the total number of dollars is the population. For example, in an audit of contracts where the value of a contract or project is measured in dollars and can vary from $1 to $1 billion or more, the population would consist of the total dollar values of all the contracts subject to an audit. Monetary-unit sampling is used to estimate and extrapolate amounts or quantities, such as the total cost overrun in a project portfolio over a given time period, or the total error in salary payments to an organization's employees for a given year.

In contrast to attribute sampling, monetary-unit sampling is rarely used in performance audits. (It is much more common in financial audits.) Calculating a sample size in a MUS context requires determining a few parameters: maximum tolerable error, the confidence level, and the expected error.

Purposeful Sampling

While generalizable sampling involves a specific approach, non-generalizable sampling (also called non-probability sampling) does not. Non-generalizable sampling uses a number of approaches with a host of characteristics that make them appropriate or inappropriate for the purpose of auditing. Traditionally, auditors have relied on one form of non-generalizable sampling: judgmental sampling, which is the reliance on professional judgment to select units of the population based on their auditing experience. This ambiguous and non-directive approach can often lead to findings that can be easily challenged and is not recommended in a performance audit. Another non-generalizable sampling approach, known as purposeful sampling (Patton, 2015), tries to interject both rigor and objectivity into the sampling approach that auditors use. For the purpose of this Guide, we will use the term "purposeful sampling" to describe this approach. Note that it does not include the exclusive reliance on professional judgment.

The rules regarding sample selection and size for generalizable samples are based on statistical rules regarding probability and error estimation. However, the rules regarding sample selection and sample size in purposeful sampling are based on developing a rational argument that links the method of selection to the purpose of the investigation. Conclusions from purposeful sampling will often be of two general types. The first type is that the sample findings provide sufficient evidence of material error. In the second type, the sample provides significant insight as to the nature or cause of material error. Consider the three following hypothetical examples of purposeful samples and conclusions:

"Having selected a small sample of the most well-funded senior residences, we observed significant cases of elderly abuse. As a result, we can reasonably assume the level of care at other less well-funded institutions is likely poorer."

"Having selected a small sample of contracts with high public visibility and high value, we have observed ineffective and insufficient oversight practices resulting in either project failure or delays. Therefore, we can reasonably assume that other less risky or costly contracts are also experiencing ineffective and insufficient oversight."

"Having selected a small sample of IT modernization projects, some that have failed and some that have succeeded, we have observed that one of the key elements of success has been a centralization of decision making for day-to-day production decisions. As a result, we can reasonably conclude that decentralization of day-to-day production decisions is one major cause of IT modernization failure."

In each case, the sampling approach is specifically tailored to the purpose of the investigation. The purposeful bias of the sample is used to strengthen the meaningfulness of the results, and while the results are never mathematically extrapolated to the population, the results logically speak to the likelihood of systemic problems. While there are several explicit purposeful sampling strategies, it is the rational link between the method of selection and conclusion that contributes to the validity and meaningfulness of the findings.

Therefore, purposeful sampling involves the introduction of an explicit bias in the sample's selection, with the specific intent of isolating and selecting information-rich cases that will be particularly useful to gain insight and understanding. For instance, auditors can focus on items of high value or deliberately select specific segments of a population. The major strength of purposeful sampling is its potential to communicate a powerful narrative, providing a perspective few might be aware of by selecting critical cases illustrating a program's operations under a variety of conditions. **Text Box 2** provides an example of an audit that used a purposeful sampling approach.

Text Box 2 – Example of a Purposeful Sampling Approach

Audit: Office of the Auditor General of Canada—Capital Projects—Yukon Hospital Corporation; published February 2013

Sampling approach:

"We used a targeted selection of 10 contracts from a total of 26, awarded between April 2009 and May 2012. The selected items were chosen to achieve equivalent coverage over the three capital projects (Crocus Ridge Residence, Watson Lake Hospital, and Dawson City Hospital) and to provide enhanced coverage of high-value contracts (7 of 8)."

Sample size for purposeful sampling also differs from the mathematically driven process of generalizable sampling. Ultimately, sample size for purposeful sampling is driven by two factors. The first is the principle of redundancy; that is, when the examination of new records or examples begins to reflect the same issues and problems already uncovered and no new information is coming to light. The second factor is whether or not the auditee is willing to accept the observation as valid. In most cases, if the negative audit observations are accurate, the auditee is willing to accept them at face value if the alternative is that the auditor will return and review more cases and uncover additional negative findings.

In performance audits, auditors gather evidence from multiple sources to build an argument and conclusions that can withstand scrutiny. The results from purposeful sampling are only one part of that argument. The specific sampling methodology used (there are many options) will depend heavily on what specific information is needed to help support the argument being built. Therefore, the reliability of the results from purposeful sampling depends not only on the method of selection, but also on the robustness of the entire argument being built and the degree to which all the pieces of evidence support the overall audit conclusions. Appendix 3 provides a comprehensive overview of methods for selecting cases for a purposeful sample.

It is a misconception that generalizable sampling is a preferred or a superior option over purposeful sampling. Both approaches have specific strengths and weaknesses. It is true that, when conducted properly, the results from generalizable sampling are autonomous and difficult to refute. Generalizable sampling also has the advantage that its results can be extrapolated to the whole population. However, not all populations or circumstances are suitable for generalizable sampling, and performance auditors often have to rely instead on a purposeful sampling approach to obtain the data they need to reach audit conclusions.

Back to Table of Contents

Factors to Consider When Determining Precision of a Generalizable Sample

When a sample is taken and analyzed, a value of interest is obtained. However, because the value is derived from a sample of a larger population, it is not guaranteed that the value obtained exactly represents the population as a whole. Rather, the value obtained is a *best estimate* of the population value. The **precision** of this estimate varies according to the sample size.

Depending on the objective pursued by auditors, the required level of precision will also vary. Some objectives require only a minimally precise sample. For example, if the objective is to demonstrate whether a moderate to high level of error exists in a process or program, and the exact level of error is immaterial, then a small sample will be enough. In contrast, if the objective is to estimate the extent of **deviation** from an expected result, then the exact level of error will be relevant.

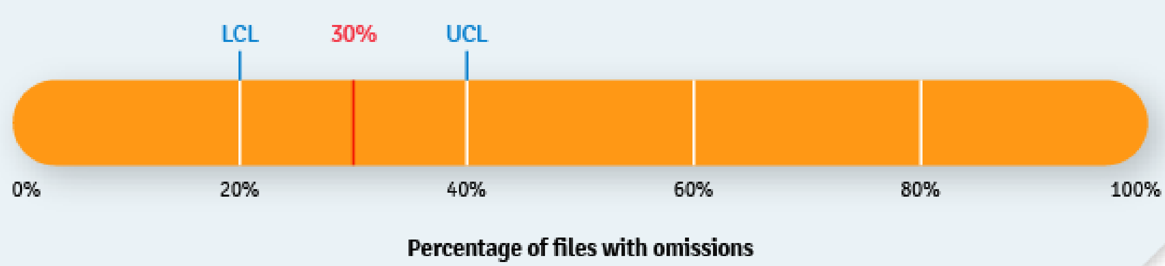
Two essential measures are used to provide information on how effective a sample is likely to be as an estimator of a population. These measures are the confidence interval (CI) and the confidence level (CL). Both are required to describe the precision of any generalizable sample estimate. (These notions do not apply to purposeful samples.)

Confidence Intervals

The confidence interval is the range of values over which we can be reasonably confident that the true population value resides, given the result from a sample. Smaller, or narrower, confidence intervals mean that there is greater precision around an estimate of a population. For example, the narrower confidence interval of ±5% is more precise than the wider confidence interval of ±10%. In the first case, we may be confident that the true population value lies from 5% below the sample value to 5% above. In the second case, we would only be confident that the population value rests within the interval from 10% below (lower confidence limit) to 10% above (upper confidence limit)—a much wider range.

For instance, let’s assume that after a hypothetical examination of a sample of 50 files, the auditors found that 15 (or 30%) have major omissions. If the confidence interval is ±10%, then the highest likely value is 40%. This is called the upper confidence limit (UCL). Likewise, the lowest likely value is 20%. This is called the lower confidence limit (LCL). These concepts are illustrated in **Figure 6**.

Figure 6 – Confidence Interval



While confidence intervals are often expressed with both a lower bound and an upper bound (e.g., ±10%), this is not the only option. Sampling may be done with either a one-sided or a two-sided confidence interval. Most statistical reporting in the media or by statistical agencies relies on two-sided confidence intervals because they show more complete information. Given the result from a sample, it shows both how low the true population value is likely to be, and how high. But auditing is different because the focus is often only on whether or not the population value exceeds a certain amount (for example, a maximum level of non-compliance in a process or program). In such cases, it is not necessary to take into account whether the extent of non-compliance in a population is lower than the estimate provided by the sample. The concern is rather with how much greater the non-compliance rate might be in the population than the rate estimated by the sample value.

One-sided sampling provides less information than two-sided sampling, but the benefit is that sample sizes are cut in half. This makes for more efficient audits by eliminating work that is not necessary to conclude on the audit objective. Most audit sampling software, such as IDEA and ACL, automatically computes sample sizes based on one-sided estimates. Whereas two-sided confidence intervals are normally expressed as ±10, for example, one-sided confidence intervals show only the "+" or the "–" sign, such as +10.

Confidence Levels

The confidence level is the extent of the certainty that the true population parameter of interest falls within the confidence interval. For example, if the level of precision of a sample estimate is a ±5% confidence interval with a 90% confidence level, then the true population parameter has a 90% chance of being within ±5% of the estimate. One of the most common misconceptions among auditors using audit sampling is that the confidence level relates to the sample value. It does not. Rather, it relates to the confidence interval surrounding the sample value. For instance, in **Figure 6**, with a confidence interval of ±10%, we would be 90% certain that the value is between 20% and 40%.

Expected and Observed Error Rates in Performance Audits

For most performance audits, the auditor has no real basis for estimating the error rates (expected error rate) before performing the examination. Unlike financial audits, performance audits are not routinely repeated and results from several past audits are not available to predict the outcome of the current audit. As a result, performance auditors must choose another strategy for selecting a value for expected error because it is a required input variable for calculating a sample size (as part of a generalizable sampling approach).

It can seem counterintuitive, but the result of the examination, the **observed error rate**, does have an impact on the precision of the findings. This is because this precision is influenced by the amount of variance within a population, and the variance within a population reflects the rate of error. Variability within a population is at its highest when the error rate is about 50%. For any given sample size, the accuracy of estimates will be lowest, less precise, with a finding of about 50% error rate. As the error rate decreases, variability is reduced, and the estimate becomes more precise. More information on expected and observed error rates is in Appendix 1.

Precision Standards for Performance Audits

Confidence level and confidence interval are closely interrelated notions, and both help to determine what the sample size should be. Audit institutions have rules about what confidence levels and confidence intervals are acceptable for audit purposes. For example, at the Office of the Auditor General of Canada, the performance audit practice requires a confidence interval of no more than 10% and a confidence level of no less than 90% for audit sampling. Sample sizes are typically computed so that particular targets (or minimum standards) for confidence interval and confidence level are achieved.

In situations of high materiality where there is a need for a very accurate measure of error rate, a confidence interval of 5% with a confidence level of 95% may be more appropriate. Examples of this situation are estimates of revenue collection for a large cost-recovery program or of amounts resulting from overpayments for a large social program. However, achieving this higher standard requires larger sample sizes and thus completing the audit work will require more time and resources.

In general, small samples (35–45 units) have a confidence interval of about 10% and a confidence level of 90% depending on population size and observed error. Generalizable samples for performance audits should not have levels of precision any lower than this. Smaller sample sizes lower the confidence level and widen the confidence interval, which means, respectively, that there is less assurance in the results, and less precision as well.

It is a good practice for audit institutions to be transparent about the standard used in each audit that relied on a generalizable sample, to prevent undue reliance on the sample by users of the report. At the Office of the Auditor General of Canada, for example, sampling parameters are typically reported in the "About the Audit" section at the end of performance audit reports by saying something like this: "Where generalizable sampling was used, sample sizes were sufficient to conclude on the sampled population with a confidence level of no less than 90 percent and a confidence interval (margin of error) of no greater than +5 percent."

