MAPS | Methodology for Assessing Procurement Systems

Guide on quantitative indicators in MAPS

2026

The MAPS Initiative is financed by

AFRICAN DEVELOPMENT BANK GROUP
GROUPE DE LA BANQUE AFRICAINE DE DÉVELOPPEMENT

European Investment Bank

IDB

Ministry of Foreign Affairs of the Netherlands

Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

Swiss Confederation

AFD GROUPE

European Commission

IsDB البنك الإسلامي للتنمية
Islamic Development Bank

Contents

Acknowledgements 3

I. Purpose and use of this guide 4

II. Planning the data required for quantitative indicators in MAPS 5

i. Planning the data collection 5

ii. From planning to evidence collection 10

III. Systems-based quantitative indicators 11

i. Establishing source and indicator coverage 11

ii. Choosing the processing route 11

IV. Sample-based quantitative indicators 12

i. Defining the sampling frame 12

ii. Choosing the sampling design 12

iii. Determining the sample size 14

iv. Selecting and accessing the cases 14

v. Reviewing the cases consistently 15

vi. Estimating the quantitative indicators 15

V. Survey-based quantitative indicators 16

i. General survey principles 16

ii. Supplier survey – Defining the supplier population and sampling frame 16

iii. Supplier survey – Selecting respondents and determining the number of invitations 17

iv. Supplier survey – Designing the questionnaire 17

v. Supplier survey – Conducting the survey 18

vi. Supplier survey – Analysing the survey results 18

vii. Adapting the survey approach to other respondent populations 19

VI. Using the results of quantitative indicators 19

i. Presenting quantitative results 20

ii. Using quantitative results in the qualitative analysis 20

Glossary 23

Acknowledgements

This guide is the result of work undertaken by Carlos Mauricio Ortiz Niño, international consultant and public procurement expert with specialised expertise in quantitative data and statistical analysis. Drawing on both his professional experience and consultations with lead institutions and assessors involved in MAPS assessments, Mr Ortiz developed the technical content that forms the basis of this guidance.

The material prepared by Mr Ortiz was subsequently reviewed, reorganised and edited by the MAPS Secretariat to ensure consistency with the MAPS framework and alignment with existing MAPS methodological and guidance documents. The resulting publication seeks to preserve the technical rigour of the original work while presenting it in a format that integrates seamlessly with the broader suite of MAPS guidance materials.

In addition to acknowledging the valuable contribution of Mr Ortiz, the MAPS Secretariat wishes to express its sincere gratitude to the representatives of lead institutions and experienced MAPS assessors who generously shared their knowledge, practical insights and lessons learned from the implementation of MAPS assessments. Their contributions helped ensure that this guide reflects both sound technical practice and the realities of conducting assessments in diverse contexts.

This includes Pierre Morin, Rajesh Kumar Shakya, Devesh Mishra, and Amilcar Bilale.

I. Purpose and use of this guide

Quantitative indicators are an important component of the MAPS framework. They provide a means for substantiating the qualitative analysis and strengthening the overall analytical basis of an assessment. Both MAPS MAIN and each of the four supplementary modules contain quantitative indicators linked to specific assessment criteria. Where relevant data are available, these indicators provide important evidence for the assessment of those criteria.

This guide provides practical guidance on planning, obtaining, processing, analysing and using quantitative indicators in MAPS. Owing to the specialised subject matter, the guide is more technical in nature than most other MAPS guidance documents and uses a number of statistical and analytical concepts that are not used elsewhere in the MAPS framework. For ease of reference, a glossary of key terms is provided at the end of the guide.

The guidance applies to MAPS MAIN and, as relevant, to supplementary modules. Most examples in this guide are drawn from MAPS MAIN because it provides the broadest recurring set of quantitative indicators, but the same planning principles can be applied to module-specific indicators and respondent populations.

The guide distinguishes between MAPS requirements and practical implementation choices. Unless a numerical value, procedure or threshold described in the guide is explicitly identified as a MAPS requirement, it should be understood as a practical reference that can be adapted to the assessment context, including situations in which data availability, access, time or resources are constrained. The guide does not introduce new methodological requirements or modify requirements established by the MAPS framework and associated guidance. Rather, the approaches described here generally reflect good practice and, where relevant, illustrate how quantitative indicators might ideally be approached when suitable data sources, sufficient resources, appropriate technical expertise and adequate time are available.

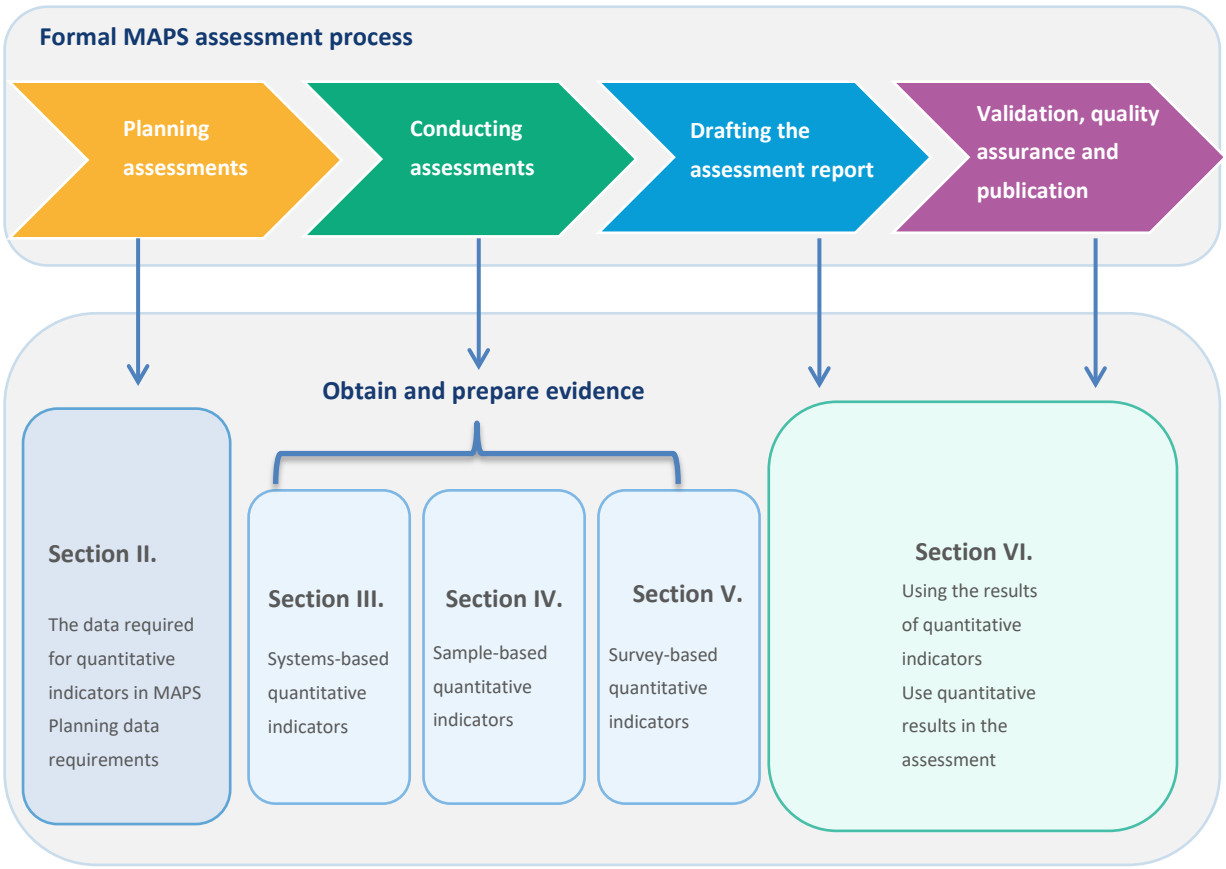
In this respect, the guide often describes a best-case approach to the use of quantitative indicators. In practice, assessments may face limitations that make some of the suggested approaches difficult or impossible to implement fully. This does not, in itself, prevent an assessment from being conducted in accordance with the MAPS methodology, nor should the absence of ideal conditions be interpreted as preventing the use of quantitative indicators where they can still contribute useful evidence.

