Data Collection and Sampling Guidelines:

International Programme 2025-2028

1. Introduction

The South Programme of the International Programme 2025-2028 of Interaction encompasses a total of 35 different indicators. For each indicator, specific guidelines have been created. However, these specific guidelines do not address the general issues on data collection and sampling outlined in this document. The main objective of these data collection and sampling guidelines is to ensure the quality of the data we need for reporting, but above all, to support the improvement of our projects and organizations. Data quality could be divided in five standards[1]:

- **Validity**: data effectively measure the intended indicators.
- **Integrity**: data are free from errors, duplicates, and missing values.
- **Precision**: data have an adequate level of detail, including appropriate disaggregation (by age, gender, etc.).
- **Reliability**: data collection methods are consistent to ensure stability and reproducibility.
- **Timeliness**: data are available when needed and up to date.

"Data you collect will never be free of bias. Thus, you need to determine, with the help of your stakeholders, what quality and quantity of data is "good enough" for your decision-making, learning and accountability needs" (MEAL DPro (2019), p. 74)

Although these guidelines are primarily intended to be used by partners of the international programme, we believe they could also be useful to other Interaction members dealing with similar data collection issues. If you have any questions or notice any inconsistencies in this document, please do not hesitate to contact Interaction. We hope these guidelines will help us to improve the quality of our data.

[1] Adapted from: A Guide to the MEAL DPro (2019) by Catholic Relief Services, the Humanitarian Leadership Academy, and Humentum

2. Theory of Change and Logframe

A thorough Theory of Change was developed with all Interaction's members taking part in the IP 25-28. This Theory of Change was consequently used to build the common Logframe for the whole programme. The following table provides an overview of the logical framework structure, illustrating the relationships between Outputs, Outcomes, and Impact with respective indicators examples:

Logframe of the International Programme 2025-2028

		Type of Indicator	Example(s)
Impact	**Sphere of Concern/Interest:** The Impact represents long-term, broad-scale changes that rely on our multiple outcomes and are influenced by various additional factors and stakeholders.	Qualitative indicator	Stories collected with the Most Significant Change technique (MSC)
Outcomes	**Sphere of influence:** While we do not have direct control over outcomes, we can influence them through our outputs.	Mostly percentage indicators (%)	% of farmers increasing their production / % of population experiencing moderate or severe food insecurity (FIES)
Outputs	**Sphere of control:** We have direct control through our activities.	Mostly numeric indicators (#)	# of farmers trained
Activities & Inputs			Plan, fund and organise training sessions for farmers

3. Numeric and Percentage Indicators

20 indicators are framed as numeric indicators (mostly Output indicators) and 11 as percentage indicators (mostly Outcome indicators). Numeric indicators have simplified specific guidelines, as they do not necessitate complex computation or survey. On the contrary, percentage indicators require more thorough guidelines to explain computations and survey questions. As they are mostly collected with surveys, they also require rigorous random samplings, unless the project population is small enough to survey all participants/households (e.g. 50 participants/households). The indicator guidelines can be found on our website: https://interaction-schweiz.ch/unsere-wirkung/.

➢ Person-reach-tracking OC 1-3 indicators

In addition to Output and Outcome indicators, we use a person-reach-tracking indicator. To effectively track the number of unique persons reached each year per Outcome, we suggest that partners maintain some kind of central database. This database should allow:

- Distributing project participants by age and gender,
- Measuring project outputs framed as numeric indicators (e.g. # farmers trained in agroecology)
- Avoiding double counting.

For each Outcome, the database could be organized in a table format as follows:

OC 1	Male	Female	>18 yearsold	<18 yearsold	Health care consultation, screening or treatment provided	Participated in support group sessions	Etc.
Person 1	0	1	0	1	1	1	
Person 2	1	0	0	1	0	1	
Person 3	0	1	0	1	1	1	
Etc.							

Note: 1 = Yes; 0 = No

This kind of database could be built in an Excel sheet or in an advanced monitoring software. it should simplify the determination of sample frames (see below); however, in large and complex projects, maintaining such a database—and consequently the sample frames—can become nearly impossible.