More information on sample size for a generalizable sample and how it influences precision is in Appendix 1.

Back to Table of Contents

Limitations and Risks of Sampling in a Performance Audit Context

Limitations

Sampling in a performance audit context could be limited by the availability of time and resources. Selecting a sample, analyzing the sample to draw audit observations, and validating these observations with the auditees is often time consuming and expensive. In practice, this means that samples must be kept relatively small to perform the audit in a timely manner. Because detecting small variations requires large sample sizes, performance audits tend to focus on detecting large variations that indicate high levels of material error, such as non-compliance. For example, if auditors are analyzing a large social program, it would be more beneficial to investigate unusual variations between regions (which could be done with a reasonably sized sample) instead of searching for single instances of overpayments (which would require a much larger sample).[2]

Also, when sample size is limited by resource availability, there may be instances when it is not possible for an audit team to use a generalizable sampling approach to draw audit conclusions that meet a desired confidence level. Sample size calculation is a very important aspect of generalizable sampling. If the required sample size is too large to be practical for an audit team, then the team will not be able to draw a sample enabling extrapolations to the population as a whole. In this instance, the audit observations would be valid only for the smaller sample that the team could afford to analyze. This issue and factors driving it are discussed further in **Part 2** of this guide, in the subsections on optimizing for homogeneity and on time and resources available to complete the audit.

Another limit that audit offices may face with regard to sampling is the availability of expertise. Not every audit institution can afford to have a sampling specialist on staff and not every auditor is sufficiently knowledgeable about sampling methodologies and sampling software applications. A lack of expertise may limit an audit office's ability to use sampling methodologies effectively in its performance audits. One solution to this problem is to purchase expertise piecemeal by using consultants for specific mandates. (When used occasionally, this option can be less expensive than having a statistician or data scientist on staff.)

Risks

The limitations discussed above can also be sources of risk. For example, an audit team that does not have enough expertise on sampling may fail to consider an important aspect when developing a sampling plan or it may apply a rule of thumb in selecting a sample size without fully understanding all the implications. A lack of expertise may result in using an inadequate sampling methodology and producing audit observations that cannot be used for the intended purpose, therefore potentially wasting precious time and money. For these reasons, the level of expertise required should be assessed in the early phase of an audit.

Data quality is also an important risk. A successful sampling exercise is highly dependent on accurate and complete data. Without such data, sampling cannot result in reliable and credible observations. This issue is further discussed in the subsection on quality of available data in **Part 2** of this guide.

---

[2] Data analytics in conjunction with sampling can be used for detecting non-compliance and rare but material anomalies using large datasets. However, the success of a data analytic approach is always limited by the comprehensiveness and quality of available machine-readable data.

Sample size can also be a source of risk. Samples that are too small may result in findings that are easily challenged. Conversely, samples that are too large may result in dedicating too many resources to a specific line of enquiry in a performance audit, thereby decreasing the budget available for other lines of enquiry or other audits. Determining the right sample size to provide the information required to address a specific audit question is therefore a crucial success factor for audit teams.

Another risk that may arise from sampling relates to complexity. A complex sampling methodology may increase the risk of audit failure. This is because as a sampling methodology becomes more complex, it may also become more difficult to implement properly or to explain findings in plain language to the auditees, parliamentarians, or stakeholders. This has an impact on the ability to defend the resulting audit observations. It may also become more difficult to clearly state the methodology and its results in the final audit report. In contrast, simple and straightforward sampling methodologies are easier to explain and less likely to be successfully challenged by the auditees.

All these risks are subcomponents of the overall audit engagement risk as defined by audit standards used by performance auditors (CSAE 3001, A14). These standards recognize the risk that the procedures performed by the practitioner will not detect a significant deviation. Due to its very nature, sampling introduces potential error that a census would not. Like other evidence collection techniques available to auditors, sampling always entails a risk of leading to inaccurate observations, incorrect conclusions, or a plain failure to detect significant deviations in the subject matter. This risk can, of course, be mitigated by consulting experts, using prudent and deliberate sampling approaches, and leveraging existing methodological tools and resources such as the suggestions in this Practice Guide.

Back to Table of Contents

Practice Guide to Sampling Methodology for Performance Audits

Part 2 – Audit Methodology

Introduction to Sampling Methodology

This part of the Practice Guide is structured according to the five main steps that audit teams should follow once they have decided to use sampling as part of their performance audit's evidence collection strategy. These five steps are:

Step 1. Analyzing the Population – Once a population has been identified and accurate and complete data has been obtained from the auditee on this population, the audit team needs to analyze this data to determine the population's level of homogeneity.

Step 2. Selecting a Sampling Approach – This step requires audit teams to specify the objective of their sampling and to select a sampling approach (generalizable sampling or purposeful sampling) that will align with their objective.

Step 3. Preparing a Sampling Plan – This step involves making several important decisions, such as determining the sample size and selecting a sample method (i.e., how the sample will be drawn from the population).

Step 4. Executing the Plan – This is when the team puts the plan into action and revises the plan as circumstances warrant.

Step 5. Reporting the Results – The final step in the sampling process involves explaining and documenting the sampling methodology, reporting the findings derived from the sample, and stating whether the analysis is subject to any limitations (e.g., whether the findings from the sample can be extrapolated to the entire population).

This part of the Practice Guide follows these steps in order and covers both generalizable and purposeful sampling. In practice, they can be implemented sequentially, iteratively, or simultaneously.

Step 1 – Analyzing the Population

The first step in any performance audit sampling process should be a comprehensive review of the available **administrative data** (i.e., data collected by government institutions or agencies for public administration purposes) relevant to the audit topic. This applies to both generalizable and purposeful sampling.

Sampling Advice #1:

Investigate and understand the population of interest

Reviewing the available data will allow auditors to identify populations to assess whether these populations are sufficiently homogeneous, and to determine whether the data is of sufficient quality to make sampling possible. Completing this review will help auditors to take decisions about sampling objectives and appropriate sampling methodology. The review in itself may also provide useful evidence that might serve to support audit observations or as contextual information in the audit report.

There is no standard process for conducting a review of administrative data, but some basic techniques and approaches are commonly used by auditors (more on this below). Professional data analysts can also help through a wide range of standard exploratory techniques for initial review of population data.

Performance auditors can use two types of approaches to analyze populations: exploratory analysis and data analytics.

Exploratory Analysis

The goal of exploratory analysis is to gain a full understanding of the nature of a population and the data available to analyze it. Exploratory analysis is especially important before sampling as a means of assessing the level of heterogeneity of a population. Auditors typically use exploratory analysis techniques to examine the distributions of both continuous and categorical variables and the relationships between variables, as well as to identify outliers and anomalies that may warrant special attention. Auditors need to rely on analysis of distributions of categorical and continuous variables of interest available in administrative databases, as well as on the opinions of experts familiar with the population and the related data. Once this is done, variables of interest can be assessed for importance and a risk-based approach to examination can be developed for each.

Quality of available data

Auditors must ensure that they can rely on the data they receive from the auditees. They need to obtain reasonable assurance that the listing of the items in a population is accurate and complete.

Accuracy means that the information is correct—devoid of systemic errors and flaws that could lead to misleading results. To obtain this assurance, auditors have a number of methods at their disposal. For example, they can:

- compare information recorded in the sampling frame with original source information;
- implement tests to check for invalid[1] or out-of-range values; or
- examine the controls over the information contained in the sampling frame (Do they exist? Are they implemented? Are they effective?).

If concerns about quality of data are observed when attempting to conduct sampling activities, the auditors may report them as they would be identified reliably and credibly. In some instances, it’s also possible to obtain from the audited organization an attestation that their data is fit for use. Although this could provide a layer of additional credibility, some underlying audit work would still have to be undertaken it to validate it.

Completeness means that no information is missing in the records in the sampling frame. To obtain this assurance, auditors can do the following:

- Conduct a gap analysis, searching for gaps in fields that encode dates, locations, ID numbers, registration numbers, etc.
- Examine missing values.
- Compare with population totals reported elsewhere (e.g., number of centenarians reported by Statistics Canada based on census data versus the number of centenarians reported in databases of universal social programs).
- Do a sampling from an independent record of the population (e.g., original paper files).

Sampling Advice #2:

Ensure that you can rely on the available data

[1] Values that violate mathematical laws. For example, dividing a number by 0.

In some circumstances, external auditors can also rely on the work of internal audit to obtain assurance regarding the quality and completeness of the population data (CSAE 3001, 60).

In order to ensure that the data meet these quality requirements, the auditors should test controls over the preparation, security, and maintenance of the information.

Basic methods of exploratory analysis

Basic methods of exploratory analysis depend on the type of data being examined. For the most part, performance auditors will be dealing with combinations of categorical and continuous data. Other forms of data, such as text and geospatial data, do offer rich opportunities for analysis, but rarely form the basis of sampling frames in an audit environment. Exploratory analysis for most performance audits can be done with a limited number of basic tools, such as:

- descriptive statistics such as mean, **median**, variance, **skewness** (continuous variables)
- bar charts (categorical variables) and histograms (continuous variables)
- frequency distributions (categorical variables)
- cross-tabulations (relationships between pairs of categorical variables)
- scatter diagrams and bivariate correlations (combinations of continuous variables)
- box plots (combination of categorical and continuous variables)

As with any performance audit, the results of these tests can be combined with other audit procedures (such as inspection of procedural documents, benchmarking, or interviews of auditees) to understand program operations and the factors affecting them.

Optimizing for homogeneity

A population may be composed of distinct subgroups that are subject to different controls or different characteristics and therefore may need to be sampled separately.

To optimize a sample for homogeneity, it is essential to understand how many distinct groups exist in a population. (Of course, the number of groups can vary depending on the definitions of group membership.) A basic analysis using a bar chart can allow auditors to determine the number of distinct groups and the relative size of each group. The idea is to arrive at a manageable number of groups that still allows for a reasonable description of the range that exists in the population; there must also be a plausible rationale for expecting a relationship between the groups and the sampling results. In **Figure 7**, we show how this could be done by reorganizing the age group data used previously in **Figure 2**. The data, grouped by **strata** of 5 years in option 1, can also be presented by strata of 20 years (option 2), thus reducing the number of groups.

The analytical approach used to explore populations and to determine the number of subgroups is often iterative. For example, when analyzing the distribution of a population using a basic histogram, an audit team may need to make half a dozen attempts before it can determine the correct number and width of strata needed to adequately display the distribution of the population. It is worth noting here that the shape of the distribution is a useful indicator of the level of heterogeneity. Normal distributions (i.e., bell-shaped) usually indicate a low level of heterogeneity, while a skewed distribution (often seen when agreements in a population are broken down by value, as in **Figure 8**) will tend to indicate a high level of heterogeneity. This analysis is important, not only to break a population into meaningful subgroups, but to also to avoid, depending on the purpose of the sampling, creating artificial subgroups.

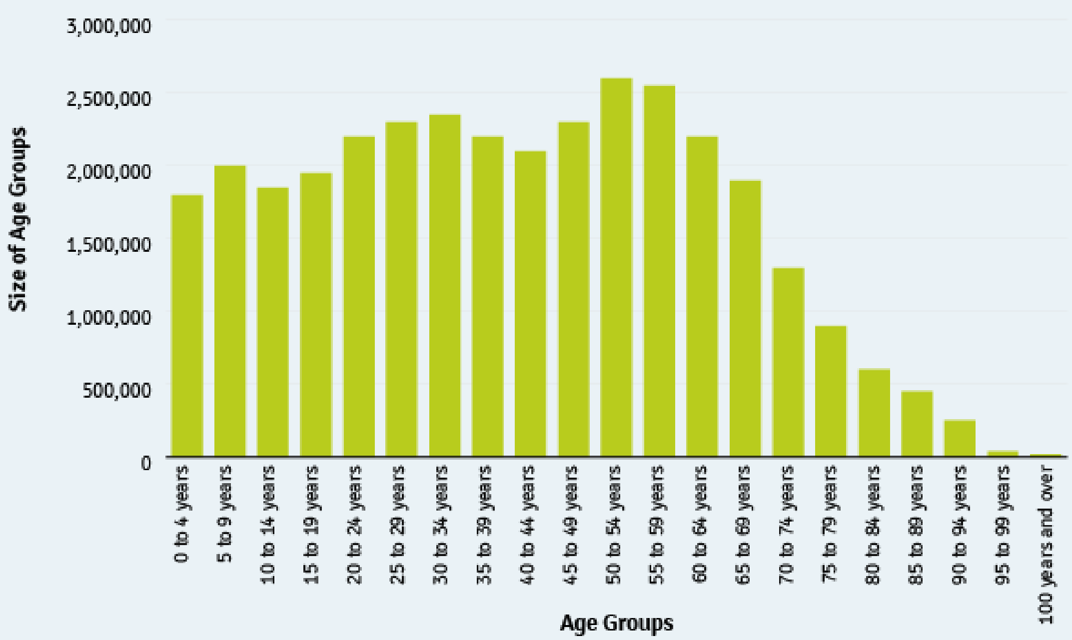
Once the subgroups are adequately identified, each one should be sampled and tested separately. Each subgroup should be categorized as representing a low, medium, or high risk of deviation or exception (i.e., non-compliance, overpayments, or other, based on the audit objective) and be sampled accordingly. Trying to assess an entire heterogeneous population with a single estimate of a variable under scrutiny will result in a conclusion that does not adequately capture the complexity of the entire population.