Accordingly, the guide should be understood as a practical reference rather than a prescriptive manual. Assessment teams are encouraged to adapt the approaches presented here to the realities of their assessment context, while remaining transparent about any limitations, assumptions, methodological choices and data constraints encountered during the assessment process.

Quantitative work should be planned as an integral part of the assessment rather than as a parallel exercise. Planning involves identifying the applicable quantitative indicators, potential data sources and feasible approaches to data collection and analysis and should inform the concept note. During the assessment, evidence is obtained and prepared for analysis, and quantitative results are subsequently interpreted and linked back to the relevant assessment criteria alongside qualitative findings and gap analysis.

How the guide sections fit into the formal MAPS assessment

Integration of the Guide structure into the formal MAPS workflow



II. Planning the data required for quantitative indicators in MAPS

Quantitative data collection should be planned for the assessment as a whole. Data needed for quantitative indicators located in different parts of the MAPS methodology often come from the same systems, administrative records, procurement files or respondent populations. Planning these requirements together reduces duplication and helps maintain consistent definitions, periods and populations across the assessment.

The assessment team should therefore translate the applicable quantitative indicators into a source-based plan for the assessment. Data requests can be consolidated by source, variables that require procurement-file review can be brought together in one case-review instrument, and questions addressed to the same respondent population can be combined in one survey instrument. This source-based planning should be reflected in the concept note, while remaining open to revision as feasibility becomes clearer.

i. Planning the data collection

At the beginning of the assessment, the assessment team may establish a quantitative indicator inventory covering the quantitative indicators in the MAPS MAIN and/or supplementary modules as applicable. The inventory is an optional working tool for planning and following quantitative work throughout the assessment; it is not a new MAPS reporting requirement and does not replace the mandatory MAPS indicator matrix.

The inventory is developed in three stages (see table 1 for illustration):

1. The quantitative indicator overview annex in MAPS MAIN or the relevant module (Annex 2) is used as the basis for the inventory. The annex lists all quantitative indicators relevant for the assessment, along with the suggested data source. For ease of reference, each quantitative indicator can be given an identification code corresponding to its associated sub-indicator and assessment criterion
2. At the concept note stage, the team completes fields needed to define how each quantitative indicator is expected to be addressed, including measure clarifications (if needed), expected applicability, planned country source, planned period and expected coverage or constraints. These planning fields should be closed at the end of the planning stage so that the original approach remains traceable.
3. As the assessment is conducted, implementation fields are filled out, keeping for the record the planning fields. These implementation fields record the source ultimately used, actual period used, the result and any material limitations or comments. The inventory therefore provides a traceable record from the MAPS requirement, through the planned approach, to the evidence ultimately obtained.

This description of the inventory as well as the illustrative example below is provided as a pragmatic suggestion for how to build a working tool for planning and following quantitative work throughout the assessment. I should be adapted to the context of the specific assessment.

Table 1: Example – Quantitative indicator inventory

MAPS identification			*Planning baseline*					*Actual implementation*		*Results*	
ID	**Indicator text**	**MAPS suggested source**	**Expected applicability**	**Measure / operational clarification**	**Planned country source**	**Planned period**	**Expected coverage**	**Source used**	**Period used**	**Result**	**Limitations/ comments**
QI-4b-01	Invoices paid on time (in %).	PFM systems	Expected to apply	On time means payment within the applicable legal deadline; denominator: invoices due for payment during 2025.	Treasury payment database	2025	Invoices processed through the central treasury.	Treasury payment database	2025	Estimated: 80% (9,600 of 12,000 invoices).	Invoices paid outside the central treasury are excluded.
QI-7b-03	Bids submitted online by micro, small and medium-sized enterprises (in %).	e-Procurement system	Expected N/A							Not Applicable. Online bid submission was not available during the assessment period.	
QI-9b-01	Average time to procure goods, works and services.	Sample of procurement cases	Expected to apply		e-Procurement system, subject to validation; case review if required	Assessment period	Goods, works and services in the e-Procurement system.	Procurement files (stratified sample of 150 cases)	2025	Estimated: goods 42 days; works 85 days; services 51 days.	Incomplete system dates required a sample-based estimate instead of full-system data.

MAPS identification			*Planning baseline*					*Actual implementation*		*Results*	
ID	**Indicator text**	**MAPS suggested source**	**Expected applicability**	**Measure / operational clarification**	**Planned country source**	**Planned period**	**Expected coverage**	**Source used**	**Period used**	**Result**	**Limitations/ comments**
QI-10b-01	Number of registered suppliers as a share of total number of suppliers in the country (in %).	e-Procurement system / supplier database	Expected to apply	Denominator: active enterprises identified as potential suppliers in the national business register.	Supplier register + business or tax register	2025	National supplier register and comparable active enterprises in the business register.	Supplier register and national business register	2025	Estimated: 15% (18,000 registered / 120,000 potential suppliers).	Business register measures potential rather than confirmed suppliers; denominator is approximate.
QI-13c-01	Share of suppliers that perceive the challenge and appeals system as trustworthy (in % of responses).	Survey	Expected to apply		Supplier survey	Survey period	Registered suppliers across regions and firm-size groups.	Supplier survey (472 valid responses)	2025 survey	Estimated (): Very trustworthy 12%; somewhat trustworthy 49%; not very trustworthy 27%; not at all trustworthy 12%.	40% response rate; possible non-response bias. Unregistered suppliers not covered.
QI-13c-03	Outcome of appeals: dismissed; in favour of procuring entity; in favour of applicant (in %).	Appeals body	To be confirmed		Appeals body records	2025	All appeals decided by the national review body during 2025.	Appeals body administrative records	2025	Not estimated. Reliable percentages by outcome could not be calculated.	Decision outcomes were recorded inconsistently; some cases lacked usable classification.