➢ Random Sampling

Random sampling is a statistical method used to select a subset of project participants/households (i.e. a sample) from a project population in such a way that everyone has an equal chance of being chosen. This method ensures that the sample is representative of the project population, reducing bias and allowing for generalizations to be made about the project population from the sample.

⚠ Generalizations made about the project population cannot be considered as representative of the whole population.[2]

A few key steps are followed to determine a random sample. First, we need to define the project population, the sample unit and the sample frame:

- **Project population**: the entire group about which we want to make a generalisation.
- **Sample unit**: the individual element selected from the project population to be part of the sample and for which we will collect data (usually individuals or households).
- **Sample frame**: the numbered list of all unique units in the project population.

Example:

- **Project population**: 1000 women who were sensitised on hygiene.
- **Sample unit**: a woman who was sensitised on hygiene.
- **Sample frame**: a list with the name of every woman who participated in the sensitisation session, each associated with a number.

⚠ In many project situations in the Global South, it is not feasible to have an adequate Sample Frame because precise demographic data do not exist! However, the whole process of randomization is very dependent whether a sampling frame is available or not, so we will cover both situations in depth in the following pages.

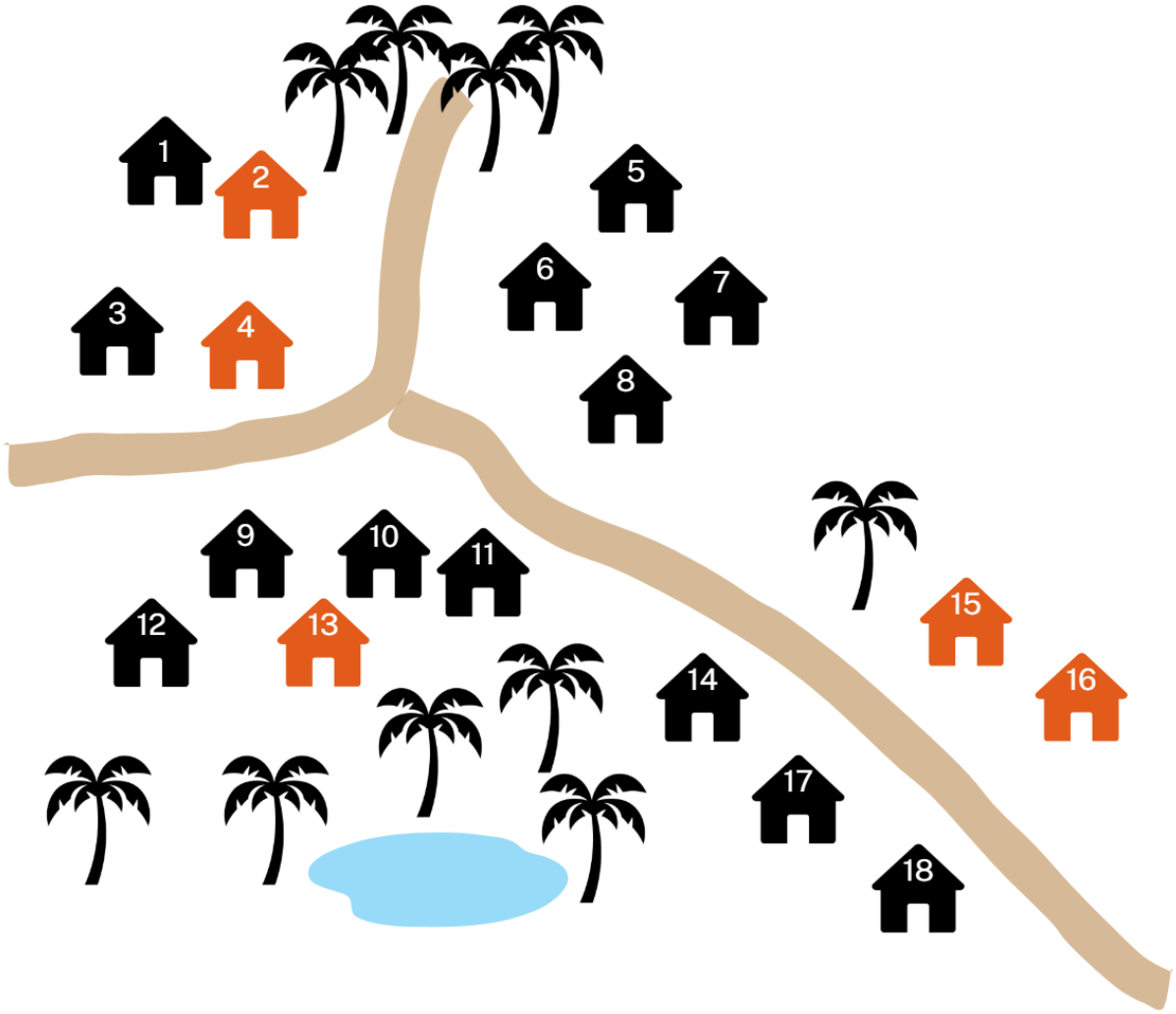
---- With Sample Frame ----

1. **Choose your random sampling method:**

There are different methods to determine a random sample:

- **Simple Random Sampling:**
  The most basic form of random sampling. Each member of the population is assigned a unique number, and numbers are drawn randomly.

[2] The project population often has specific characteristics (e.g., female rural farmers), making it distinctly different from the general population (e.g., all inhabitants of a region or country).



In this example, households N°2, 4 ,13 ,15 & 16 were randomly selected

One easy way to do that is to use the following Excel function [3]:

If we need to determine X random numbers to choose among a population of 20 project participants:

Minimum Value Maximum Value

=INDEX(UNIQUE(RANDARRAY(X^2;1;1;20;TRUE));SEQUENCE(X))