In the example in **Figure 8**, it is assumed that the auditors want to review the contribution agreements awarded by a department for the 2019–20 fiscal year. The funding level varies greatly between agreements. Many have relatively low funding levels (e.g., less than $300,000), while a few have high funding levels (e.g., $1 million to $2 million). How these contribution agreements are managed, and the risk associated with them, likely differs depending on the funding level of each contribution.

In this case, the contribution agreements should be batched into different populations based on funding level. As shown in **Figure 8**, three subgroups can be identified.

Figure 7 – Example of a Population Analysis Using Different Groupings

Grouping Option 1



Grouping Option 2

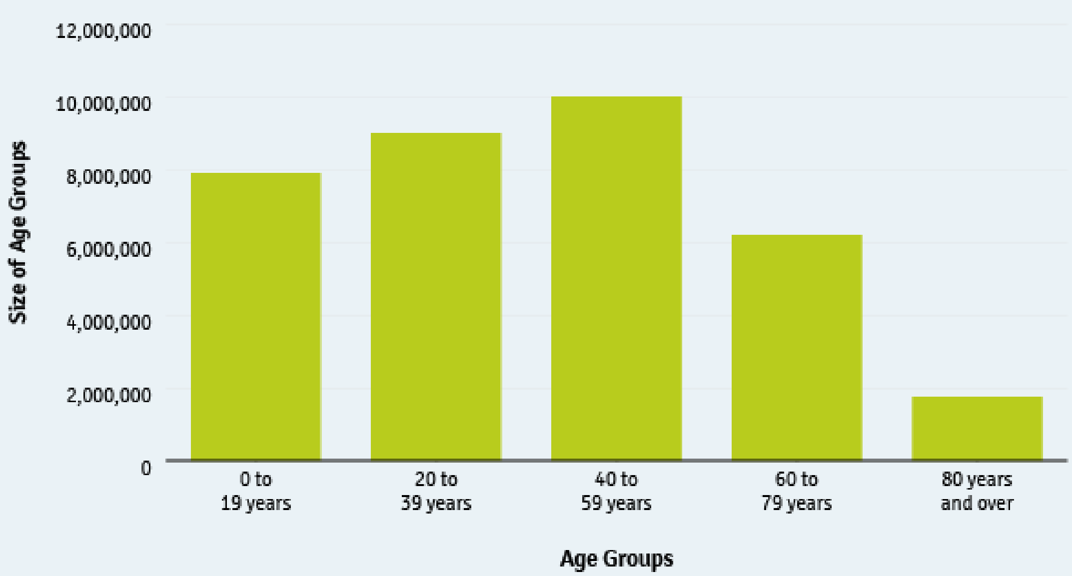


Figure 8 – An Example of Subgroup Identification

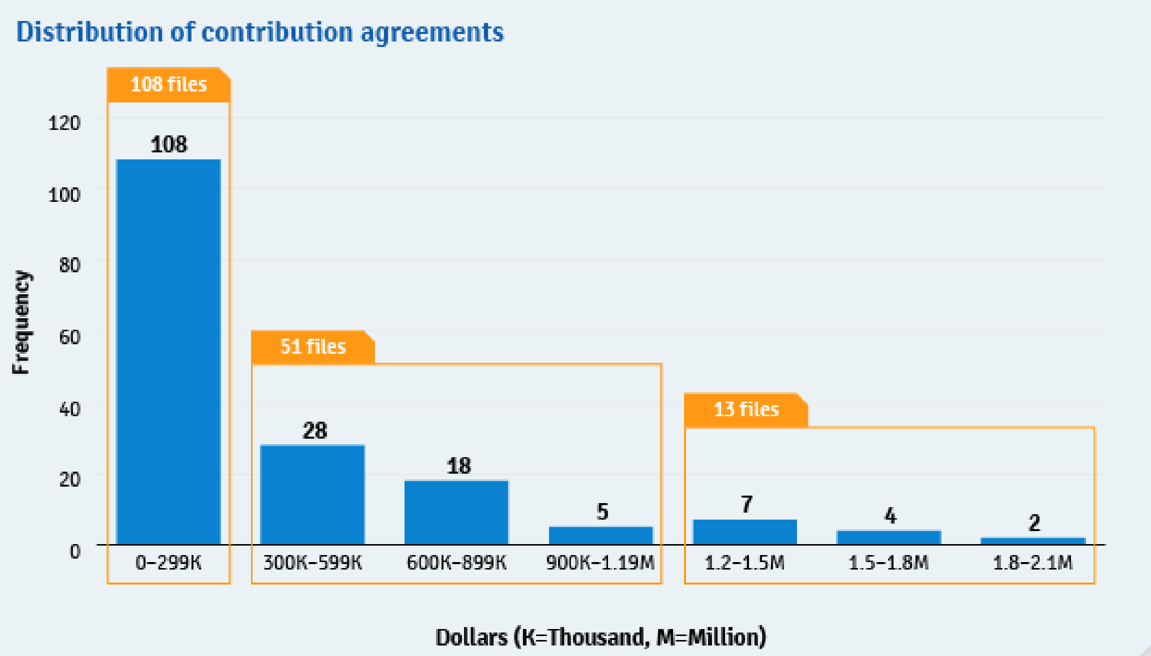


Table 1 provides a more detailed description of the differences among the three subgroups identified in **Figure 8**, as well as justifications for using different sampling strategies using a risk-based approach.

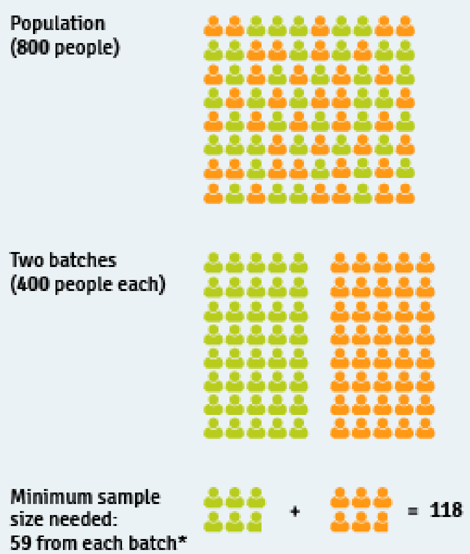
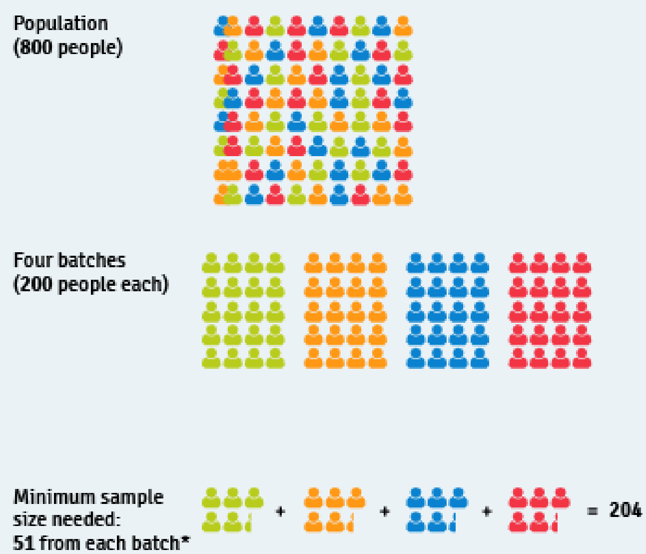
Table 1 – Comparison of Subgroup Characteristics

Subgroups	Characteristics	Sampling Strategy	Comments
Awards with a value between $0–$300,000	108 files – Low risk	Small sample	Population is large and risk associated with the files is low. Therefore, a sample with a high level of precision is not needed. (It could also be decided that no assessment is needed at all in light of the low risk level.)
Awards with a value between $300,000–$1.2 million	51 files – Moderate risk	Large sample	Smaller population with moderate risk. A sample with a high level of precision is justified.
Awards with a value between $1.2 million–$2.1 million	13 files – High risk	A sample is not needed—all files are to be examined	Very small population associated with high risks. A census (100% of the files) is warranted because it will provide the highest level of precision.

Analyzing the distribution of a population is especially important because the distribution (i.e., how many distinct groups compose the population) has implications for the number of samples that auditors will need to select and has a direct influence on how much time and resources will be required to complete the audit. To minimize the cost of audits, auditors should analyze populations with the goal of creating the smallest number of relatively homogenous subgroups within each population, while taking into account the level of precision required. **Table 2** compares two subgrouping strategies, highlighting the advantages and disadvantages of each one.

Table 2 – Example of the Effect of Using Different Numbers of Subgroups for a Survey

*Sample size calculated with confidence interval of 10% and confidence level of 90% using an online sample size calculator.

Two Subgroups	Four Subgroups
	
Pros	**Pros**
• Fewer members to survey (118) • Less costly	• More homogeneous subgroups • Able to report estimates for all subgroups
Cons	**Cons**
• Less homogeneous subgroups • Able to report estimates for only two subgroups	• More members to survey (204) • More costly

As illustrated in **Table 2**, having more subgroups implies increasing the costs of conducting a survey for an audit because, overall, more units (members, in this case) need to be selected and examined. This situation involves a trade-off for audit teams between the costs and homogeneity of the samples. Because there are often, in an audit environment, not enough resources to examine more than one or two small samples, the opportunity for auditors to drill down to subgroups may be rare. Therefore, it is incumbent on auditors to assess the level of heterogeneity of the populations before examination.

Data Analytics

Data analytics is a substantially different approach than exploratory analysis.

With exploratory analysis, the purpose is to gain a thorough understanding of the population's main characteristics. Information from exploratory analysis provides necessary information for developing a sampling plan. With data analytics, the intent is much broader. It can be used, for instance, to discover trends in a population, to detect anomalous activities and identify high-risk items, or to make inferences and predictions about a population. Leveraging artificial intelligence and other technological innovations, many software applications have been developed to conduct increasingly sophisticated analyses of population data in an audit context.

One main advantage of modern data analytics applications is that they can allow auditors to analyze 100% of the units in a population, thus making sampling unnecessary. However, this is only possible when a number of requirements are met. Ultimately, data analytics is dependent on the availability of data that is

- reliably collected,
- provided in a format that enables analysis by the auditors' software, and
- of sufficient comprehensiveness that it allows for making audit observations and conclusions.

In situations where administrative data is incomplete, poorly organized, or mostly available in non-electronic formats, a data analytical approach will be of little value and traditional sampling approaches will remain useful tools for auditors.

However, depending on the circumstances, it may be possible to combine data analytics and sampling methods. For example, it may be possible to use data analytics to do preliminary population analysis and identify groups of outliers or high-risk cases and then to use a sampling approach to examine in detail a number of these cases. (For a more extensive discussion of data analytics in performance audits, see European Court of Auditors, 2020.)