The questions below provide a practical sequence for completing the planning fields of the inventory. Entries may be refined during planning as source feasibility and expected coverage become clearer.

1. What exactly needs to be measured?

Before looking for a source, the assessment team should be clear about what the quantitative indicator is asking to measure. Depending on the indicator, this may require clarifying the phenomenon or condition, target population, unit of observation, numerator and denominator, reference period, and any relevant definition. The objective is not to redefine the MAPS indicator, but to understand exactly what evidence is needed to calculate it.

The distinction between observable measures and perception or experience measures is particularly important at the planning stage because it affects the appropriate source, respondent population, instrument and the possibility of using another source as an equivalent measure. For example, Invoices paid on time (in %) measures observed payment performance, while the proportion of suppliers who consider payments to be timely measures supplier perception or experience. Both may be useful, but they answer different questions.

The distinction also matters when results point in different directions. A high share of invoices paid within the applicable period and low supplier satisfaction with payment timeliness are not necessarily contradictory. Where a perception measure is used because an observable measure cannot be obtained, it should be identified as a proxy rather than presented as the quantitative indicator itself.

The inventory should therefore establish the required measure before assigning a country source. The Measure / operational clarification field should be completed only where clarification is needed to make the indicator operational; it should not restate every MAPS indicator.

2. Is the quantitative indicator applicable?

Once the required measure has been clarified, the assessment team should determine whether the quantitative indicator is applicable within the assessment scope and reference period. Note that the applicability of a quantitative indicator within a specific assessment scope and reference period is distinct from the applicability of assessment criteria under the MAPS methodology. The latter is defined by the methodology and applies only to specific assessment criteria explicitly identified as potentially non-applicable. The applicability of a quantitative indicator, as used in this guide, is a separate operational concept for planning and conducting data collection and analysis.

A quantitative indicator is Not Applicable (N/A) where there is no legitimate population, event, condition or process to which the indicator applies within that scope and period. Where such a population, event, condition or process exists, the indicator is applicable regardless of whether the evidence required to estimate it is ultimately available.

During planning, applicability may be recorded as expected to apply**,** expected N/A, or to be confirmed. For example, the indicator measuring the share of online bids submitted by micro, small and medium-sized enterprises (MSME) is calculated over the total number of bids submitted online. If online bids were submitted during the reference period but none was submitted by an MSME, the indicator is applicable and the result is 0%. If no bids were submitted online during the period, there is no relevant population over which the share can be calculated and the indicator is therefore N/A. If online bids were submitted but the information needed to identify whether bidders are MSMEs is unavailable, the indicator remains applicable but cannot be estimated.

3. Which source or sources can provide the required evidence?

Unless an indicator is clearly N/A, the assessment team should identify the potential source or sources that can provide the required evidence in the country. A source is where the evidence resides; the collection or processing route is how the evidence will later be obtained or produced. Planning should identify the most plausible source and make a preliminary assessment of feasibility, which may include early contact with the responsible institution, without confusing this step with full data collection.

Some indicators require more than one source to be calculated. Number of registered suppliers as a share of total number of suppliers in the country (in %) may, for example, require the supplier register for the numerator and a tax or business register for the denominator. The inventory should identify such combined source requirements from the outset.

The MAPS suggested source is the starting point for this assessment, but it should be validated against the evidence available in the country. For example, even though MAPS identifies a sample of procurement cases as the source for Average time to procure goods, works and services, a sufficiently complete and reliable e-Procurement source containing the dates and procurement-method information needed to calculate the same measure may provide a more appropriate country source.

After identifying a potential source, the assessment team should consider whether it is likely to provide evidence with sufficient availability, access, completeness, coverage and reliability for the required measure. Where a material risk of non-estimation exists, a complementary or alternative route should be identified where feasible. An alternative source is not the same as a proxy: an alternative should preserve the required measure, while a proxy measures something different and must be reported as such. An alternative route is not required for every quantitative indicator.

A survey may serve as an alternative source only if it can measure the same required measure required by the quantitative indicator. If the original indicator concerns an observable condition or outcome, but a survey can capture only stakeholder perception or experience, the survey does not provide an equivalent estimate. Its result should instead be treated as a proxy or complementary source of evidence, while the original quantitative indicator remains Not estimated.

4. What period should be covered?

Where MAPS specifies a reference period, that period takes precedence. Where no period is prescribed and comparable data are available, a multi-year series can add value by showing whether performance is persistent, improving, deteriorating or associated with an exceptional period. Three to five years may be a useful practical reference when feasible, but comparability is more important than length.

The period should nevertheless be comparable. A shorter series may be more informative where, for example, a new e-Procurement system was introduced during the period, the relevant variable only began to be recorded recently, or a reform materially changed the definition of the measure.

Exceptional observations also need to be interpreted rather than automatically removed. A major regulatory reform, an emergency, or a significant fiscal or social event may legitimately produce an unusual value in a particular year. When such an observation reflects actual conditions, it should normally remain in the analysis and the event should be explained in a note accompanying the indicator. Potential breaks in the series or exceptional years should therefore be flagged while the inventory is being prepared.

ii. From planning to evidence collection

Once the planning fields of the inventory have been completed, quantitative work can be organised by evidence source rather than by individual MAPS indicators. Indicators relying on the same information system or institution can be consolidated into a common data request; information to be obtained from procurement files can be incorporated into a single case-review instrument; and questions addressed to the same respondent population can be combined into a common survey instrument.

The planning fields should remain unchanged as the assessment proceeds. Actual sources, periods, coverage, status and results are recorded separately in the implementation/final fields, allowing the assessment team to trace where implementation differed from the original plan.

Sections III to V provide guidance on obtaining and preparing these different forms of quantitative evidence.

III. Systems-based quantitative indicators

Where the quantitative indicator inventory identifies an information system, administrative database or register as a potential source, the assessment team should first verify that it covers the relevant population and period and contains the information needed to calculate the indicator.