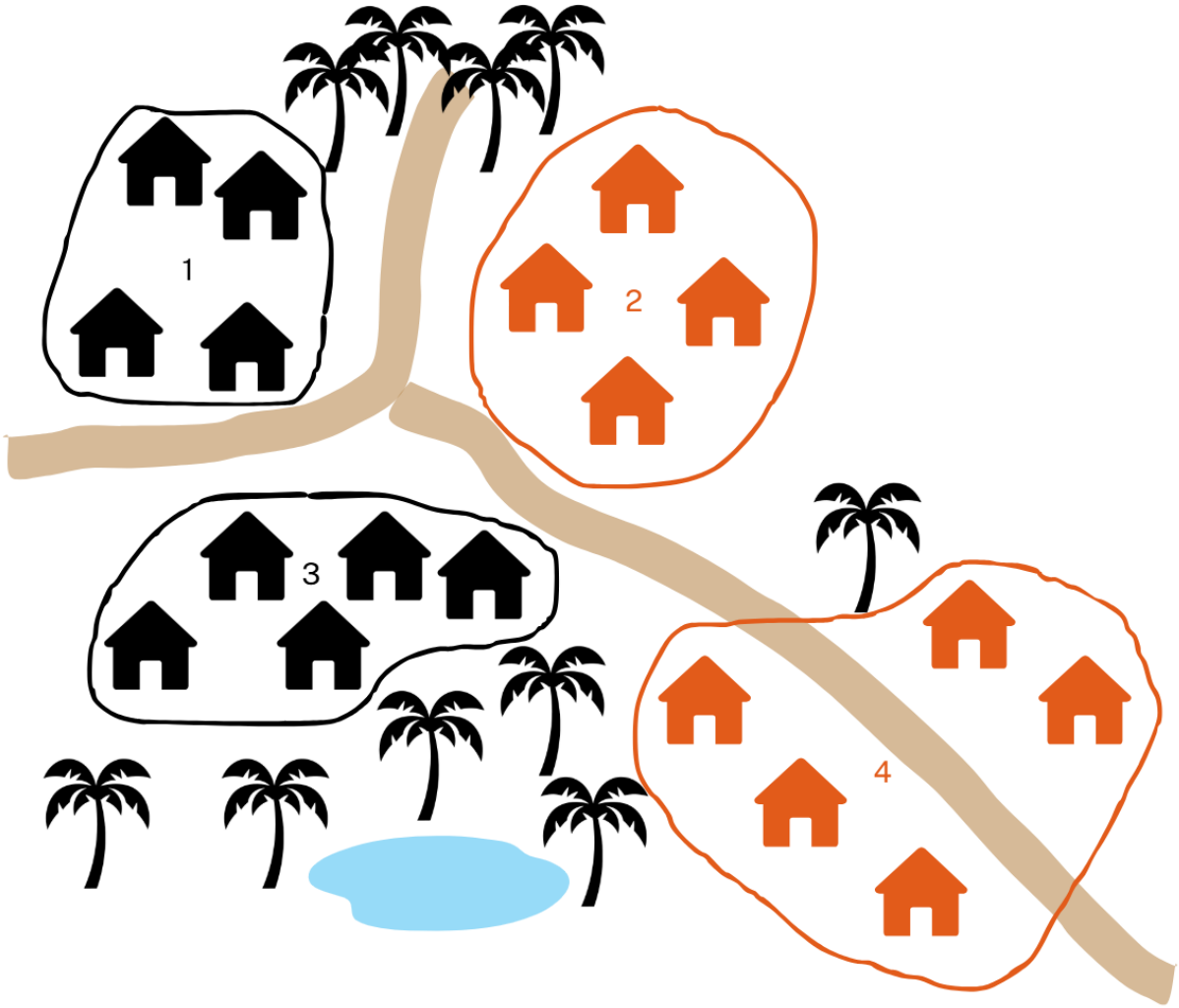
Number of random numbers needed

Once you have your random sample, make sure to copy and paste the values to prevent Excel from generating a new set of random values whenever you make changes of perform other actions in the spreadsheet!

[3] If you need further explanation, you can watch the following