Back to Table of Contents

Step 2 – Selecting a Sampling Approach

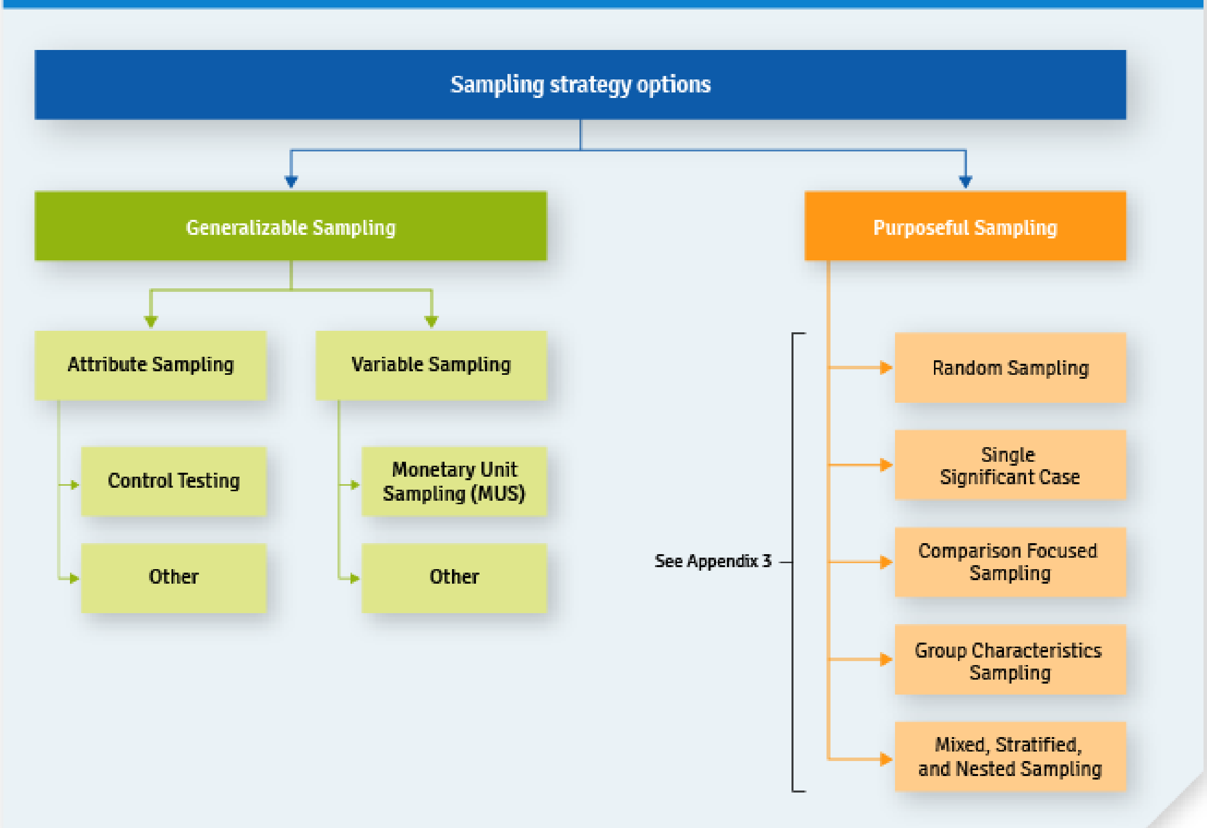
A very important decision auditors must take once they have determined that they can and will be using sampling to gather audit evidence is whether they should use a generalizable sampling approach or a purposeful sampling approach. This decision is governed by the need to determine which approach best enables the efficient collection of sufficient and appropriate evidence to meet the testing objective.

Selecting the right sampling approach will depend on

- the nature and complexity of the questions auditors need to answer,
- how generalizable to the entire population the sample needs to be,
- the characteristics of the population being studied, and
- the resources available.

This section explores the factors that auditors must take into account to reach this decision. **Figure 9** (an expanded version of **Figure 5**) provides a non-exhaustive overview of the options available to auditors. More methods are in Appendix 2 and Appendix 3.

Figure 9 – Sampling Approaches: Overview of Options Available to Auditors



Defining the Objective of the Sample

Auditors resort to sampling because they have a question they want to answer. This question can take many forms and its precise nature helps to determine what sampling methodology would be most appropriate to obtain an answer.

A central question to consider when defining the objective of a sample and selecting an appropriate approach is whether the audit team will need to extrapolate the results of its analysis to the entire population to be able to conclude on the audit objective. If yes, this necessarily implies a generalizable sampling approach (either attribute or variables sampling).

> **Sampling Advice #3:**
>
> Your sampling methodology must be aligned with your audit questions

For example, if auditors want to estimate the frequency of a certain characteristic in a population, such as non-compliance with a policy, then a generalizable sampling approach would be most appropriate because there is a need to extrapolate the result of the sample to the entire population. In this case, the best approach would be attribute sampling because of the binary nature of each sampling unit (each unit is either compliant or non-compliant). Alternatively, auditors may want to assess a program's impact in a population using a continuous variable such as dollars or size. In that case, a variables sampling approach would be appropriate.

A related question is whether the audit team wants to provide a quantitative measure of the sample's accuracy; that is, the confidence interval. If so, this also implies using a generalizable sampling approach. Similarly, if the audit team does not think it will be able to support its findings from the sample with other sources of information, then being able to provide a quantitative measure of sampling risks by using a generalizable sample would help the team to increase the credibility of its findings.

Another approach applies in cases where auditors seek to identify the causes and repercussions of specific problems without having to quantify, extrapolate, or measure their extent. In such cases, auditors can use their judgment to select sample units that exhibit characteristics of interest and there is no need to select the sample randomly or to extrapolate the analysis results to the entire population. A purposeful sampling approach is the most appropriate choice in these circumstances. A purposeful sample is also recommended if the auditors are aiming to demonstrate extreme variations or exemplary occurrences.

Other Factors to Consider When Selecting a Sampling Approach

The level of homogeneity of the population

Is the population relatively homogenous? If it is, a generalizable sampling approach using a simple random sampling strategy, or a more sophisticated method may be adequate. Otherwise, a purposeful sampling might be more appropriate. When auditors analyze the population, they will collect information that is key to making the right decision about strategy.

The availability of corroborating evidence obtained from other sources

When determining the sampling approach, like any other evidence collection strategy, auditors consider the global context of the audit approach. Is the evidence resulting from the sample indispensable to conclude on the audit objective? Does it complement other sources of evidence, or does it have to be interpreted in conjunction with other qualitative (testimonial or documentary) evidence to provide a more complete and convincing picture of the program's performance? The answers to these questions require considerable professional judgment that must take into account the evidence collection strategy as a whole.

The availability of statistical expertise and software

The ability to properly develop complex sampling plans is helped by the availability of statistical expertise, on staff or acquired through contracts. Practice guides, such as this one, or other resources (including those in the Bibliography section of this guide) are of course good tools to guide auditors, but they also have limits. In some instances, auditors may be limited in their capacity to implement more sophisticated approaches, either because of resource constraints or because they do not have an appropriate statistical software (or do not know how to use their software properly).

A specialized software application is a practical tool to assess key population characteristics and calculate sample sizes and other sampling parameters. Some information on this type of software is in Appendix 1.

The time and resources available to complete the audit work

All the above factors have a time or a resource component. Does the audit team have time to use the optimal sampling approach? Can a more modest approach achieve the same goal? Can the audit team afford hiring the expertise needed or purchasing a licence for the required software? Overall, does the audit team have enough resources to examine the number of sampling units needed to meet the requirements to conclude on the audit objective with sufficient and appropriate evidence? Ultimately, the decision may come down to a cost–benefit analysis.[2]

To guide auditors when they navigate these factors, **Table 3** summarizes the advantages and disadvantages of generalizable and purposeful sampling approaches.

[2] Auditors can rely on a sample already extracted by other auditors, assuming that it was done for a sufficiently compatible audit on the same population. The decision to rely on such a sample should be supported by an analysis of the sampling plan. (See the section on Documenting the Plan and Its Results for a list of items that should be documented.)

Table 3 – Advantages and Disadvantages of Generalizable and Purposeful Sampling Approaches

	Generalizable Sampling	Purposeful Sampling
Advantages	▪ Software applications are available to determine an optimal sample size. ▪ Yields statistically valid results (i.e., can extrapolate the results to the entire population). ▪ Sample is selected in an unbiased manner. ▪ Can quantify uncertainty (i.e., confidence interval and confidence level). ▪ Lends credibility to audit conclusions and results can stand on their own.	▪ Is flexible. ▪ Is well suited for some common performance audit circumstances, such as small population sizes. ▪ Uses the auditor's judgment to select sample size. ▪ Does not require specialized statistical knowledge or software. ▪ Allows for reasonable reliability at a reasonable cost.
Disadvantages	▪ Can be costly and time consuming. ▪ Not well suited for some common performance audit circumstances such as small population sizes. ▪ May require training to use software application.	▪ The results are not statistically valid (i.e., cannot be extrapolated to the entire population). ▪ It introduces (intentionally) a bias in the sample. ▪ No objective measure of sampling risk is provided (i.e., no confidence interval or confidence level). ▪ There is a risk that the sample size may not be optimal or adequate. ▪ Credibility needs to be enhanced by corroboration from other sources of evidence. ▪ Effectiveness depends on the audit team's skills.

Source: Modified from The Internal Auditor Professional Services Limited (n.d.), Sampling for Effective Internal Auditing, and from J. L. Colbert (1991), Statistical or Non-statistical Sampling: Which Approach Is Best?

Back to Table of Contents

Step 3 – Preparing a Sampling Plan

Once a sampling approach has been selected, the next step is preparing a detailed sampling plan. Audit teams that decide to use a generalizable sampling approach should consult the section Key Elements of a Sampling Plan – Generalizable Sampling. Audit teams that decide to use a purposeful sampling approach should consult the section Key Elements of a Sampling Plan – Purposeful Sampling.

Key Elements of a Sampling Plan – Generalizable Sampling

Before starting the audit's examination phase, the scope, methodologies, and objectives are documented and reviewed with the audited organization. If generalizable sampling is to be used during the audit to gather evidence, then it is suggested that a sampling plan could be developed and, if appropriate, discussed with the auditees before the examination phase. A good generalizable sampling plan should cover the following seven key elements.

Key element 1: Population analysis

Where generalizable sampling is concerned, there is usually some form of administrative database with a reasonable amount of quantitative data that can serve as a sampling frame. This data can also be used to perform basic population analysis. The plan should provide an assessment of the data quality.

The plan could also include histograms and bar charts that adequately describe the distribution of data and potential sources of heterogeneity. If this analysis indicates that there is enough heterogeneity to be a concern, it needs to be documented. If needed, the plan could contain a strategy to segment the population into more homogenous groups.

Key element 2: Objectives of sampling

The plan could indicate that this is a generalizable sampling approach and explicitly state what aspects of the sample are being extrapolated to the population. (Rarely is there a single type of error or deviation being measured.) This section of the sampling plan lists what population parameters are being estimated.

Key element 3: Level of accuracy desired

The plan could describe the level of accuracy desired for each segment of the population being sampled. Auditors should determine whether this level needs to be low, medium, or high for each segment of the population based on the expected quality of controls in place and the impact if risk is materialized. This determination also has to be contextualized based on other evidence available that in part mitigate the risk of inadequate controls. In some circumstances, sampling may be providing supplementary evidence to other audit procedures and this may also affect the level of accuracy needed.

Key element 4: Established expected error level

Although performance auditors may be able to rely on the experience of other auditors (internal or financial audits) on similar topics and in the same entity they plan to audit, they will often have to deal with a limited availability of empirical information that can be used to predict the level of error. In such instances, auditors should choose, and document in their plan, a level of error that is a high estimate and for which their **audit recommendations** would not change if the level of observed error were higher than expected. If the sample is insufficient to serve the purposes intended, it may be necessary to extend the sample.

Key element 5: Reportable findings

Before making any observations, auditors could state in their plan what conclusion would be appropriate for various levels of error that might be observed and reported.

Key element 6: Sample size calculation

Auditors could calculate an appropriate sample size for each segment of the population that has been defined and that they intend to sample from. The plan could indicate the parameters used to determine sample size and the specific software or method of calculation used. More information on how to determine the sample size for a generalizable sample is in Appendix 1.