The team should then decide whether to process the data directly or request the required statistics from the responsible institution. In either case, the evidence should be checked for material issues, and the source, coverage, processing route and any limitations should be documented.

i. Establishing source and indicator coverage

Information systems do not necessarily cover all procurement activity within the scope of a MAPS assessment. Coverage may be limited because some entities or levels of government do not use the system, certain procurement regimes or methods are excluded, transactions below or above particular thresholds are not recorded, or reporting requirements are not fully observed. The assessment team should therefore establish what part of the relevant population and procurement activity is represented by each system-based source.

Where procurement expenditure is a relevant dimension of coverage, comparing the value recorded in the information system with total procurement expenditure can help assess how much of the procurement activity within the assessment scope is represented. If an official estimate of total procurement expenditure is not available, budget execution data may provide a practical order-of-magnitude denominator. The purpose is not to create a new official procurement expenditure statistic, but to obtain a reasonable basis for assessing the coverage of the source.

Coverage of a particular quantitative indicator may be narrower than the overall coverage of the source. For example, a system may contain most procurement procedures but lack a variable required to calculate a specific indicator for some records.

Limited coverage does not make a source unusable, but it limits the population to which the resulting indicator can be generalised. Where the uncovered portion is material, the assessment team should consider whether another administrative source, procurement-case review or another equivalent source of evidence is needed to complement the system-based evidence.

ii. Choosing the processing route

Once the coverage of the source is understood, the assessment team should decide whether to obtain and process microdata directly or ask the data-owning institution to produce the required indicators or statistics. In this Guide, microdata means individual or record-level observations, such as procurement procedures, contracts, bids, invoices or suppliers, rather than aggregate statistics.

Direct processing is appropriate where record-level data can be obtained and the assessment team has the capacity to work with them. The request should specify the relevant period and population, unit of observation, required variables and, where necessary, the identifiers needed to link datasets, such as a common contract or supplier identifier. Before estimating the indicators, the team should perform proportionate checks for issues that could materially affect the result, including missing values, duplicates, implausible dates or amounts, inconsistent coding, discontinuities or changes in coverage or recording practice. Material outliers should be examined rather than automatically excluded.

Processing by the data-owning institution may be preferable where access to microdata is restricted, the source is complex or the institution is better placed to interpret and process the records. In that case, the request should specify the indicator or statistic, definition, period, relevant population or unit of analysis, numerator and denominator where useful, and only those disaggregations that are analytically relevant.

When results are produced by the data-owning institution, the assessment team should perform limited ex post validation: confirm that the requested period, population and definitions were applied, assess plausibility and coherence with other available information, and seek clarification for material unexplained differences. This does not require reproducing the institution’s calculations. The processing route and institutional

provenance should be recorded in the inventory and, where relevant, in the Concept note, indicator matrix and assessment report.

Regardless of the processing route, system-based quantitative indicators should be interpreted within the coverage supported by the source, and any material limitations should be carried forward when the results are presented and used in the assessment.

IV. Sample-based quantitative indicators

This section applies where the quantitative indicator inventory identifies procurement-file review as the appropriate route for obtaining the required evidence.

Where procurement-file review is needed, all relevant quantitative indicators should be addressed through a single procurement-case review and sampling exercise rather than through separate samples for each indicator. In MAPS, procurement cases are the procurement transactions or proceedings, together with the corresponding files or records selected for review. A single case review can therefore collect the variables needed to estimate several quantitative indicators across the assessment.

i. Defining the sampling frame

Before identifying the sampling frame, the assessment team should define the population of procurement cases that the sample is intended to represent, taking into account the assessment scope, reference period and quantitative indicators to be estimated. For example, if the assessment covers central-government procurement during 2023-2025, the sample should represent procurement cases from that population and period rather than all procurement recorded in the country.

The next step is to identify the sampling frame: the list from which procurement cases will actually be selected. Where a sufficiently complete case-level list is available, for example from an e-Procurement system or procurement register, cases can be selected directly from that list. The team should verify that the frame covers the relevant procurement activity and contains the information needed for selection, particularly the procurement method.

If no sufficiently complete case-level list exists, a two-stage approach may be used. First, procuring entities are selected from an appropriate list of entities. Each selected entity then provides the complete list of eligible procurement cases for the relevant period, and the procurement cases are selected from those lists in the second stage. Selected entities should not choose which cases are reviewed.

The sampling design and sample size should be determined only after the sampling frame has been established. Increasing the number of cases can improve the precision of the results, but it cannot compensate for procurement activity that was excluded from the frame and therefore had no chance of being selected.

ii. Choosing the sampling design

Once an appropriate sampling frame has been identified, the sampling design should reflect three related dimensions: the selection structure, whether the frame is stratified, and the probabilities used to select units. Procurement cases may be selected directly from a case-level frame or through a multistage structure where entities must be selected first. Within either structure, the relevant frame may be sampled as a whole or divided into strata, and units may be selected with equal or unequal probabilities.

Direct selection of procurement cases

Where a sufficiently complete case-level frame is available, procurement cases should be selected using a probability-based method.

In MAPS assessments, stratified random sampling by procurement method will often be the preferred design. Procurement method is directly relevant to several quantitative indicators and commonly reflects differences in procedure, complexity, timelines and, in many systems, value thresholds. For example, MAPS requires *Average time to procure goods, works and services* to be calculated for each procurement method used. Stratification helps ensure that the main procurement methods contain enough observations to support such estimates.

The assessment team should identify the procurement-method strata that are analytically relevant. Small or exceptional methods that do not require separate analysis may be grouped transparently in an "Other" stratum. Additional stratification variables, such as entity type, year, contract type or value range, should be introduced only where they address an analytical need not adequately captured by procurement method. Excessive cross-stratification should be avoided.

Simple random sampling is also a valid option, particularly where the procurement population is relatively homogeneous. However, because cases are selected from the full frame without ensuring representation of particular procurement methods, the resulting sample may contain too few observations from less frequent methods to support reliable estimates at that level.

A separate design decision concerns the probability assigned to individual units during selection. Cases may be selected with equal probabilities or with unequal probabilities where there is a methodological reason to give some units a higher chance of selection. Probability proportional to size (PPS)[1] is one such approach: the probability of selection is linked to a defined measure of size, such as contract value.

PPS can be applied within either an unstratified or a stratified design. For example, procurement cases may first be stratified by procurement method and then selected within each stratum with probabilities proportional to contract value.

PPS does not mean selecting only large cases or entities. All eligible units may remain selectable, but units with a larger value of the chosen size measure have a higher, known probability of selection. Because these probabilities are unequal, they must subsequently be taken into account when the quantitative indicators are estimated.

.