tutorial: https://www.youtube.com/watch?v=iUFaqlkelJw

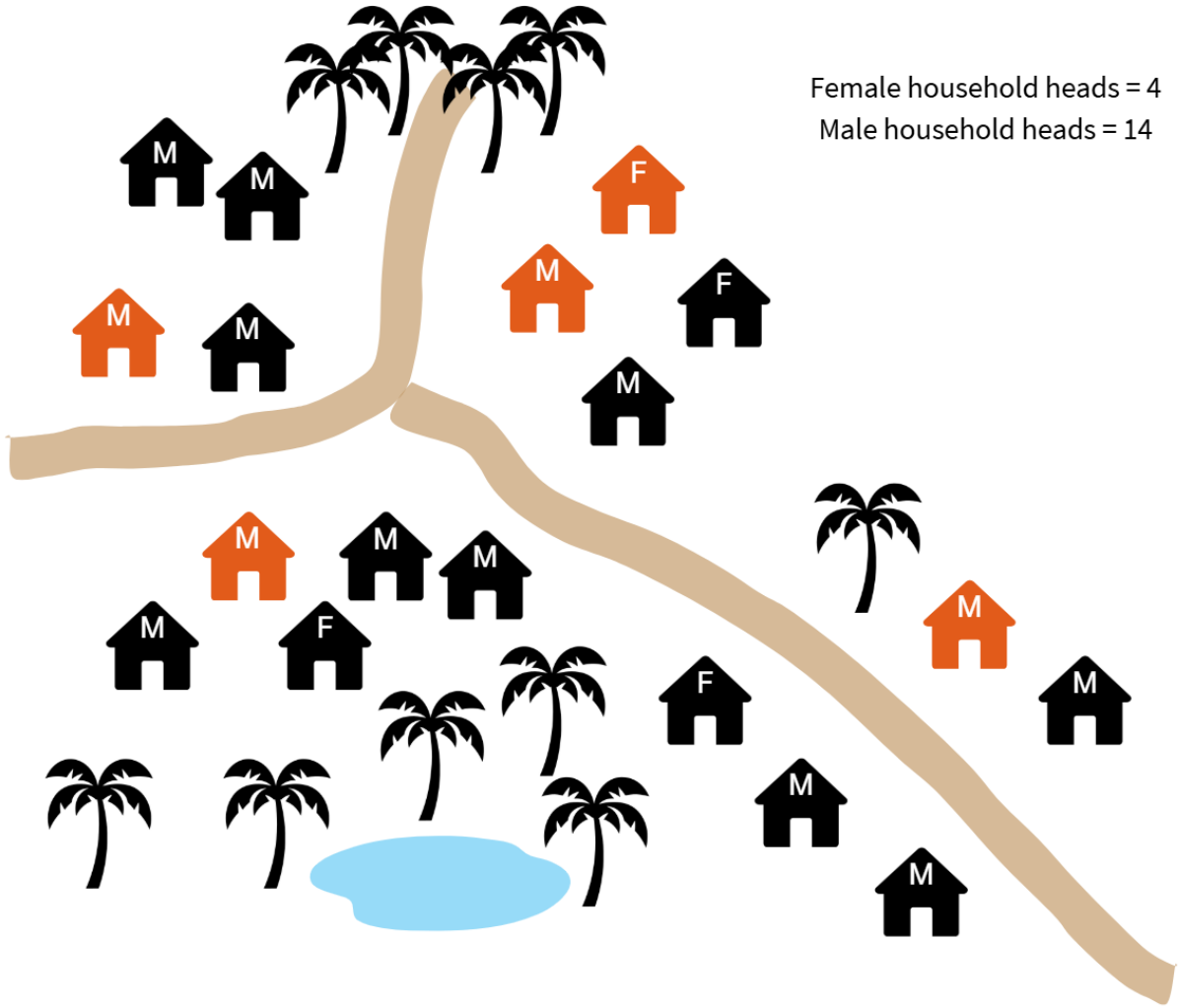
- **Systematic Random Sampling:**
  A starting point is randomly selected, and then every n-th member of the population is chosen. Example: Selecting every 5th participant from the list of project participants after randomly picking a starting point.

