CHAPTER 5

DATA ANALYSIS

Once you finish collecting the data, you should start analysing it. This means using all the research material (information obtained with various methods) and answering evaluation questions as well as valuing the evaluated project according to chosen criteria. Therefore, at this stage, it is worth going back to the evaluation concept, which acts as a compass, leading the evaluator through the entire research process (not only information collection, but also data analysis, drawing conclusions and formulating recommendations).

The **purpose of data analysis** is:

- Compilation and verification of collected information,
- Description, assessment and juxtaposition of the quantitative and qualitative data that is obtained (checking how reliable and consistent they are),
- Identification and explanation of various cause and effect relationships that will allow you to understand the mechanisms of the studied phenomena,
- Interpretation of the obtained evaluation findings in relation to wider knowledge about the subject of the evaluation (evaluandum),
- Obtaining detailed answers to evaluation questions and credible valuing of the evaluandum according to chosen criteria,
- Drawing conclusions from the collected information and formulating useful recommendations based on it.

In the data analysis, you should bear in mind the principle of **triangulation**, i.e. the compilation of data obtained from various sources, using various research methods, by different researchers. Thanks to this, you have the opportunity to supplement, deepen and verify respective information in order to obtain a full picture of the evaluated project.

Although during data analysis the actions undertaken are common to both types of data (quantitative and qualitative), such as **reduction, presentation and concluding**, the obtained findings are in a different form for each of them. The comparison of these data is presented in the table below.

OPERATIONS	QUANTITATIVE DATA	QUALITATIVE DATA
data reduction	Calculating percentages, averages, and other measures	Selection, simplification, generalisation, summary
presentation of findings (in a consistent form)	tables, charts, diagrams	Text studies, summaries, diagrams, matrices, networks of connections
drawing conclusions	Statistical testing of hypotheses concerning the relationship between the studied phenomena	Noticing patterns, regularities, deviations, developing explanations

Before starting the data analysis, it is necessary to check whether all research materials have been **anonymised**, i.e. there are no personal data (names, surnames, addresses, including e-mail addresses, telephone numbers etc., as well as contextual information enabling the identification of research participants). Interviewees who participated in the qualitative part of the research (IDIs, FGIs) are given pseudonyms, e.g. taking into account the features important for the researcher. The personal information concerning research participants should be **separated** from the content data provided by them.

There are four main stages of data analysis:

1. Selection and ordering of the collected research material - during this stage, the correctness and completeness of the data are checked, the reliability of every piece of information is verified (thanks to triangulation), and data that is not useful for the purpose of the evaluation is removed. You should collect all the information and facilitate its further analysis - recordings of the interviews can be transcribed or written down in accordance with a previously prepared scheme (which includes a summary of the respondents' statements). In the case of a survey, you should remove uncompleted questionnaires from the analysis, etc.

2. Constructing analytical categories (selecting the type of encoding and data coding - their categorisation and classification) - this means assigning codes / "labels" to each piece of information obtained, representing specific categories of information, thus allowing for the organisation of the research material.

- In the case of **closed-ended questions,** the answer codes take a **numerical form** (e.g. "female" = 1, "male" = 2), which allows you to analyse the obtained data using statistical programs (or spreadsheets). First, you need to create a coding instruction that contains the names of codes and the numbers which were used in the questionnaire to identify answers given by the respondents to particular questions. Paper surveys require manual coding - to do this, you need to number the answers in the questionnaire, code the answers and enter this information into the database. Electronic surveys are coded automatically.

- In the case of **open-ended questions** and other qualitative data, the codes for particular answers have **verbal form** (e.g. "training organisation", "conducting a training"). Codes for qualitative data can be planned before or after reading the entire material. The first method is called "top-down" coding, which results from a good knowledge of the research problem and / or its grounding in a given theory. The second method is open coding ("bottom-up"), which consists of categories identified in the collected material (e.g. relating to research questions). In both cases, you need to develop a coding scheme that will organise the codes (establish a code hierarchy, superior / collective and detailed codes), so that you can present the collected information in a consistent form.

Table 4. Example of qualitative coding

Assessment of training modules (code):	
Relevance of each training (subcode)	*Adapting the training to the recipients' needs, adapting the transferred knowledge to the level of their competence.*
Organisation of training (subcode)	*Duration (too long / too short), amount of transferred information (too little / too much), assessment of the type of classes (lectures, workshops), theory-practice ratio, amount of time devoted to discussion and trainees' questions.*
Assessment of the trainers (subcode)	*The way of transferring knowledge, using examples, encouraging participants to ask questions, exchange of experiences by trainees.*

The information corresponding to the given codes can be summarised in one table (Tool 7, next page), which will facilitate the search for **similar or common** elements for the research participants as well as information **that differentiate them**. It also allows you to see the relationship between the interviewees' characteristics or situation and their statements.

3. **Analysis and interpretation of the obtained findings** (explanation and assessment by the researcher of a particular issue / problem)