Two-stage selection

Where no sufficiently complete case-level sampling frame is available, but a reliable list of procuring entities exists, procurement cases may be selected through a two-stage design.

In the first stage, procuring entities are selected probabilistically from the entity-level frame. Depending on the structure of the procurement system, entities may be stratified by characteristics such as entity type or procurement volume. PPS may also be used where entities with larger procurement volumes should have a higher probability of selection. Where such distinctions are not analytically important, simple random sampling may be sufficient.

Each selected entity should then provide the complete list of eligible procurement cases for the relevant period and assessment scope. These lists form the second-stage sampling frames. Procurement cases are then selected probabilistically from each list, normally using stratification by procurement method where this is relevant to the quantitative indicators being estimated. Selected entities should not choose which cases are reviewed.

The two stages do not need to use the same selection method. Two-stage sampling describes the structure of the selection process, while stratification, simple random sampling and PPS are methods that may be applied within either stage. For each selected procurement case, the probability of inclusion reflects the selection probabilities at both stages and should be retained for subsequent estimation.

[1] The OECD Glossary for Statistical Terms defines PPS as "*A sampling procedure whereby each unit in the universe has a probability of selection proportional to the size of some known relevant variable. In the case of establishments, size is usually defined in terms of employment or output.*"

Sampling frame and corresponding design options

Available sampling frame	Selection structure	Practical starting approach
Complete case-level frame with procurement method available	Direct selection of procurement cases	Stratified random sampling by procurement method
Complete case-level frame; stratification adds little analytical value	Direct selection of procurement cases	Simple random sampling may be sufficient
Complete case-level frame where inclusion probabilities should vary with a measure of size	Direct selection of procurement cases	PPS, with or without stratification
Entity-level frame only	Two-stage selection: entities, then procurement cases	Stratification and/or PPS at first stage; probabilistic case selection at second stage

iii. Determining the sample size

The sample size should be determined only after the sampling frame, relevant procurement-method strata and selection design have been defined. As a practical reference for MAPS assessments, the sample should normally include at least 30 randomly selected procurement cases in each analytically relevant procurement-method stratum, or all eligible cases where a stratum contains fewer than 30.

This commonly used rule of thumb provides a consistent and workable analytical base for each relevant procurement method without requiring a separate statistical sample-size exercise. . The total sample therefore depends on the number of relevant strata: two strata would normally imply about 60 cases, three about 90, while an assessment focused on a single procurement method may require only about 30 cases. A total sample below 100 cases can therefore be entirely appropriate. However, a larger sample would generally be expected where data availability allows for it.

The 30-case reference is a practical planning benchmark rather than a minimum established in the MAPS methodology. The strength of the resulting evidence depends not only on the number of cases reviewed, but also on whether the sampling frame adequately covers the procurement activity of interest and whether cases are selected probabilistically.

A larger sample improves the amount of evidence available, but it cannot correct a problem in who had the opportunity to be selected. For example, reviewing 100 rather than 30 cases from a frame that excludes an important part of procurement activity still provides no evidence about the excluded cases. Similarly, increasing the number of cases selected discretionarily does not make the sample statistically representative. These limitations must therefore be addressed through the sampling frame and selection design, not simply by increasing the sample size.

iv. Selecting and accessing the cases

Once the sampling design and sample size have been defined, cases should be selected according to the probability-based procedure established in the sampling plan. Cases should not be chosen because they appear representative, interesting or easy to access.

A reserve set should be selected at the same time and using the same selection procedure. Reserve cases can replace selected cases that later prove ineligible or whose files cannot be obtained. Where the sample is stratified, replacements should come from the same stratum. Selecting reserves in advance avoids discretionary replacement once the review has started.

After selection, the assessment team should obtain the corresponding procurement files. Where cases are selected directly from an e-Procurement system or another case-level frame and supporting documents are available electronically, the selected files can normally be retrieved and assigned directly to reviewers.

In a two-stage design, access requires an additional step. Selected procuring entities first provide the complete list of eligible cases, from which the second-stage sample is drawn, and then provide the files corresponding to the selected cases. Where records are not centrally available in electronic form, access may require a combination of electronic requests and on-site review.

v. Reviewing the cases consistently

The reliability of sample-based indicators depends on reviewing all selected cases consistently. All reviewers should use the same structured instrument, definitions and decision rules. Where several reviewers participate, these should be calibrated at the start of the exercise.

The review instrument should consolidate the variables needed for the sample-based quantitative indicators and use standardised fields, response options and controls where useful. A spreadsheet or equivalent structured data-entry tool will usually be sufficient. Each procurement case should have a unique identifier.

Where interpretation is required, the review instrument should capture observable variables rather than ask reviewers to calculate or judge the indicator directly. For example, to estimate *Time overruns (in %; and average delay in days)*, reviewers should record only the initial contractual duration, the final actual duration and the corresponding units of time. Once all cases have been reviewed, these variables are standardised centrally to a common unit, normally days. For each contract, the difference between the final and initial duration is then calculated. This produces the absolute overrun in days and the relative overrun as a percentage of the initial duration.

Quality control should begin while reviews are under way so that ambiguous fields or decision rules can be corrected early. A subset of completed reviews should then be validated against the source documentation. Around 10% of completed reviews can serve as a practical starting reference. Where possible, validation should be performed by another team member; where only one reviewer is available, a structured second-pass review can be used. If material discrepancies are identified, the relevant rules should be clarified, affected reviews corrected and validation expanded as needed.

vi. Estimating the quantitative indicators

The estimation method should reflect the sampling design. Results reported separately for a procurement-method stratum can be calculated directly from the randomly selected cases in that stratum. When results from different strata are combined into an aggregate estimate, however, their different population sizes and selection probabilities must be taken into account.

For the basic stratified design used in this Guide, the sampling weight for each stratum can be calculated as the number of eligible cases divided by the number of sampled cases in that stratum (N/n). For example, if 300 open competitive procedures and 60 direct procedures are represented by 30 sampled cases each, the respective weights are 10 and 2. Without weighting, both strata would contribute equally to the aggregate result even though they represent very different shares of the population.

Sampling weights reflect the probability of selection and should not be confused with contract or procurement value. The same sample can therefore support indicators based on case counts as well as indicators involving monetary values.

When procurement cases do not all have the same probability of being selected, this difference must be reflected in the estimation so that cases with a higher chance of selection do not have disproportionate influence on the population result. PPS is one situation in which this occurs, because larger cases or entities are deliberately given a higher probability of selection.

The same principle applies in a two-stage design. A procurement case can enter the sample only if its procuring entity is selected first and the case is then selected within that entity. Its overall probability of selection therefore depends on both stages. If PPS is used at either stage, those unequal probabilities must also be incorporated. The sampling records should retain the information needed to calculate the overall probability of selection for each case. Population-level estimates can then apply the corresponding sampling weights.