> **Sampling Advice #4:**
>
> Properly estimate sample size, especially for generalizable sampling

Key element 7: Sampling method

The plan could describe the method of sampling for each segment of the population that has been defined and that the audit team intends to sample. The plan could indicate whether the method of sampling is simple random, systematic random, proportional stratified, non-proportional stratified, or some other method of sampling. (More information on unbiased methods of selection for generalizable sampling is in Appendix 2.) Auditors could be specific with respect to how random selection is accomplished and ensure that the method used can be substantiated as unbiased. Sampling software does offer log files that can substantiate unbiased sampling (i.e., the software keeps track of the **random seed** and all steps taken to determine the sample). Ideally, another person should be able to repeat the sample selection by using the same seed, software, and population. Otherwise, auditors could ensure that more than one person witnesses the selection process and is willing to substantiate that the chosen method was conducted appropriately.

Key Elements of a Sampling Plan – Purposeful Sampling

Before beginning the examination phase, the audit's scope, methodologies, and objectives are documented and reviewed with the audited organization. If purposeful sampling is to be used during the audit to gather evidence, then a sampling plan could be developed and given to the auditees before the examination phase. It is suggested that a good purposeful sampling plan should cover the following five key elements.

Key element 1: Population analysis

The plan could include the results of the population analysis conducted using the administrative data sets obtained from the auditees (including an assessment of the quality of the data sets). Whereas generalizable sampling sometimes requires a segmentation of the population in subgroups, this is not usually the case for purposeful sampling.

Key element 2: Objectives of sampling

The plan could indicate that a purposeful sampling approach will be used and explicitly state the sample's specific objective. Auditors could specify the assumptions, assertions, or hypotheses that are being tested, and what type of information is required to either confirm or deny these.

Key element 3: Sampling method

Based on the type of information being sought and the assumptions being tested, auditors could select, and document in the plan, a specific sampling method or mix of methods that will allow for definitive conclusions. Determining the specific method or mix of methods of sampling should be a collective decision that is discussed at length within an audit team. This will ensure that a firm selection rationale is established and that the chosen approach will provide reliable interpretation of observations. More information on basic methods of sample selection for purposeful sampling is in Appendix 3.

Key element 4: Reportable findings

Before making any observations, auditors could state in their plan what conclusion and recommendations would be appropriate for various levels of error that might be observed. Any conclusion based on audit observations needs to clearly explain how it is supported by the sampling methodology.

Key element 5: Sample size

The plan could provide a rationale for the initial sample size and the criteria for ending the selection process or expanding the sample.

Sample size for purposeful sampling is not a straightforward calculation. Sample size depends on many factors, including the method of sampling, the importance of the subject matter, the redundancy of findings, and the clarity of results. Even initial findings can influence sample size. More information on determining sample size for purposeful sampling is in Appendix 4.

Back to Table of Contents

Step 4 – Executing the Plan

During the examination phase of a performance audit, the auditors conduct the procedures they have described in their audit plans, including sampling procedures, and document the results of these procedures. However, in some cases, it may be impossible to implement the sampling procedures exactly as planned and then the auditors need to modify the plan to adjust to the circumstances.

Revising the Plan as Needed

The execution of a sampling plan, however well conceived and sound, always requires exercising professional judgment. New information on a sample's parameters can always be brought to the auditors' attention during the audit, and this information could affect the sample's size and reliability. New information can be obstacles to properly implementing the planned audit procedures, which may require modifying or abandoning them. Auditors should also be vigilant when they cannot examine some selected items. If a selected item is missing and it cannot be determined what happened to it, it should normally be considered a deviation.

In a generalizable sample, auditors must monitor the items as they are selected and analyzed and note all observed deviations. If the observed error or deviation rate is higher than the auditors planned for, this indicates that the sample size may not be large enough to render reliable conclusions (since the expected error rate is one factor influencing sample size). It could also mean that the information collected cannot be relied on and the sampling should be stopped. In both cases, changes to the sampling plan may be necessary.

> **Sampling Advice #5:**
>
> Evaluate your sample size during and after sampling for accuracy and sufficiency

For purposeful sampling, auditors may need to rethink the selection criteria based on whether or not the selected items are relevant to the particular investigation required by the audit. For example, if an audit aims to examine controls around both sole source and competitive procurements and the auditors take a purposeful sample of the highest value procurements, but soon find that their sample does not include many (or any) sole source procurements, then their efforts will be of limited value. In such a case, the auditors would have to rethink their sample. (One option could be to take a sample of high-value competitive procurements and a separate sample of sole source procurements.)

Documenting the Plan and Its Results

Performance audit methodology requires auditors to document their audit plans and the results of their audit procedures. This requirement applies to sampling plans and their results. It is a good practice for the working papers to include enough detail to describe clearly the sampling objective and the sampling method used. It is suggested that enough information is provided to allow another audit team to replicate the method used and the results that were obtained.

The items that auditors could include in the documentation of their sampling tests are:

- a description of the test's purpose (the question that needs to be answered)
- a definition of what constitutes a deviation, an error, or an exception condition
- a definition of the sampling unit and of the population, including the source of the data used and, when relevant, an assessment of the coverage error
- a description of steps taken to ensure that the population data was complete and reliable
- a description of how the population's homogeneity was assessed and of what decisions were taken based on the results (e.g., defining several populations, using stratified sampling)
- when applicable, the desired confidence intervals and confidence levels, the tolerable deviation or exception rate, and the expected population deviation rate
- the chosen sample size and how it was determined
- the sampling approach used (generalizable or purposeful) and the sampling parameters
- in generalizable sampling, reasonable evidence that the sample was selected in an unbiased manner or attestations from two or more individuals regarding the steps taken for random selection
- confirmation that the results of relevant IT controls testing were considered before starting sample testing
- a list of the selected sample items
- a description of the examination procedures performed on the selected items
- the results of the examination procedures on the selected items, including, when relevant, a description of procedures conducted to assess any observed deviations
- a short statement of the overall conclusion in relation to the objective of the sample
- a description of how the sample results contributed to forming the audit's conclusion (including, for generalizable samples, whether the results were extrapolated to the whole population)

It is suggested to keep this information in the audit files for each performance audit that relies on sampling to obtain audit evidence.

Of course, this is a lot of information, and it cannot all be included in the final audit report. The next section, Reporting the Results, explains what should be presented in the report.

Back to Table of Contents

Step 5 – Reporting the Results

During the reporting phase of a performance audit, auditors produce a report that presents their audit observations and conclusions. Audit reports vary considerably in scope and nature. In addition, the formats and writing styles of performance audit reports are specific to individual audit offices. As a result, there is no standard way to present audit findings. However, there are a few general principles that auditors should apply when reporting audit findings that were obtained through generalizable or purposeful sampling.

Describing the Sampling Methodology

Transparency is a good practice when explaining the methodology and approach used by the audit team. It is also indirectly required by audit standards used by performance auditors that call for, where appropriate, a description of any significant inherent limitations associated with the measurement or evaluation of the underlying subject matter against the applicable criteria. Transparency ensures clarity while informing recipients of the audit report of the audit findings' rigor and overall reliability. This is especially true when a form of sampling has been used. While it is not necessary to report all the details that should be included in the working papers (see the section Documenting the Plan and Its Results), readers should be able to form an educated opinion with respect to the reliability and soundness of the sampling methodology used. It is suggested that the key elements that could be included in the audit report are:

- sampling approach (generalizable or purposeful)
- size of the population and size of the sample
- confidence interval and confidence level (for generalizable sampling)[3]
- explicit description of rationale for method of selection (for purposeful sampling)[4]

An example of a sampling methodology description is in **Text Box 3**.

[3] This provides the information required regarding significant inherent limitations of the measurement as required by CASE 3001 73.
[4] Idem.

Text Box 3 – Describing a Sampling Methodology: An Example

Audit: Office of the Auditor General of Canada – Preventing Illegal Entry into Canada; published Fall 2013

Description of the sampling methodology:

"In addition, we conducted file reviews of the following representative samples of border-related activities:

- a sample of 102 flights (45 from the United States and 57 from other international destinations) from a population of 61,684 carrying 5.6 million passengers arriving in Canada between 1 September and 30 November 2012, and a sample of 306 passengers from the population of approximately 15,000 on those 102 flights, to assess whether complete advance passenger information was provided to the Agency;
- a sample of 49 targets from a population of 998 targets issued in March and April 2013, to assess whether air passenger targets were working as intended; and
- a sample of 49 cases from a population of 1,427 Integrated Border Enforcement Team reports and a sample of 43 cases from a population of 476 Marine Security Enforcement Team reports in the RCMP's Police Reporting and Occurrence System and its Police Records Information Management Environment databases for the fiscal years 2011–12 and 2012–13, to assess whether those teams were able to respond to known border occurrences.

The results for each of the random samples is considered accurate to within 10 percent, 9 times out of 10."

Reporting the Results from a Generalizable Sample

The results from a generalizable sample are relatively easy to report. The intent of generalizable sampling is relatively narrow: extrapolating sample results to the entire population. While there may be variations in how a sample is generated, its result and the interpretation of this result should remain fairly uniform. Plus, most people have an intuitive sense of what a generalizable sample is.

This does not mean that auditors should not be careful when reporting the result of a generalizable sample. To illustrate this point, consider these two ways of reporting a finding obtained using generalizable sampling:

- *We found that 18 of 20 Income Assistance files we reviewed (90%) did not meet one or more key program requirements.*

or

- *We found that 90% of the Income Assistance files did not meet one or more key program requirements.*

Only the first version provides information on the extent of testing done.

Sampling Advice #6:

Report only at the level your sampling approach allows for

Reporting the Results from a Purposeful Sample

Greater care needs to be taken when reporting the results from a purposeful sample than from a generalizable sample. Most people are familiar with generalizable sampling and how it is used to measure the level or magnitude of population characteristics. This familiarity may be the reason why the results from purposeful samples are incorrectly reported as if they were the results from a generalizable sample. The results from a purposeful sample cannot be used to extrapolate to the population as a whole. An explicit clarification of this limitation should always be included in the description of the selected purposeful sampling method.

What can be reported depends entirely on how the sample was selected. Each purposeful sampling method introduces bias into the sample for the intent of learning something specific about the population. The reporting of results should clearly state that intent and be specifically limited to it.

When reporting the results of a purposeful sample, it helps to focus on six key questions about the methodology and the evidence. These are:

1. How was the sample selected?
2. What type of information was this sample expected to yield?
3. What did the auditors look for?
4. What did the auditors find?
5. What can be inferred from the results given how the sample was selected?
6. What other information corroborates this inference?

How was the sample selected?

While many people understand generalizable sampling, the same cannot be said of purposeful sampling. One reason that explains this difference is that the variation in methods is much larger for purposeful sampling (see Appendix 4 for descriptions of various methods), and this variation has a large influence on how to interpret results. For these reasons, it is important to describe in simple language to the reader of an audit report how the purposeful sample was selected.

What type of information was this sample expected to yield?

The key to using purposeful sampling appropriately is to provide a logical rationale between the method of selection and the types of conclusions one can make. Auditors should describe how their method of selection can logically allow them to draw a specific conclusion. Examples are in **Text Box 4**.

Text Box 4 – Theoretical Examples of Descriptions of Sampling Methods and their Potential

Example 1 – Selecting a sample with the widest possible variation in cases:

Generating a sample that includes the widest possible variation of cases will allow the audit to conclude that the potential for risk exists in practically any area of the population and is not restricted to a narrow segment as previously assumed.

Example 2 – Using a single index case to demonstrate the existence of a serious threat:

The common practice of open USB ports on government computers has the potential to result in widespread introduction of computer viruses throughout government computer systems. We identified one case that shows that the risk is real. The user in this case had a high level of computer security knowledge and his computer configuration had a reasonable level of software protection. Despite these advantages, the use of open USB ports and common use of flash drives resulted in a system-wide infection.

What did the auditors look for?

The report should describe the type and extent of examination performed. Examination of cases in purposeful sampling is usually broader and more extensive than in generalizable sampling.

What did the auditors find?

The report should describe the findings objectively. It is a recommended good practice for auditors to avoid reporting rates or statistical proportions derived from a purposeful sample, especially when small numbers are involved, because readers may assume the findings can be generalized to the population. It is better to report raw figures (not percentages) and use examples to illustrate typical findings. If rates or statistical proportions are reported it is recommended to specify clearly that they relate to the sample, not the population.

What can be inferred from the results given how the sample was selected?