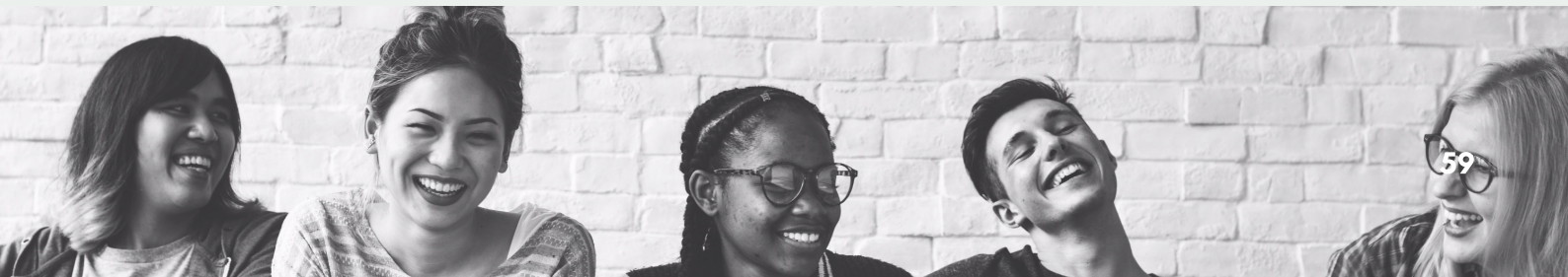
Data analysis is an important element of evaluation because it allows you to summarise the findings and find common and divergent elements in the collected materials. It is worth choosing and describing the method of data analysis at the stage of planning the evaluation. Data obtained during evaluation can be analysed in a number of ways. The simplest distinction is division into:

- **Quantitative data analysis** (numbers, answers to closed questions) - for simple analyses you can use, for example, MS Excel, and for more complex analyses statistical programs, such as SPSS or Statistica, operated by specialists, whose services can be used if necessary.

PRACTICAL TIP
For small groups, quantitative data should not be presented in the form of percentages, i.e. informing that 20% of respondents in a group of ten have a particular opinion. Better to use absolute numbers and say that it is two people.

- **Qualitative data analysis** (e.g. text, interview statements) - for simple analyses, it is enough to compile the data in a chart / matrix, and for more extensive research material, it is worth using programs that facilitate the analysis, e.g. QDA Miner, OpenCode, Weft QDA.

Some of them are briefly presented in the table at page 60.



Tool 7. Table for summarizing information from interviews

	Person 1	Person 2	Person 3
Information about the interviewee (place of employment, work experience, participation in other training courses on this subject)			
Adequacy of training to the needs of the interviewee (subject, level, method of conducting training, assessment of training materials, missing elements, organisation method)			
Impact and sustainability of training effects (usefulness of acquired knowledge and skills, their impact on the professional situation and personal development of the interviewee)			
Summary (strengths and weaknesses of the training, additional comments, recommendations - what should be changed in the future and how)			

Table 5. Options for analysing numerical data

Analysing numerical data such as cost, frequency or physical characteristics.	MS Excel functions
Frequency tables: arranging collected data values in ascending order of magnitude, along with their corresponding frequencies, to ensure a clearer picture of a data set.	FREQUENCY
Cross Tabulations: obtaining an indication of the frequency of two variables (e.g., gender and frequency of school attendance) occurring at the same time.	Pivot tables functions
Correlation: a statistical technique to determine how strongly two or more variables are related.	CORREL
Measures of central tendency: a summary measure that attempts to describe a whole set of data with a single value that represents the middle or centre of its distribution (i.e. arithmetic mean, median, mode)	AVERAGE MEDIAN
Measures of dispersion: a summary measure that describes how values are distributed around the centre (e.g. standard deviation, variance).	STDEV VAR

In **Table 5** you can see options for analyzing numeric data and in **Table 6** are options for analyzing textual data.

The tables are our own elaboration based on:

Peersman, G. (2014). Overview: Data Collection and Analysis Methods in Impact Evaluation, Methodological Briefs: Impact Evaluation 10, UNICEF Office of Research, Florence.

Table 6. Options for analysing textual data

Analysing words (spoken or written), including interviews, documents and open questions in questionnaires.
Content analysis: reducing large amounts of unstructured textual content into manageable data relevant to the research / evaluation questions.
Narratives: construction of coherent narratives of the changes occurring for an individual, a community, a site or a programme or policy.
Timelines: a list of key events, ordered chronologically.
Thematic coding: recording or identifying passages of text or images linked by a common theme or idea, allowing the indexation of text into categories.

IMPORTANT TIP

When analysing the data, it is very important to determine what changes have occurred as a result of the project and what role respective activities played in them.

Therefore, it is necessary to answer the question to what extent the project activities influenced the achievement of the assumed result indicators and what was the role of project activities among other factors influencing the expected changes (see chapter 2.5).

When analysing data, it is worth referring to the previously described theory of change adopted as part of the description of the project logic. When planning the change at the beginning of the project, you made certain assumptions about the conditions that must be met (resources provided, implemented activities) in order to achieve the given results, i.e. you have planned the cause-and-effect chain. Evaluation verifies our theory of change - it can confirm it or show some gaps in it (e.g. missing / redundant elements) and recommend improvements for the future.

There are three general strategies for **causal inference.** Using a combination of these strategies can help to increase the credibility of the conclusions drawn:

Table 7. Data analysis approaches for causal attribution with various options

Counterfactual approach	Developing an estimate of what would have happened in the absence of the programme or policy; this implies the use of experimental and quasi-experimental methods (so also a control or comparison group) or modelling.
Consistency of evidence with causal relationship	Identifying patterns that would be consistent with a causal relationship, and then seeking confirming and disconfirming evidence. It includes, among others: • Achievement of intermediate outcomes, • Checking results against expert predictions, • Checking timing of impacts, • Comparative case studies, • Checking consistency with existing literature, • Process tracing (developing alternative hypotheses and checking them), • Qualitative comparative analysis.
Ruling out alternatives	Identifying possible alternative causal explanations, and then seeking information to determine if these can be ruled out. Options include: • Process tracing, • Ruling out technical explanations, • Modelling, • Identifying possible explanations, their verification and possible ruling out.

Own elaboration based on: Rogers, P. (2014). Overview: Strategies for Causal Attribution, Methodological Briefs: Impact Evaluation 6, UNICEF Office of Research, Florence.