Where probability sampling is not feasible, a purposive or otherwise practicable set of procurement cases may still be reviewed to generate useful qualitative evidence. Such a review can help identify recurring issues, implementation problems or illustrative examples that enrich the assessment.

The findings should not be presented as statistically representative of the wider procurement population or used to estimate population-level quantitative indicators. They should instead be reported as qualitative evidence derived from the cases reviewed, with appropriate caution about their broader applicability.

V. Survey-based quantitative indicators

i. General survey principles

The quantitative indicator inventory identifies which quantitative indicators require survey evidence and the respondent population associated with each indicator. Questions addressed to the same respondent population should normally be consolidated into one survey instrument rather than organised separately by pillar.

The main respondent populations considered in this Guide are:

- Suppliers/firms.
- Public procurement buyers/procurement workforce.
- Other, including module-specific populations (for example, the PROF module may require evidence from training or education providers)

Survey evidence should remain distinct from qualitative consultation methods. Interviews and focus groups can provide useful contextual or explanatory evidence, but they do not replace a survey where the quantitative indicator requires a population-level quantitative measure. Where the MAPS Methodology identifies interviews alongside surveys as a possible source, that evidence should be used according to the nature of the indicator rather than treated as equivalent by default.

ii. Supplier survey – Defining the supplier population and sampling frame

The target population of a supplier survey extends beyond firms already participating in public procurement. This is important because perceptions of the procurement system may vary substantially according to the level of interaction with it. Firms that have never registered, suppliers that registered but never submitted a bid, bidders that have not obtained a contract, and government contractors may face different barriers and have different experiences.

In practice, however, a complete sampling frame of all potential suppliers will rarely be available. The assessment team should therefore distinguish between the broad target population and the population for which a formal sampling frame can actually be constructed.

The formal sampling frame will normally be based on the supplier registry, e-Procurement records or another administrative source that identifies registered suppliers. By definition, this frame does not include potential suppliers that are not registered in the procurement system, even though they form part of the broader target population of the survey.

Within the supplier registry (or the equivalent list used as the formal sampling frame) firms may have different levels of participation in public procurement:

- **Registered only:** suppliers registered in the system that have not submitted bids.
- **Registered bidder:** registered suppliers that have submitted one or more bids but have not been awarded a contract.
- **Registered contractor:** registered suppliers that have been awarded one or more government contracts.

Potential suppliers that have not registered in the procurement system present a different challenge. They may be relevant to the assessment, particularly for understanding barriers to entry, but they will normally fall outside the formal sampling frame. Rather than attempting to construct a comprehensive list of all potential suppliers, the assessment team can extend access to the same survey through open or targeted dissemination

channels, such as chambers of commerce, business or supplier associations, professional organisations or other mechanisms capable of reaching firms outside the procurement registry.

Where possible, the survey should record how each response was obtained so that responses from the formal sampling frame can be distinguished from those received through open or supplementary outreach. Responses obtained through open or supplementary outreach can provide valuable evidence from firms outside the formal sampling frame. However, because these firms are self-selected and their probability of inclusion is unknown, their responses should not be combined with the probability sample to produce a single statistically representative estimate. They can instead be analysed separately as complementary evidence on potential suppliers not covered by the formal frame.

iii. Supplier survey – Selecting respondents and determining the number of invitations

Where a formal supplier sampling frame is available, respondents should be selected. The purpose of stratification is to ensure that each analytically relevant group is represented by enough completed responses to support separate analysis.

Within the formal frame, participation status and firm size are particularly useful stratification variables. Participation status helps ensure that the survey captures suppliers with different levels of interaction with the procurement system, whose perceptions and experience may differ accordingly. Firm size helps prevent the sample from being dominated by the most numerous size category and supports comparison of the experience of smaller and larger firms. Additional characteristics, such as region or sector, should be used for stratification only where they are relevant to the assessment.

As a practical planning reference, the assessment team should normally seek at least 30 completed responses in each analytically relevant stratum. This is a commonly used rule of thumb that provides a workable basis for analysing results within a group, but it is not a statistical guarantee of representativeness or precision. Larger samples may be needed where greater precision or more detailed comparisons are required.

The number of suppliers invited should then be determined from the target number of completed responses and the expected response rate:

Number of invitations = target completed responses / expected response rate

The calculation should normally be made for each stratum. For example, if three strata each require around 30 completed responses and the expected response rate is 25%, approximately 120 suppliers should be invited from each stratum, for a total of 360 invitations.

There is no standard response rate set by MAPS that can be assumed across assessments. The expected response rate should reflect available experience and factors such as the quality of contact information, survey length, mode of administration and the relationship of respondents with the procurement system. Where response rates are expected to differ materially across strata, the number of invitations should be adjusted accordingly.

During the survey, the assessment team should monitor both the number and composition of completed responses. If a relevant stratum is responding below expectations, additional invitations or reminders may be needed to obtain a sufficient number of completed responses before the survey closes.

iv. Supplier survey – Designing the questionnaire

The supplier questionnaire should be designed as a single instrument covering the relevant quantitative indicators across the assessment, rather than as separate questionnaires by MAPS pillar. Its structure should reflect differences in suppliers' experience with the procurement system, so that respondents answer only the questions they are in a position to answer.

The questionnaire should remain as short and focused as possible. Branching helps reduce unnecessary questions, initial characterisation and screening questions can be used to determine the relevant path through the questionnaire. For example:

Example – Questionnaire branching

All respondents
→ general perceptions of accessibility and barriers to participation

Registered in the procurement system?
→ registration experience

Submitted a bid?
→ bidding, competition and participation experience

Awarded a government contract?
→ contract execution and payment experience

Used the challenge or appeals mechanism?
→ experience with challenge and appeals

Questions should be formulated in language that suppliers can readily understand and should refer only to circumstances they can reasonably assess from their own experience. The wording of the MAPS quantitative indicator does not need to be reproduced literally where a simpler question will generate a clearer and more reliable response.

Response options should, where possible, use standardised categories or scales that can be processed directly into the required quantitative indicator. Open-ended questions may be included where they provide useful context but should remain secondary where the primary purpose is quantitative estimation. Characterisation questions should likewise be limited to variables that serve a clear purpose, such as questionnaire branching, analysis by firm size or participation status, assessment of respondent composition or application of sampling weights.

v. Supplier survey – Conducting the survey

Electronic administration will often be the most practical way to conduct a suppliers survey. For the formal probability sample, selected suppliers should normally receive individual invitations through the available contact information. At the same time, the same questionnaire may be disseminated through open links to chambers of commerce, business associations or other channels to reach potential suppliers outside the formal sampling frame. The source through which each response was obtained should remain identifiable.