Inference from purposeful samples is dependent on what was found, how the sample was selected, and the uniformity of the results. While interpretation of findings from a generalizable sample is a relatively simple exercise of extrapolation, interpretation of findings from a purposeful sampling requires the establishment of a logical argument. Given the manner in which the sample was selected, what can the results tell the readers with certainty? It is also important to remind the readers of what cannot be inferred from the finding, such as a prediction of population error rates.

What other information corroborates this inference?

Finally, remember that evidence from purposeful sampling has a higher need for corroboration from other sources of evidence (what is sometimes called triangulation of evidence) than findings from generalizable sampling. When designing the audit approach, auditors should plan for multiple methods of inquiry to provide information from different perspectives, which will help in forming a cohesive argument.

An example of a good description of a purposeful sampling approach is in **Text Box 5**.

Text Box 5 – Example of Reporting Findings from a Purposeful Sample

Audit: Office of the Auditor General of Canada – Indian and Northern Affairs Canada – Meeting Treaty and Land Entitlement Obligations; published November 2005

Findings from the audit (excerpt):
Land selection files illustrate the need for improved management practices

7.48 We reviewed land selection files in both Saskatchewan and Manitoba to gain insight into particular problems and assess departmental performance against stated commitments. To accomplish this, we extracted two purposeful samples. In the first sample, we asked both Saskatchewan and Manitoba to identify their best cases and more problematic cases, for a total of 24 files. The function of this sample was to determine what the Department determines as successes and failures. In our second sample, we selected 44 files representing a cross-section of land selection situations with specific characteristics, such as the location (urban or rural, northern or southern), affiliation (part of multi-party framework agreement or an individual agreement), and status (complete or still in process). These cases were reviewed and compared with the best and problematic cases.

7.49 In both samples, we found that most files contained the key documents required to the point that the files had progressed in the process. The majority of completed files were from Saskatchewan, whereas Manitoba had a significant number of files still in the process. Overall, we found little evidence of communication with individual First Nations, such as notes-to-file and records of meetings—documentation necessary for properly managing selections.

7.50 Most significant was the considerable variation in file management methods by individual project officers. This ranged from detailed and comprehensive checklists, to periodic notes-to-file, to very little file management whatsoever. This finding is of particular concern when coupled with the level of staff turnover noted among project officers responsible for the selections in each sample. Some files had as many as four project officers in six years. If information is not systematically collected and recorded, then it quickly becomes an issue of high risk when coupled with staff turnover and the Department's unreliable data systems.

7.51 Within the sample of best and problematic cases, almost all of the files that the Department considered a success had proceeded or were proceeding through the process with few complications. However, we noted that the processing times were much faster in Saskatchewan than in Manitoba. Conversely, most of the files that the Department considered problematic were more complex, involving third-party interests, difficulties with municipalities, and procedural matters (for example, land selections deemed ineligible under the framework agreements).

7.52 In our second sample, we found that the majority of files had the same characteristics as the problematic cases in our first sample. Very few cases had gone through the process without complications, and most were delayed at some point in the process. Delays were most often caused by third-party interests, such as concerns of municipalities or issues related to natural resources (minerals or oil and gas). When the files remained stalled for several years, there were no plans to resolve the impasse. Conditions attached by the Department to regional approval-in-principle, a document indicating the region's support for selections, were another cause of delays. Given that most of the files in our second sample exhibited the same characteristics as the problematic files in the first sample, we believe there are critical issues that must be addressed in planning and file management.

Back to Table of Contents

Appendix 1 – Sample Size Determination for Generalizable Sampling

Sample Size

Several factors influence the sample size required for an audit:

- level of detail for reporting
- required level of precision
- expected or observed error rate
- population size

Level of detail for reporting

Level of detail is the degree to which auditors wish to drill down into data and report on segments of the sample as well as overall results. Obviously, a high level of detail requires increasing the total sample size. In situations where a very large overall sample is required, efficiencies can be gained through **stratification** (i.e., a sample per subpopulation) and planning the minimum sample required for each reported finding. Weighted proportions can be used when calculating overall findings.

Any reported finding needs to have a large enough sample that meets the auditor's requirement for precision and reliability of results. Auditors should realize that starting with a minimal sample size does not allow for any drilling down of results. If detailed results are required for appropriate observations and recommendations, then sufficiently large samples must be planned to accommodate reporting of subsets of sample data. But if not, a small sample may be sufficient.

Required level of precision

The level of precision needed of any sample depends on the topic of the line of enquiry. If the objective is to determinate, for example , if there is some non-compliance with a specific requirements and the exact level of error is immaterial, then a small sample will be enough.

In situations where smaller levels of error need to be detected or in situations of high importance (e.g., human health and welfare are directly affected), a confidence interval of 5% with a confidence level of 90% would be more appropriate. In the most extreme cases, where the subject is of high materiality and there is a need for a very accurate measure of error rate, then a confidence interval of 5% and confidence level of 95% would be more appropriate. **Table 1A** provides further information and examples for each of these situations.

Table 1A – Three Levels of Precision for Audit Samples Based on Relative Materiality

Sample Precision	Context	Confidence Interval	Confidence Level	Sample Size* EE=5%/ EE=20%
Low	Auditors are interested in detecting moderate to high levels of error in areas of moderate importance. An approximate measure of error is sufficient to justify a recommendation. Example: An audit of a grants and contribution program where standard and effective controls are in place and there is no indication of major non-compliance. A generalizable sample is used to confirm effective use of controls.	10%	90%	34 / 45
Moderate	Situations where even low levels of error have significant material impact and a precise measurement of error is required to determine what recommendation to make. Example: Audits of procurement practices including tendering of contracts, management of invoices, and use of acquisition cards. The potential for overpayments and wrongdoing is high given the nature of procurement with large numbers of transactions and the potential for individuals to gain financially either through direct theft or influence peddling. Even limited overpayments can have major impacts (both on finances and on the reputation of government departments).	5%	90%	76 / 125
High	Situations where errors result in significant impact to human health and welfare. A high confidence and precision in results are required to justify recommendations. Example: An audit of maintenance and sanitation of medical equipment where improper procedures and inconsistent application of controls can result in multiple infections, long-lasting health impacts, and death.	5%	95%	99 / 180

*Two sample sizes are provided assuming a large population and two different levels of expected error (EE): 5% expected error and 20% expected error. Population size used for these calculations is 1,000 units. Sample sizes were calculated using CaseWare Analytics IDEA, version 10.

Expected or observed error rate

There are two strategies auditors can use to decide on an estimate of expected error. The first is to assume the highest possible level of variance; that is, an error rate of 50%. This maximizes the sample size. It is an easy and safe strategy, but also the costliest.

An alternative strategy is to consider how different levels of error affect conclusions and recommendations. Starting at 0% error, auditors postulate conclusions and recommendations for increasing levels of error. At some point, the increasing level of error no longer has any impact on conclusions or recommendations. Depending on the situation, that point might be an error rate of 10%, 20%, or 30%.

To illustrate the effect of expected error, several scenarios are depicted in **Table 1B**. Three options for expected error are presented: 5%, 20%, and 35%. As the expected error increases toward 50%, the respective sample sizes also increase—and so does the amount of variance. In order to maintain the same level of precision (5% confidence interval), the sample size must increase to compensate. For these examples, we have chosen a 5% confidence interval at a 90% confidence level as the required level of precision.

Table 1B – Impact of Expected Error on Audit Observation

Expected Error (%)	Required Confidence Interval (5%)	Sample Size*	Observed Error (%)	Observed Confidence Interval (%)	Audit Observation
5% Error	±5%	76	3 (4%)	±4.4%	Predicted error rate is 4%
			12 (16%)	±6.4%	Predicted error rate is greater than 5%
20% Error	±5%	125	24 (19%)	±4.9%	Predicted error rate is 19%
			50 (40%)	±5.7%	Predicted error rate is greater than 20%
35% Error	±5%	154	44 (29%)	±4.7%	Predicted error rate is 29%
			72 (47%)	±5.1%	Predicted error rate is greater than 35%

*Sample sizes were based on the following parameters: Population size of 1,000, confidence level of 90%, and confidence interval of 5%. Sample sizes were calculated using CaseWare Analytics IDEA, version 10.

Two potential observed error rates are presented for each case; one below the expected error, and one higher than the expected error. When the observed error is below the expected error, the observed confidence interval is smaller, more precise, than the required level. As a result, the observed error rate can be reported in an unqualified manner.

When the observed error rate is higher than the expected one, the observed confidence interval is now larger, less precise, than the required confidence interval. As a result, the sample is no longer large enough for the specified confidence interval and confidence level; it is not reportable at the accepted confidence level and confidence interval. A possible solution may be to increase the sample size, time and resources permitting. It is important to keep in mind that a higher error rate indicates other potential audit issues, which may warrant reassessing the overall risk assessment.

A lower expected error does result in a smaller sample size, and is therefore less costly, but it also limits the range of possible audit observations. The strategy of always choosing a 50% error rate provides the greatest flexibility in reporting but it also maximizes the sample size and increases the audit's cost. A more reasonable strategy is to choose the lowest possible error rate that would trigger the most drastic recommendation; that is, the point at which an error becomes so large that the most drastic recommendation is warranted.

Population size

Sample size is only mildly influenced by population size, and usually only when the population is very small. This is because variance in a population quickly reaches a plateau as the size of the population increases. That is, the amount of variance in a population of 1,000 is about the same as the variance in a population of 3 million. As a result, both populations require similar sample sizes despite the differences in population size. For extremely small populations, the required samples sizes diminish slightly. Regardless of population size, sample size should always be calculated in the same manner. A good audit data analysis software will have a built-in utility for sample size calculation.

Calculating Sample Size

It is advisable to use a sample size calculator for determining appropriate sample size for any generalizable sample. However, there are major differences between available utilities or applications.

Online sample size calculators

Online sample size calculators from reputable sources are accessible and easy to use. Available features vary, but most allow the user to adjust population size, confidence level, and confidence interval. However, not all online utilities allow the user to adjust the expected error. The major difficulty with practically all of these utilities is the distribution of variance that is used to make the calculation. Almost exclusively, online utilities use a basic t-distribution for calculating sample size. This method is more appropriately used for normally distributed data. These online calculators should not be used for calculating sample sizes for audits, especially when the expected error is less than 30%.

The input variables for attribute sampling can easily be translated into terms used by the IDEA utility. **Table 1C** summarizes and compares the input variables needed for attribute sampling and how to translate these into the input variables that appear in the IDEA sampling utility (version 10). The value of tolerable error has to be calculated based on the desired confidence interval and the expected error level.

Table 1C – Adapting Attribute Sampling Variables for Common Software Sampling Utilities

	Typical Input Variables for Attribute Sample Size Calculation				**Input Variables for IDEA Attribute Sample Size Calculator**			
Input Variables	Population Size	Confidence Limit	**Confidence Interval**	Expected Error	Population Size	Confidence Limit	**Tolerable Error**	Expected Error
Formula	N	CL	**CI**	EE	N	CL	**CI+EE**	EE
Ex 1	500	90%	**5%**	10%	500	90%	**15%**	10%
Ex 2	500	90%	**5%**	15%	500	90%	**20%**	15%
Ex 3	500	90%	**10%**	20%	500	90%	**30%**	20%

Back to Table of Contents

Appendix 2 – Unbiased Methods of Selection for Generalizable Sampling

Auditors must strictly adhere to using random selection to ensure the integrity of sample results and to defend the results during a clearance or fact validation process with the auditees. Random selection tools include random numbers tables, computer-generated numbers, and utilities to generate randomly selected samples. All these options can be used to ensure that sample selection is unbiased.

There is a tendency to consider haphazard or arbitrary selection (i.e., trying to create a true **random sample** by haphazardly choosing items) to be equally good or good enough to ensure unbiased selection. However, despite the honest intentions of auditors to select a sample that will fairly examine the population, haphazard selection methods are not appropriate for statistical sampling. Haphazard selections can introduce bias into a sample. Factors such as convenience, interest, or prominence do influence selection decisions when not controlled through an objective process.

There are several methods of objectively selecting a sample.

Simple Random Sampling

This involves collecting a sample in a way that ensures every member of the population has an equal chance of selection.