- **Cluster Random Sampling:**
  The population is divided into clusters (usually based on geographic or natural groupings). Entire clusters are randomly selected, and all members of the selected clusters are included in the sample. If each cluster contains too many people, a representative random sample can be selected within each cluster.



In this example, neighborhoods (i.e., clusters) N° 2 & 4 were selected, all households within these two neighborhoods will be surveyed

- **Stratified Random Sampling:**
  The project population is divided into homogeneous subgroups (strata) based on certain characteristics (e.g., age, gender). A random sample is then taken from each subgroup proportionate to its size in the population.



In this example, 4 male household heads were randomly selected and 1 female household heads were selected, respecting their proportions in the population

2. Determine your sample size:

For a sample to be representative of the population, it needs to be large enough. However, if the sample size is too large, we will waste precious time and money to collect all data. In the examples above with the little village, the sample sizes may be too small to represent the entire population accurately.
Several formulas exist to calculate the minimum sample size. For our purposes, we suggest that the Yamane's simplified formula for sample size is sufficiently reliable:

$$n = \frac{N}{1 + N \cdot e^2}$$

Where:

- n = sample size
- N = Priority population size
- e = level of precision/margin of error (usually 5%, or 0.05)

Example for a priority population of 1000 women sensitised on hygiene:

$$n = \frac{1000}{1 + 1000 \cdot 0.05^2} = 285.71$$

Thus, according to Yamane's formula, we would need to survey **285** women to obtain a representative sample. Alternatively, the IndiKit calculator provides an easy way to determine sample size: Indikit Calculator

Example: Using a 95% confidence level and a 5% margin of error, the required sample size for a population of 1,000 is **278**.

Both methods yield very similar results. However, Yamane's formula or the IndiKit calculator have limitations. For instance, they do not account for population variability; the more diverse the priority population, the larger the sample size needed to achieve the same level of precision.

3. **Randomly select the sample units**

Example: We can use the Excel function mentioned above to generate 278 random numbers. Using the list of the 1000 women who participated in the sensitization session, we can then identify the 278 women to be surveyed.

---- Without a sample frame ----

1. **Choose your random sampling method:**

Without a sample frame, other strategies need to be put in place to be able to collect relevant data:

- **Use a Proxy List:**
  If you cannot obtain a full list of the population, you might consider using a proxy. This could be a partial list, such as community, public institutions or service providers records that can approximate the full population.

- **Random Walk Method:**
  Use a random selection of streets, directions, or households (e.g., by rolling dice or drawing lots to determine which way to walk or how many doors to skip).

- **Snowball Sampling :**
  Start with a few randomly selected participants (through any means available) and ask them to refer others from the same population. While this introduces some bias, it's one method when other options are limited.

- **Systematic Random Intercept :**
  Approach every n-th person or use a random time interval (e.g., every 10 minutes, the next person who passes).