The invitation should come from a recognised institution and clearly explain the purpose of the survey, who is being invited to participate and how responses will be treated. Where confidentiality or anonymity is offered, the collection process and subsequent handling of the data should be consistent with that commitment.

A response period of around one to two weeks will often be sufficient for an electronic survey if accompanied by active follow-up. Rather than leaving the survey open indefinitely, the assessment team should use reminders during the response period and shortly before closing.

Where electronic administration is not feasible, the survey may be conducted through other channels, such as telephone, in-person administration or assisted completion. The choice of channel should reflect the country context and the characteristics of the respondent population, while preserving the sampling design, the questionnaire and the consistent treatment of selected respondents. Alternative modes of administration should not introduce discretionary respondent selection or other changes that compromise the statistical integrity of the survey.

Throughout the response period, the team should monitor completed responses by the main sampling strata. If a stratum is responding below expectations, additional suppliers can be invited using the same probability-based selection procedure. This helps the survey approach the planned number of completed responses in each stratum without compromising the sampling design.

vi. Supplier survey – Analysing the survey results

Analysis should begin with the completed survey responses. For the formal probability sample, the assessment team should calculate the number of completed responses in each sampling stratum and the achieved response rate. These achieved responses should then be used to calculate the sampling weights applied to population-level estimates.

Population-level estimates should be weighted so that each stratum contributes according to its share of the supplier population. For a basic stratified design, the weight for each stratum can be calculated as the number of suppliers in the sampling frame divided by the number of completed responses in that stratum (N/r). For example, if one stratum contains 600 suppliers and produces 60 completed responses, while another contains 150 suppliers and produces 30 responses, the respective weights are 10 and 5. Applying these weights restores the relative contribution of each stratum when estimating results for the overall supplier population.

Use the sampling weights when estimating results for the supplier population as a whole. When reporting results for a specific stratum, calculate them directly from the completed responses in that stratum, provided that there are enough responses to support the analysis.

Results should not be disaggregated beyond what the achieved responses can reasonably support, and material limitations arising from low or uneven response should be reported.

Responses obtained outside the formal sampling frame through open or supplementary outreach should remain separately identifiable and should not be incorporated into weighted population estimates based on the formal frame. They may be reported as complementary evidence, as described earlier in this section.

Finally, survey results should be interpreted according to what they measure. Differences between survey findings and system-based or procurement-case indicators are not necessarily inconsistencies; they may reflect genuine differences between recorded performance and suppliers' experience or perception.

vii. Adapting the survey approach to other respondent populations

The supplier-survey approach should be adapted where MAPS quantitative indicators are directed to other respondent populations. Two cases require particular consideration.

For **public procurement buyers or the procurement workforce**, the population can usually be identified through procuring entities, staff rosters, e-Procurement users or similar administrative sources. Stratification is generally less relevant than ensuring adequate coverage of the officials who perform procurement functions. A particular concern is respondent confidence: officials may be reluctant to provide candid views if they believe their individual answers can be identified by their employer or supervisory authorities. The survey should therefore provide anonymity or appropriate confidentiality where needed and avoid collecting identifying information that is not necessary for the analysis.

The **PROF module** requires survey evidence from education and training providers. In this case, the main challenge is often constructing the sampling frame itself. The assessment team should first identify the relevant providers as comprehensively as possible, using desk research, institutional directories, accreditation or education records and other available sources. The resulting list can then be validated and expanded as providers are contacted. Once the frame has been established, the identified providers can be invited to participate in the survey using the general questionnaire-design, response-monitoring and analysis principles described above.

VI. Using the results of quantitative indicators

Once the quantitative evidence has been collected and processed through the workflows described in Sections III–V, the assessment moves from Conducting the assessment to Drafting the assessment report in the formal MAPS process. At this stage, the quantitative indicator inventory initiated during planning should first be completed with the results actually obtained.

For each quantitative indicator, the final fields should record the source actually used, the period covered, material coverage or limitations, the final status and the quantitative result. These fields should be added alongside the planning information rather than replacing it, so that the inventory preserves the link between what was initially planned and what was ultimately produced.

The final status should distinguish between indicators that were Estimated, Not Applicable, or Not estimated. The quantitative result should answer the indicator directly and concisely. More detailed calculations, breakdowns and methodological explanations can then be presented where they are useful for interpretation.

i. Presenting quantitative results

Each estimated quantitative indicator should begin with a direct statement of the result. The reader should not need to derive the answer from a table, chart or methodological explanation.

Additional information should be included only where it adds analytical value. Depending on the indicator, this may include the absolute numbers behind a percentage, a relevant disaggregation, a time trend, comparison with an equivalent source, or a material limitation affecting interpretation.

Absolute numbers can help show the evidentiary basis of a percentage. For example, a result of 80% has a different meaning if it represents 4 of 5 observations rather than 400 of 500.

Disaggregation can also reveal differences concealed by an aggregate result. Depending on the indicator, relevant breakdowns may include procurement method, year, entity type, firm size or another category supported by the evidence. Similarly, a multi-year series can show whether a result is stable, improving, deteriorating or associated with an exceptional period.

Where more than one source provides evidence on the same phenomenon, equivalent estimates may be compared where this helps validate or interpret the result. Evidence of stakeholder perception may also be presented alongside an observable measure, provided that the two are not treated as equivalent.

The level of detail should therefore vary across indicators. Some may require only a direct result, while others may justify a short explanation, table or breakdown. Material qualifications concerning source, period, coverage, number of observations or responses, proxies or other limitations should accompany the result where needed.

Example – Presenting a quantitative indicator

Quantitative indicator
Bids submitted online (in %).

Direct answer
In the most recent year, 68% of bids were submitted online.

Underlying magnitude
This corresponds to 13,600 of 20,000 bids received during the year.

Trend
The share increased from 54% three years earlier to 68% in the most recent year.

Relevant breakdown
Electronic submission reached 81% in open competitive procedures, compared with 37% in other procedures where electronic submission was available.

The direct result answers the quantitative indicator. The additional information is included because it helps explain the magnitude, trend and variation behind the aggregate result; not every indicator requires all of these elements.

If the quantitative indicator was either N/A or not estimated, as explained in Section II, it should be explicitly recorded in the report, including an explanation of why.