Systematic Random Sampling

After randomly selecting a starting point in the population between 1 and n, every n^{th} unit is selected, where n equals the population size divided by the sample size.

Cluster Sampling

Cluster sampling is a generalizable sampling technique where the population is divided into multiple groups (clusters). Random clusters are selected with a simple random or systematic random sampling technique for data collection and data analysis.

Proportional or Non-proportional Stratified Sampling

The population is subdivided into homogenous groups, for example regions, size, or type of establishment. The strata can have equal sizes or may be based on a higher proportion in certain strata. The specific type of sample selection used depends on the nature of the population, sampling frame, and analytical requirements.

For most circumstances in performance auditing, a simple random sample is sufficient. If obtaining audit observations requires a highly complex methodology, whether sampling or otherwise, it can add significantly to the effort required to communicate and defend the results.

Back to Table of Contents

Appendix 3 – Basic Methods of Sample Selection for Purposeful Sampling

There are many different ways to select a sample for purposeful sampling. This appendix provides examples of basic methods in each of four main categories:

- single significant case
- comparison-focused sampling
- group characteristics sampling
- mixed, stratified, and nested sampling methods

It also includes two illustrative case studies in which the context of a real audit report is used as a basis from which to consider how an additional sample method might have been selected and applied.

This material is largely based on Patton (2015). The U.S. Government Accountability Office (2017) also provides a useful document that includes information on many of these sampling techniques, as well as examples.

Single Significant Case

The first class of purposeful sampling methods includes methods where just one example is enough to provide the evidence necessary to make a reliable conclusion. In an audit environment, these critical case situations can be used to exemplify how a weakness in a control can be exploited or to give a first-hand account from an auditee’s perspective.

Types of single case sampling methods include:

- Index case: The first documented case to manifest a phenomenon
- Critical case: A case that allows for logical generalization to other situations
- High-impact case: A case that resulted in a larger than normal level of material impact
- Self-study: An examination of one’s own experience when subjected to a process, program, or service
- Teaching case: A well-documented case study that provides a revealing perspective on the topic
- Exemplar case: An in-depth examination of a case that illustrates an important dimension and allows for a more thorough examination

Index Case

A single highly informative case could be the first known instance of a new phenomenon. Fox example, given the rapid advancement of information management and information technology, it is not unreasonable that controls developed even a just a decade ago might not be enough to deal with new IT security risks. With an emerging IT risk, it is often the first discovered case that results in new controls. The low incidence rate is not the determining factor in deciding whether to address the threat. The potential importance of *not* addressing the threat immediately upon discovery is what matters here.

Critical Case

Critical cases are typical cases where one or more major controls failed to mitigate or to avoid risk. This failure demonstrates that if the controls failed in this instance, they can fail anywhere. Although critical cases do not allow for statistical generalization, they are a form of plausible logical generalization.

The sinking of the Titanic can be considered a critical case. It was not the first case of individuals perishing from a vessel sinking at sea and there was not much unique about this case. Indeed, there was not much to learn about sinking vessels from this case. In fact, it was its ordinary circumstances that made this a critical case. This particular vessel should not have been lost in an ordinary sinking. It was designed to remain afloat long enough to allow for the rescue of its passengers. Hence, it only needed enough lifeboats to ferry the passengers to a rescue vessel. As a result of this ordinary sinking of an extraordinary vessel on its maiden voyage, both the British and U.S. governments introduced several new regulations to mitigate the risk of catastrophic sinking. Why? If it could have happened to the Titanic, then it could happen to any vessel.

High-Impact Case

High-impact cases deal less with the control of risk and focus more on the level of importance. Given that the effort to control for risk is a function of both likelihood and importance, a program may dedicate little effort to control for some risks that are considered to result in minor exposure or harm. A high-impact case can demonstrate how the level of importance has either changed or was underestimated.

Self-Study

The self-study is substantially different from other single-case methods because the researchers or auditors need to subject themselves to the program or service of interest to obtain the necessary first-hand perspective that normally only program recipients or service clients would have. In this method, an auditor adopts the role of the client and tries to navigate normal procedures and controls with the aim of gathering first-hand evidence for programs involving the delivery of services to people. Alternatively, this method, often used in investigations, could also be used by an auditor adopting the role of a criminal or terrorist with moderate means and resources in order to discover vulnerabilities of programs that could pose a significant and immediate threat to public safety. The potential ethical risks of this approach should be considered and may require appropriate safeguards (e.g., documentation of ethical risks, implementation of ethical safeguards such as privacy and "no-harm" stance).

Case Study #1: Self-Study

Audit: U.S. Government Accountability Office – Use of Covert Testing to Identify Security Vulnerabilities and Fraud, Waste, and Abuse; published November 2007

Posing as private citizens, investigators purchased sensitive military equipment—including ceramic body armour inserts, guided missile radar test sets, and microcircuits used in F-14 fighter aircraft—on the Internet from the Department of Defense's liquidation sales contractor. This demonstrated and documented an existing threat to public safety.

Teaching Case

Teaching cases are not necessarily high impact or exemplar, but what they might lack in dramatic appeal, they make up for in how well they are documented and understood. Having a case that has been well dissected allows the rest of the world to explore it and learn from it. Unlike other types of single case sampling methods, the teaching case is a combination of finding a potential case and exploring it with enough depth so as to create a teaching case. A teaching case is equal parts opportunity and research.

Case Study #2: Teaching Case

Audit: Office of the Auditor General of Canada – Audit of the Privacy Commissioner; published September 2003.

When exploring the risks associated with poor leadership, one might not consider looking to a small department or agency given that levels of importance are always considered to be higher within large departments. However, the case documented by the Office of the Auditor General of Canada's 2003 audit of the Privacy Commissioner is a perfect teaching case. Despite the small size of the Office of the Privacy Commissioner, the extent of abuse and impact on lives identified in this audit remains an outstanding example of the negative and long-lasting impact poor leadership can have.

The working atmosphere in the Office of the Privacy Commissioner was described in the report as a "reign of terror"; there were repeated instances of humiliation, inappropriate comments, intolerance, and verbal abuse. Favouritism was also common: there were cases of individuals rewarded with over-classification of positions, biased hiring practices for friends, inappropriate cash-outs of vacation leave, and abuse of travel and hospitality privileges. The amount of documentation of this case and the impact it had is what makes it a teaching case. This particular case can be used for years to come to help individuals understand the impact and anatomy of abuse of power.

Exemplar Case

An exemplar case provides a more in-depth perspective to a problem and allows for an analysis of how a problem repeatedly manifests itself when no intervention is applied. As with the teaching case, such cases are rare in the field of auditing given that the periods covered by most audits are relatively short.

Generic Example #1: Exemplar Case

To see whether federal policies caused problems in port operations, a national audit office could examine the port of a large metropolis, which is diverse and has a high work volume. Problems would be likely to show up at this site and at others; if no problems were observed at the port the large metropolis, problems were unlikely at other sites.

Source: GAO (2017)

Comparison-Focused Sampling

Comparison-focused sampling takes advantage of naturally occurring variation within a population to gain insight into the conditions that lead to either success or failure. Cases can be selected based on specific criteria for either comparison to the norm or comparison between two or more groups. Criteria for selection can include outliers, intense manifestation of a phenomenon, positive deviations, or a particular criterion of interest.

Types of comparison-focused sampling include:

- Outlier sampling: Cases with unusually high manifestations of a phenomenon
- Intensity sampling: Information-rich cases that exhibit a phenomenon intensely
- Positive-deviation comparisons: Examinations of cases that demonstrate novel solutions to problems
- Matched comparisons: Comparisons of two groups that differ along an important dimension in an attempt to understand or explain the difference
- Criterion-based case selection: The selection of cases that meet an important criterion and then the performance of a comparison to the general population of cases
- Continuum sampling: The selection of a sample with cases that lie along a continuum and the making of observations regarding observed differences and relationships

Outlier Sampling

Outliers are cases that do not conform to the norm. Normally, it is the outlier that is discarded in favour of cases that more accurately represent the norm. When trying to extrapolate from a sample to a population, it is appropriate to be wary of outliers. Using a baseball analogy, it is the pitcher's wild throw that deviates from the usual controlled trajectory within or just outside the strike zone. However, if one were interested in assessing the quality of a catcher, which situation would one rather observe: The catcher's ability to catch the pitch perfectly thrown to exactly where the catcher indicated, or how he or she reacts to wild, unpredictable throws?

Outlier situations allow auditors to assess the effectiveness of mitigation strategies and gain a better understanding of how both people and controls perform under stress.

Intensity Sampling

Intensity sampling is likely to be a standard sampling practice for many audit situations where a particular phenomenon is of interest and auditors wish to better understand its causal nature. For example, how does having a dedicated data management specialist affect departmental performance measurement efforts? Does employee engagement result in more innovation and change in program delivery methods? In trying to answer questions like these, auditors can seek out situations where these phenomena appeared in an intense fashion.

Positive-Deviation Comparisons

Typically, auditing is thought to be focused on the discovery of deviations and exceptions. However, some of the most powerful recommendations are likely to come from examining exemplary performance, not poor performance. Where typical performance across a population ranges from poor to mediocre, having one or two cases of stellar performance (i.e., where challenges have been surmounted) can provide valuable insight. For how long have these cases shown positive results? What process was used to define the challenges or problems? What change management approach was used to bring about corrective action? How long did it take? How can this process be transferred or replicated?

Matched Comparisons

A matched comparison method is used to create two separate groups, with relatively similar characteristics, except for one major difference along a dimension of interest. This method is used to highlight factors that may explain the difference.

Generic Example #2: Matched Comparison

The goal of an audit could be to assess the success of implementing a monitoring program to track performance in education institutions, and to provide recommendations for improvement. The audit could rely primarily on the following:

- literature reviews
- consultations with managers
- a review of program criteria, policies, and procedures
- an analysis of performance measurement data

If, for the sake of argument, we assume that the monitoring program was experiencing a high failure rate (e.g., only half the educational divisions were successfully monitoring readiness to learn), then one of the audit objectives might have been to isolate some of the factors associated with success.

Under these circumstances, a matched comparison approach would be appropriate. A dozen divisions could have been selected: half that succeeded at conducting the monitoring and half that had not. Besides this main difference, the divisions could be matched based on other important criteria, such as rural/urban areas, socio-economic status, new immigration demographics, or any other major characteristic.

Interviews would focus on the main enabling factors and challenges of implementing the monitoring program. It is possible that the comparison of two such groups might yield insight into how to improve program compliance and lead to more robust recommendations.

Criterion-Based Case Selection

Criterion-based selection is appropriate whenever a characteristic is known to be correlated with material error, but the underlying causal relationship is unclear. Criterion-based selection is the selection of any case that exhibits the characteristic in question. The purpose of the examination is to better understand the cause of error and provide appropriate recommendations.

Continuum Sampling

While criterion-based selection results in a fairly homogenous sample of cases, the goal of continuum sampling is the opposite. With continuum sampling, the auditor purposely ensures the sample includes cases that fall along the entire spectrum of at least one variable of interest. The goal of this type of sample is to help illustrate the relationship between the variable of interest and error.

Group Characteristics Sampling

Group characteristics sampling focuses on discovering something meaningful about the nature of a group, or just a particular narrow stratum within a group. It is meant to demonstrate that a particular challenge (or success factor) exists regardless of other characteristics. It can also be used to demonstrate that a phenomenon exists by demonstrating its existence within a small randomly chosen sample.

Types of group characteristics sampling include:

- Maximum-variation sample: Selecting a wide range of cases to demonstrate a pattern that cuts across an entire population
- Homogenous sampling: Selecting only cases within a narrow range of variation due to their importance as an area for examination or due to a lack of pre-existing knowledge about a specific subgroup
- Typical cases: Using typical cases to help us understand challenges and difficulties experienced in most cases
- Key informants: Selecting individuals that, due to experience and training, have better-than-most understanding of a particular issue or problem
- Complete target population: Conducting an observation of every single case that meets a specific set of criteria
- Purposeful random sample: Using a small random sample to add credibility to findings by reducing selection bias (not to be confused with a generalizable sample)

Maximum-Variation Sample

With maximum-variation sampling, auditors select cases that represent the full variation that exists in the population: the cases are purposefully as different from each other as possible. The goal is simply to have at least one example of each type of case that exists in the population. This type of sample is useful because it vividly demonstrates any similarities across cases despite the obvious heterogeneity. This type of sampling is useful for examining large national or global programs.