- **Cluster or Spatial Sampling:**
  You can divide the geographical area into clusters (villages, blocks, or other units) and randomly select some of these clusters. Then, survey everyone or randomly select participants

within the chosen clusters. In the absence of a formal list, you can randomly sample based on geographical coordinates or grid systems. Tools like GIS (Geographic Information Systems) can assist in identifying random points within a specific area.

While random sampling without a frame can introduce some biases, applying one or a combination of these techniques can still help approximate randomness and reduce bias where a full sampling frame isn't available.

2. **Determine your sample size:**

For a sample to be representative of the population, it needs to be large enough. However, if the sample size is too large, we will waste precious time and money to collect all data. The minimum sample size can be calculated with the same methods as above (p. 7-8). However, as the total population is unknown, an approximation is used. An approximation is especially valid with large priority population as the required sample sizes are quite similar (**e.g., 370 for a population of 10000 and 377 for a population of 20000**).

3. **Randomly select the sample units:**

As no sample frame is available, we should simply start data collection using one of the sampling methods suggested above and stop when we have reached an adequate number of respondents.
Example: We randomly walk through a village, and we survey households until we have surveyed 278 people. Or we randomly selected enough clusters to reach the sample size.

➢ Regarding disaggregation of indicators

---- With Sample Frame ----

Most indicators would ideally need stratified sampling to divide by age (>18, <18) and gender (Female/Male). This means we would have 4 different subgroups: Female > 18, Female < 18, Male >18 and Male <18.

However, it may be too complicated to collect percentage indicators with 4 subgroups for the following reasons:

1. **Large sample size:** To obtain representative results with a reasonable margin of error (<5%) for each subgroup (Male <18, Female <18, Male >18, Female >18), we would require very large sample sizes meaning more time and money. See the example below:

	Population size	Sample size *	margin of error
Male > 18	350	183	5%
Female > 18	300	169	5%
Male < 18	100	80	5%
Female < 18	250	152	5%
Sum	1000	584	3%
Priority population	1000	278	5%

* Used Indikit calculator to calculate each subsample independently (CL: 95%; MoE: 5%)

If we want every subgroup to have a margin of error of 5%, we would need a total sample size of 584, which is more than twice the sample size required if we only consider the total priority population (278 with 5% of error).

2. **Complexity of surveys:** Since most % indicators will be collected through household surveys, it is challenging to capture data disaggregated by both age and gender, as generally the household head responds to the survey.

3. **Limited demographic data:** In cases partners do not have clear demographic data on their priority populations, it is impossible to calculate representative subsample sizes for these subgroups.

For these reasons, we suggest that we only focus on disaggregation by gender for % indicators. Even then, obtaining representative data for both male and female groups remains difficult. For example, with a population of 1,000, a representative sample (margin of error <5%) would be 278. But if the population is 50% male and 50% female, we would need to survey 217 males and 217 females (434) to achieve representative results for each subgroup. This increases the sample size by 56% compared to a single representative sample.

In our case, where we have limited resources for monitoring, the most feasible option would be to use proportionate stratification. This means the sample size for each subgroup is proportional to its size within the population:

			Proportionate sampling strategy		
			1) Use Indikit calculator to calculate the whole population sample size (CL: 95%; MoE: 5%) 2) Calculate each subsample with their respective proportion in the population: *(Priority population sample size / Priority population size) × subgroup size*		
	Size	Proportion	Size	Proportion	margin of error
Priority population	1000	100%	278	100%	5%
Male	750	75%	194,6	75%	6%
Female	250	25%	55,6	25%	12%

With this method, we can achieve a reasonable sample size that maintains the same proportion of males and females as in the priority population. This makes the overall sample more representative of the population, allowing for safer generalizations about the priority population. However, caution is needed when interpreting and comparing subgroups, particularly when a subgroup is small, as this leads to a higher margin of error.

---- *Without Sample Frame* ----

The situation is complex, as without a proper sampling frame, it is not possible to determine the size of the priority population or any subgroup. Similarly to situations with a sampling frame, we should limit our disaggregation to male and female only. An estimation of each subgroup's proportion can be made, for example, based on the gender distribution in the country. In this case, even greater caution should be exercised when analysing the data, both for the priority population and subgroups, as multiple biases may have been introduced.