The completed inventory provides the bridge between the source-based work undertaken during the assessment and the MAPS assessment criteria to which the quantitative evidence ultimately contributes. It remains a working and traceability tool: the findings themselves should be incorporated into the relevant assessment criteria, gap analysis and assessment report in accordance with the MAPS Methodology.

ii. Using quantitative results in the qualitative analysis

Once the quantitative result has been recorded in the inventory, it should be incorporated into the MAPS assessment criterion it is intended to substantiate. This is where the source-based quantitative work reconnects with the standard MAPS assessment process.

For each criterion, the quantitative result should be considered together with the qualitative analysis. The two forms of evidence play complementary roles: qualitative analysis explains the legal, institutional and

operational arrangements, while quantitative indicators show the extent to which relevant conditions, practices or outcomes are observed in practice.

The assessment team should then consider what the combined evidence implies for the criterion and for the subsequent gap analysis. Quantitative and qualitative evidence may reinforce the same finding, reveal an important difference between formal arrangements and observed performance, or address different dimensions of the criterion that need to be interpreted together.

When the evidence points in the same direction

The quantitative result may confirm what the qualitative analysis already indicates. In this situation, it strengthens the evidentiary basis for the conclusion.

Illustrative example

Qualitative evidence
The e-Procurement system allows electronic bid submission, and interviews with procurement officials and suppliers indicate that electronic submission has become the normal way of participating in procurement procedures.

Quantitative evidence
In the most recent year, 88% of all bids and 84% of bids submitted by micro, small and medium-sized enterprises were submitted online.

Implication
The quantitative evidence reinforces the qualitative conclusion that electronic participation is widely embedded in procurement practice.

When the evidence points in different directions without being contradictory

A formal rule, institutional arrangement or procedure may exist while the quantitative evidence shows that it is not fully reflected in practice. The two findings describe different aspects of the same system and should both be reflected in the assessment.

Illustrative example

Qualitative evidence
The procurement framework requires contracting entities to publish their annual procurement plans.

Quantitative evidence
72% of entities required to publish a plan actually did so in the most recent year, down from 81% three years earlier.

Implication
The quantitative evidence does not contradict the existence of the formal requirement. It shows that implementation is incomplete and has deteriorated over the period analysed. The assessment should therefore distinguish between the formal framework and its implementation in practice.

In such cases, the difference between the qualitative and quantitative evidence may itself be one of the most relevant findings of the assessment.

When the evidence appears contradictory

In other cases, the qualitative and quantitative evidence may appear to make incompatible claims about the same condition. The difference should be investigated before the assessment conclusion is finalised.

Illustrative example

Qualitative evidence
Interviews with procurement and finance officials consistently indicate that late payment is exceptional and that contractual payment periods are routinely met.

Quantitative evidence
Payment data for the same entities and period show that only 48% of 320 invoices were paid within the applicable contractual period.

Implication
The assessment team should first verify whether both pieces of evidence refer to the same stage of the payment process, definition of timeliness, institutional population and reference period. If those differences do not explain the discrepancy, the underlying evidence should be reviewed before the assessment criterion is concluded.

Quantitative indicators should not be treated as automatic compliance thresholds. MAPS does not establish benchmark values that determine whether an assessment criterion is met. Quantitative results instead provide empirical evidence to support the interpretation of the criterion and may also serve as baselines for national targets or future measurement.

Their purpose is therefore not simply to produce a figure, but to strengthen the evidence on how the procurement system operates in practice and to support a more robust qualitative analysis and gap assessment.

Where quantitative and qualitative evidence appear to point in different directions, the assessment team should first determine whether the difference reflects a real feature of the procurement system or a methodological issue. Definitions, reference periods, coverage, source limitations or different interpretations of the assessment criterion should be checked before drawing conclusions.

Once these issues have been examined, the quantitative evidence may reinforce the qualitative analysis, qualify it or justify revisiting it. Where both sources remain valid but describe different dimensions of performance, the assessment should preserve that distinction and explain it rather than force the evidence into a single conclusion.

Glossary

The terms below are used throughout this guide and are defined here to clarify the relationships between the main concepts used for planning, obtaining and interpreting quantitative evidence.

What is being measured and how evidence is obtained

Required measure

The specific phenomenon, condition, outcome or perception that the MAPS quantitative indicator requires to be measured. It may need to be operationally clarified without changing the meaning of the MAPS indicator.

Proxy

A measure that is related to the quantitative indicator but does not measure the same phenomenon directly. A proxy may provide useful complementary evidence, but it should not be reported as if it were the quantitative indicator itself.

Source

The system, database, register, procurement file, institution or respondent population where the evidence needed to estimate the quantitative indicator resides.

Collection or processing route

The way in which evidence from a source is obtained or transformed into the quantitative result. A single source may support different routes, such as direct processing of record-level data by the assessment team or production of the required statistic by the institution that owns the data.

Alternative source or route

A different source or processing route capable of producing the same required measure when the initially preferred option is not feasible. An alternative may involve different data or procedures, but it should preserve the meaning of the quantitative indicator.

Who or what the evidence represents

Population

The complete set of cases, entities, transactions, suppliers, respondents or other units to which a quantitative indicator refers.

Sampling frame

The operational list or source from which units can actually be selected for a sample. The sampling frame may cover the full population or only part of it.

Coverage

The extent to which a source or sampling frame represents the population and period relevant to the quantitative indicator. Coverage can differ across indicators using the same source where the variables required for a particular indicator are not available for all observations.

How observations are selected and combined

Probability sample

A sample selected through a procedure in which eligible units have known probabilities of selection. Subject to the quality of the sampling frame and the achieved evidence, this allows results to be used to estimate characteristics of the population represented by the frame.

Stratum

A subgroup of the sampling frame defined before selection because sufficient representation or separate analysis of that group is required.

Sampling weight

A factor used to restore the appropriate contribution of observations or strata when their representation in the achieved sample differs from their representation in the population or sampling frame. Sampling weights reflect the sampling design and, where relevant, the effective number of observations contributing to the estimate.

Self-selected response

A response from a participant who chooses to enter an open survey or outreach mechanism rather than being selected through the formal probability sample. Such responses can provide useful complementary evidence, but their probability of inclusion is unknown and they should not be treated as statistically representative.

How the indicator is classified and recorded

Applicable

Whether a legitimate population, event, condition or process exists to which the quantitative indicator applies within the assessment scope and reference period. Applicability is independent of whether the evidence needed to estimate the indicator is available.

Not Applicable (N/A)

Final status used when there is no legitimate population, event, condition or process to which the quantitative indicator applies within the assessment scope and reference period. Lack of data does not make an indicator Not Applicable.

Estimated

Final status used when an applicable quantitative indicator has been calculated using evidence sufficient to support the reported result.

Not estimated

Final status used when the quantitative indicator is applicable but the evidence required to produce the quantitative result could not be obtained or was insufficient.