Homogenous Sampling

As the name implies, a homogenous sample includes cases that show very little variation. Homogenous sampling is used to focus on a single, but important, subgroup within a population. This group might be new and emerging, and the nature of risk and controls associated with it may be unique. Alternatively, it may be a subgroup that represents a high-risk area and requires attention.

Typical Cases

Typical cases are best used for illustrative purposes. An examination of typical cases will help an audience understand situations they would otherwise not be knowledgeable of.

Key Informants

A key informants sampling method involves sampling from a population of individuals. Key informants, by virtue of education, training, and/or experience, have a better-than-most appreciation and understanding of the situation or program of interest. Key informant evidence is most powerful when independent expert opinions converge and describe problems and potential solutions in a similar manner.

Generic Example #3: Key Informants Case

In studying the use of transportation funds in mass transit projects, the auditors may ask transportation policy experts to provide their opinion on what would constitute an effective way of funding the right projects.

Source: GAO (2017)

Complete Target Population

A complete target population method is ideal in a situation where an unexpected and new incident occurs and either affects or is related to a limited number of cases. In a situation such as this, all affected or related cases are examined.

Purposeful Random Sample

The unbiased nature of a randomly selected sample can provide powerful and persuasive evidence. While the sample sizes are usually too small to be considered generalizable, these samples can demonstrate the existence of pervasive problems. This method is also useful in combination with other methods to increase the level of objectivity and validity of findings.

Generic Example #4: Purposeful Random Sample

In an audit of correctional facilities, the auditors may consider it necessary to collect evidence to demonstrate overcrowding. A small purposeful random sample of just a few correctional facilities would accomplish this goal. Even though such an observation would not be statistically generalizable, it would provide concrete examples of what actual conditions are like.

Mixed, Stratified, and Nested Sampling Methods

Mixed, stratified, and nested sampling methods, as the name implies, combine the use of two or more of the basic sampling methods. Several methods are mutually exclusive and difficult to combine, such as homogenous sampling and maximum variation sampling. Other sampling methods pair well together. Some examples include:

- Mixed matched and continuum sampling: Selecting two samples that differ significantly along a dimension of interest, but within each sample, a full continuum of other dimensions is represented
- Mixed homogenous and intensity sampling: Limiting the sample to a very narrow range of variation, and within that range, selecting only cases that demonstrate a phenomenon of interest with sufficient intensity
- Mixed criterion-based and purposeful random sampling: Beginning with the identification of all cases that meet the defined criterion of interest, then selecting cases from this group using a random selection method; this limits the bias introduced to the sample to only the intended bias

Other Purposeful Sampling Methods

Other purposeful sampling methods are discussed by Patton (2015); however, they are of limited value in an audit environment. This is either because the methods are intended for very specific conditions encountered in social science research or because interpretation of results depends on emerging theoretical constructs. Auditing is not an environment for theory building or testing. In order to substantiate findings, any assumptions made regarding how the sampling method is related to the interpretation of results needs to be transparent and clearly stated. Assumptions must also be reasonable and difficult to refute.

Back to Table of Contents

Appendix 4 – Sample Size Determination for Purposeful Sampling

Unlike in generalizable sampling, purposeful sampling is not supported by recognized mathematical criteria for predicting the minimum sample size that is necessary for reliable results. Obviously, for single-case sampling, sample size is not an issue. However, for other purposeful sampling methods, selecting criteria for how many cases need to be examined is important.

The most often cited criteria for sample size in qualitative research are those of saturation and redundancy. The idea is that the sample size should not be predetermined but should be allowed to be flexible and emergent. As more and more cases are examined, the level of insight and knowledge gained reaches a point of saturation and any further examination becomes redundant. However, this iterative strategy can create some challenges with respect to the budgeting and planning of audits.

Regardless of any decisions made regarding sample size, it is important to avoid judging the credibility or importance of findings based on the number of cases examined as is done with generalizable samples. Purposeful sampling is a trade-off of wider breadth for greater depth of understanding. As Patton (2015) states: "The validity, meaningfulness, and insights generated from qualitative inquiry have more to do with the information richness of the cases selected and the observational/analytical capacity of the researcher than with sample size."

As a rule of thumb, this Practice Guide suggests starting with a small sample and then increasing the sample if no uniformity of information is achieved after the initial sample is reviewed. There is no correct or universally recognized method for calculating a sample size for purposeful sampling. Some offices or firms may come up with an arbitrary number, but it is only that—arbitrary. There is no mathematical formula to justify it. What matters for auditors is to have a sample that could be defended by demonstrating that it is reasonable and sensible. They could use the rationale for selecting the various types of purposeful sampling approaches suggested in Appendix 3 as the basis to explain how the sample was selected.

Back to Table of Contents

Glossary

Administrative data
Data sets collected by government institutions or agencies for public administration purposes.

Attribute sampling
The selection of sampling units on the basis of a characteristic that can be determined by a binary choice, such as yes/no, error/no error, or on time/late. Attribute sampling is used to assess the proportion of a specified attribute in a sample and to extrapolate this proportion to the entire population being sampled.

Audit observation
The outcome of an objective evaluation of audit evidence against specific audit criteria.

Audit recommendation
A measurable statement for corrective action made by the auditors and addressed to the audited organization(s). Recommendations must address the causes of deficiencies identified in audit reports.

Census
A complete enumeration (100%) of a population. (Also known as a *certainty sample*.)

Confidence interval
The range of values in which we can be reasonably confident that the true population value resides, given the result from a sample. (Also known as the *margin of error.*)

Confidence level
The certainty with which the estimate value obtained from a sample lies within the confidence interval for that sample.

Controls
The policies and procedures designed, put in place, and operated within an organization to mitigate the risks that threaten the achievement of the organization's objectives.

Deviation
A departure from an expected value or standard.

Expected error rate
The estimate of the true rate of occurrence of the event (error) in the population being tested.

Generalizable sampling
The application of auditing procedures to a generalizable group of less than 100% of the items within a population of audit relevance such that all sampling units have a chance of selection in order to give the auditor a reasonable basis on which to draw conclusions about the entire population. (Also known as *probabilistic sampling*, *statistical sampling,* or *representative sampling.*)

Heterogeneity
The level of differentiation among sampling units within a population.

Homogeneity
The level of uniformity among sampling units within a population.

Materiality
The magnitude of an omission or misstatement of accounting information that may change or influence the economic decisions of users of financial statements. In performance audits, the equivalent concept is "importance" or "relevance" in order to include quantitative dimensions beyond strict monetary value as well as qualitative considerations.

Mean
The sum of all the values in a set of observations (either a census or sample) divided by the number of observations, indicating the typical value for a set of observations. (Also known as *average*.)

Median
A value separating the higher half from the lower half of a data sample, population, or probability distribution.

Non-generalizable sampling
A form of sampling in which auditors rely on their own judgment when choosing units of a population for examination. (Also known as *judgmental sampling*, *non-probabilistic sampling*, or *purposeful sampling.*)

Observed error rate
Actual value of the rate of occurrence of an event (error) in the population being tested.

Performance audit
An independent, objective, and reliable examination of whether government programs, activities, or organizations are performing according to the principles of economy, efficiency, and effectiveness.

Population
The total number of items from which a sample is drawn.

Precision
The accuracy of a sample estimate compared with the actual value of the population.

Purposeful sampling
A form of non-generalizable sampling based on the introduction of an explicit bias in the sample's selection, with the specific intent of isolating and selecting information-rich cases that will be particularly useful to gain insight and understanding of the topic.

Random sample
A sample in which every item in the population has an equal chance of being selected.

Random seed
A random seed is a number (or vector) used to initialize a pseudorandom number generator, i.e., an algorithm for generating a sequence of numbers whose properties approximate the properties of sequences of random numbers.

Risk

An event or action that may adversely affect an organization's ability to achieve its objectives. Assessing risk involves considering the probability (or likelihood) of the event occurring and the potential impact of that event.

Sample

A selection of items from a population.

Sample size

The number of items selected in a population.

Sampling frame

A list of the items or people forming a population from which a sample is taken.

Sampling unit

What is being sampled (files, transactions, projects, individuals, regions, etc.). Each unit is regarded as individual and indivisible when the selection is made.

Skewness

Distortion or asymmetry in a symmetrical bell curve, or normal distribution, in a set of data. If the curve is shifted to the left or to the right, it is said to be skewed. Skewness can be quantified as a representation of the extent to which a given distribution varies from a normal distribution.

Strata

Two or more mutually exclusive subgroups of a population defined in such a way that each sampling unit can belong to only one subgroup or stratum.

Stratification

The process of segregating a population into homogenous subpopulations explicitly defined so that each sampling unit can belong to only one subpopulation.

Substantive test

A procedure performed to provide audit evidence as to the completeness, accuracy, and validity of information in the accounting records and financial statements, or for compliance with standards and tests of non-automated controls.

Variable sampling

A form of sampling in which the selected sampling items are measured or evaluated in terms of a continuous variable of interest (such as dollars, distance, or time). Variable sampling is used to estimate the value of the variable within the whole population.

Variance

A measure of the degree of dispersion among a set of continuous data. The square root of variance is called the *standard deviation*.

Back to Table of Contents

Bibliography

Buskirk, T. (2016). Target Population and Sampling Frame in Survey Sampling, *The Analysis Factor*, available at: https://www.theanalysisfactor.com/target-population-sampling-frame/

Colbert, J. L. (1991). Statistical or Non-statistical Sampling: Which Approach Is Best? The Journal of Applied Business Research, 7(2), 117–120. Available at: Statistical or Non-statistical Sampling: Which Approach Is Best?

Cook, T. D. & Campbell, D. T. (1979). *Quasi-Experimentation: Design and Analysis Issues for Field Settings*. Belmont, CA: Wadsworth Publishing.

European Court of Auditors (2020, January). Big Data and Digital Audit. *ECA Journal*, available at: https://www.eca.europa.eu/en/Pages/NewsItem.aspx?nid=13397

Guy, D. M., Carmichael, D. R., & Whittington, O. R. (1998). *Practitioner's Guide to Audit Sampling.* New York: John Wiley & Sons, Inc.

Jokovich, G. (2013). Statistical Sampling in Auditing. *International Journal of Accounting and Financial Management*, 16, 892–898.

Leslie, D. A., Teitlebaum, A. D., & Anderson, R. J. (1979). *Dollar-Unit Sampling. A Practical Guide for Auditors.* Toronto: Copp Clark Pitman.

National Audit Office of the United Kingdom (2001). *A Practical Guide to Sampling*, available at: https://www.nao.org.uk/wp-content/uploads/2001/06/SamplingGuide.pdf

Office of the Auditor General of Canada (2019), *Direct Audit Engagement Manual*, available at: https://www.oag-bvg.gc.ca/internet/methodology/performance-audit/manual/index.shtm

Palinkas, L. A., Horwitz, S. M., Green, C. A., Wisdom, J. P., Duan, N., & Hoagwood, K. (2015). Purposeful Sampling for Qualitative Data Collection and Analysis in Mixed Method Implementation Research. *Administration and Policy in Mental Health and Mental Health Services,* 42(5), 533–544.

Patton, M. Q. (2015). *Qualitative Research and Evaluation Methods* (4th ed.). Thousand Oaks, CA: Sage.

Public Company Accounting Oversight Board (n.d). AS 2315: Audit Sampling, available at https://pcaobus.org/Standards/Auditing/Pages/AS2315.aspx

The Internal Auditor Professional Services Limited (n.d.), Sampling for Effective Internal Auditing, available at https://web.archive.org/web/20190713121518/http:/www.isaca.org/chapters10/Lusaka/NewsandAnnouncements/Documents/SAMPLING-FOR-INTERNAL-AUDITORS%5b1%5d.pdf

U. S. Government Accountability Office (1992). Using Statistical Sampling, available at: https://www.gao.gov/products/PEMD-10.1.6

U. S. Government Accountability Office (2017). Selecting a Sample of Nongeneralizable Cases, available at: https://www.ignet.gov/sites/default/files/files/16_Selecting_a_Sample_of_Nongeneralizable_Cases.pdf