➢ Data Collection

Data collection related to numeric indicator is straight forward as long as it is clearly defined what should be counted. The indicators guidelines provided should be sufficient to ensure accurate and consistent data collection. Most percentage indicators rely on being determined with quantitative surveys. Again, the definition of the indicator is crucial to elaborate the right question to collect data.

A few steps should guide the process of collecting adequate quantitative data through survey:

1. **Prepare the Questionnaire**
   - Design your questionnaire based on the indicator guidelines to ensure consitence and clarity.

2. **Define the Sampling Method**

- Select an appropriate sampling method and create a representative sample, as explained earlier.

3. **Select and Train Interviewers**
   - Choose interviewers carefully and train them thoroughly. Ensure they fully understand each question. Role-playing exercises can be particularly helpful in preparing them.

4. **Test the Questionnaire**
   - Conduct a pilot survey with a small group of respondents to identify and address potential issues.

5. **Conduct the Interviews**
   - Obtain the respondent's informed consent, ensuring they understand and agree to the confidentiality terms.
   - Choose the best time to conduct the interview to respect the respondent's schedule.
   - Use a language the respondent understands and create a relaxed environment where they feel comfortable.
   - Emphasize the importance of their responses and reassure them that negative answers or refusals are acceptable.

➢ Data cleaning

Data cleaning is the process of identifying, correcting, or removing errors and inconsistencies in data to improve its quality. This process ensures that the data is accurate, complete, consistent, and usable for analysis.

A few steps should guide the process of Data Cleaning:

- **Removing Duplicates:** Identifying and deleting duplicate records to ensure each data point is unique.

- **Handling Missing Data:** Find out if the raw data contains missing values, if missing values cause an issue for a specific analysis, we can disregard the respective rows/columns with them.

- **Correcting Errors:** Identifying and correcting errors in data entries, such as misspellings, incorrect values, or formatting inconsistencies.

- **Standardizing Data:** Ensuring consistency in data formats (date formats, units of measurement, etc.) across the dataset.

- **Filtering Outliers:** Identifying and deciding whether to keep or remove unrealistic extreme values that may skew analysis results.

- **Checking logical consistency:** For example, a 7-year-old boy could not have answered questions about his pregnancy.

4. Qualitative Impact Indicator: Most Significant Change (MSC)

The international programme planned to use the Most Significant Change (MSC) technique[4] to provide additional insight into the impact of the programme. A specific guideline on MSC has been created and is available on our website: https://interaction-schweiz.ch/unsere-wirkung/ . So, here we will focus only on the sampling issues related to this technique.

➢ Purposeful Sampling [5]

> ***"In many instances more can be learned from intensively studying extreme or unusual cases than can be learned from statistical depictions of what the average case is like." (Patton, 1990, p. 170)***

Purposeful sampling, also known as purposive or selective sampling, is a technique for collecting qualitative data where the MEL specialist deliberately selects participants based on specific characteristics or criteria relevant to the project or its objectives, rather than using random selection. Because the sample is selected non-randomly, results **cannot be generalized** to the entire population. Instead of seeking basic quantitative data that applies to a large population, this approach focuses on **a few information-rich cases** that can provide **deep insights** on important issues, including **unexpected findings**.

Purposive sampling can take multiple forms, among others:

Homogeneous (or Criterion or Expert) Sampling

- Purpose: To focus on a specific subgroup with similar characteristics, experiences or expertise.
- Example: Selecting participants who are all female and have local leadership duties to explore gender-specific leadership challenges.

Heterogeneous (or Maximum Variation or stratified) Sampling

- Purpose: To capture a wide range of perspectives by including participants with diverse characteristics.
- Example: Choosing participants from different age groups, genders, and socioeconomic backgrounds to explore a broad range of experiences and viewpoints on hygiene issues.

Typical Case Sampling

- Purpose: To select participants that are considered "typical" or representative of a phenomenon.
- Example: Choosing a few farmers that represent the average performance level in a region to study the effects of a new agroecological training.

Extreme or Deviant Case Sampling

- Purpose: To study outliers or unusual cases that may provide unique insights.
- Example: Selecting the top 1% of students who excelled or those who failed spectacularly to understand the factors contributing to their success or failure.

[4] Davies, R. & Dart, J. (2005), The 'Most Significant Change' (MSC) Technique: A Guide to Its Use, version 1.0
[5] Patton, M. Q. (1990), *Qualitative Evaluation and Research Methods*. Sage Publications, Newbury Park, CA.

Critical Case Sampling

- Purpose: To select a small number of important cases that can provide the most information about a specific phenomenon.
- Example: Choosing a key community where a WASH intervention has been highly successful or failed to understand critical success or failure factors.

Snowball or Chain Sampling

- Purpose: To identify participants through referrals from initial subjects.
- Example: Asking initial participants to refer others who may give important positive or negative insights on the project.

For the start of the program, we do not want to promote any specific form of purposive sampling. Instead, we encourage local MEL specialists **to test different, locally adapted options** for collecting stories. In some cases, very insightful positive or negative stories may not emerge from active searching but rather be heard in the **normal** course of fieldworkers' activities. The assumption here is that **fieldworkers** will naturally learn about change stories as part of their daily work because they have close, regular contact with the project participants.[6] Field workers would then select where to collect stories based on this knowledge.

➢ Focus group or semi-structured interviews?

Both can be used for qualitative data collection. Focus group can help create interactions between the participants and obtain perhaps divergent point of views. However, they require better facilitation skills, for example to prevent a person from monopolizing the discussion. For the MSC technique specifically, semi-structured interviews may be more suitable, as stories are quite personal.

➢ How many people should we sample for qualitative interviews?

Qualitative surveys are often criticised for their very small sample sizes. However, given the time and financial requirements of qualitative methods, it is unrealistic to expect sample sizes comparable to those of quantitative data collection. That said, large samples are not necessary to collect valuable qualitative data. Recent research suggests that:

Quality is better that Quantity :

- Min 3 Focus groups of 8-12 people

- Min 6 single interviews

- as few as three focus groups (of typically 8-12 people) are enough to capture about 80% of all relevant themes and adding more focus groups only improve marginally the results.[7]
- only 6 semi-structured interviews can provide insights on 80% of all relevant themes.[8]

However, these numbers might be only valid for relatively homogenous priority population. Nonetheless, it may be more effective to invest in improving the facilitation of interviews and focus groups rather than increasing their number.

[6] Davies, R. & Dart, J. (2005), The 'Most Significant Change' (MSC) Technique: A Guide to Its Use, version 1.0, p. 24
[7] Guest, G., Namey, E., & McKenna, K. (2017). How many focus groups are enough? Building an evidence base for nonprobability sample sizes. *Field methods*, *29*(1), 3-22.
[8] Gandy, K. (2024). How many interviews or focus groups are enough?. *Evaluation Journal of Australasia*, 1035719X241266964.

5. Final Remarks & Additional Resources

These guidelines only provide general recommendations for improving data collection and sampling practices, both processes must be adapted to the local context. Given our limited resources for monitoring and the significant variability and volatility of local conditions, we will inevitably make certain assumptions or compromises in order to collect relevant data. As a result, data gathered through the International Programme should be interpreted with caution and not considered as rigorous as findings from a formal scientific study.

Resources
INTRAC (2020). Qualitative and Quantitative Methods
INTRAC (2017). Sampling
Davies, R. & Dart, J. (2005). Most Significant Change (MSC) Guide
Humentum (2019). Monitoring, Evaluation, Accountability and Learning for Development Professionals Guide
Gandy, K. (2024). How many Interviews or Focus Groups Are Enough?
Patton, M. (1990). Qualitative Evaluation and Research Methods

Acknowledgement

Gratitude goes to the reviewers who provided meaningful input substantially improving the quality of this document: Nicola Malacarne, Mikaël Amsing, and all Interaction's members for their precious feedback.

Learn more about Interaction's international programme 2021-2024 *Learning 360*:

Programme Overview – German & French
How we understand Impact – German & French

---

info@interaction-schweiz.ch; www.interaction-schweiz.ch.

IP 25-28 MEL guidelines, version of March 2025.

Author: Thibaud ROSSEL, Programme Coordinator Monitoring, Evaluation & Learning (MEL), Interaction.
Thibaud.rossel@interaction-schweiz.ch