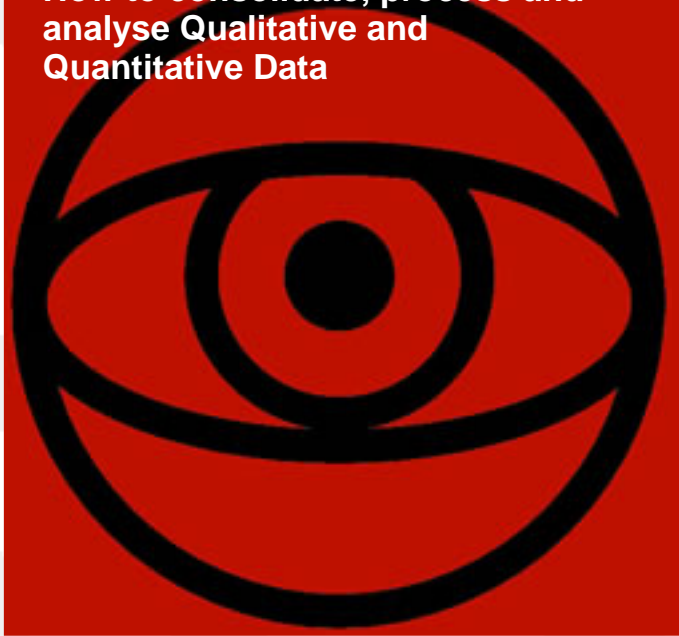


How to consolidate, process and analyse Qualitative and Quantitative Data



Monitoring & Evaluation Guidelines



United Nations World Food Programme
Office of Evaluation

- **How to consolidate and process Qualitative Data** 4
- **How to analyse Qualitative Data** 8
- **What do I need to know about Variables in order to conduct Data Analysis?** 11
- **How to consolidate and process Quantitative Data** 13
- **How to analyse Quantitative Data** 17

How to consolidate, process and analyse Qualitative and Quantitative Data

Overview

Introduction. The purpose of the module is to describe steps to take, following data collection, to prepare the raw data collected for the analysis stage. The module covers both qualitative data and quantitative data. New terms, especially those related to variables, are introduced and defined.

Why is this Module important?

This module is important because it provides a set of consecutive steps that can be followed for dealing with data collected from the field using a variety of data collection tools and instruments. It gives guidance on ensuring that the data is systematically consolidated and appropriately screened and checked for completeness and accuracy. The module also demonstrates that significantly different methods are employed for consolidating and processing qualitative and quantitative data although some aspects are overlapping. It is important to review these steps prior to embarking on data collection in order that sufficient human and cash resources are allocated to consolidation and processing tasks from the outset. Effective completion of these tasks is important as the quality of the data consolidation and processing work significantly affects the quality of subsequent analysis.

What does this Module aim to achieve?

This module has the following objectives:

- To illustrate how to consolidate and process qualitative data prior to analysis.
- To describe some basic techniques for analysing qualitative data.
- To define variable characteristics and the relationship between multiple variables during analysis.
- To describe how to consolidate and process quantitative data prior to analysis.
- To describe the basic principles for analysing quantitative data.

What should be reviewed before starting?

- Choosing Methods and Tools for Data Collection

How does this module apply to designing or implementing an M&E strategy for an operation?

This module gives staff with responsibilities in a range of M&E implementation tasks a quick review of the terms and procedures required to undertake the necessary steps to prepare raw data for analysis purposes.

Section Titles and Content Headings

- **How to consolidate and process Qualitative Data**
 - Introduction
 - Why is Data Consolidation and Processing necessary
 - Steps to follow for consolidating and processing Qualitative Data
 - An Example of Qualitative Data Consolidation and Processing from a Proportional

Piling Exercise concerning 'Sources of Food'

- **How to analyse Qualitative Data**
 - Introduction
 - What is Qualitative Data Analysis and who is involved
 - What are the Main Types of Data Analysis for Qualitative Data
 - Examples of 3 Types of Qualitative Data Analysis during and following Data Collection
- **What do I need to know about Variables in order to conduct Data Analysis?**
 - Introduction
 - What is a Variable
 - What are Variable Characteristics
 - Some Examples of Variable Characteristics
- **How to consolidate and process Quantitative Data**
 - Introduction
 - Why is Data Consolidation and Processing necessary
 - Steps to follow for consolidating and processing Quantitative Data
 - An Example of Quantitative Data Consolidation and Processing through the Application of the 6 Steps outlined above
- **How to analyse Quantitative Data**
 - Introduction
 - What is Quantitative Data Analysis
 - Guidelines for Analysing Quantitative Data
 - An Example of Quantitative Data Analysis Progression from Simple to Complex using Analysis Tables

How to consolidate and process Qualitative Data

Introduction. This section illustrates how to consolidate and process qualitative data prior to analysis.

Why is Data Consolidation and Processing necessary

As the critical first step, following data collection and prior to data analysis, raw qualitative data (e.g. interview notes, tables, charts, drawings, maps) must be processed and consolidated in order to be usable. The members of the team who conducted the interviews (WFP or implementing partner staff or consultants) process and consolidate the data. This will require some form of data cleaning, organising, and coding to so that the data is ready to be analysed and compared between different discussion groups or respondents. The degree to which it is done well can significantly affect the quality of subsequent analysis.

Steps to follow for consolidating and processing Qualitative Data

The following 5 steps provide general guidance on how to consolidate and process the majority of qualitative data. Depending on the methods used in data collection the 5 steps may need to be modified to suit the data processing needs.

Step 1: Summarise Key Points and Identify Quotations

Review data collection notes for each interview or discussion session. It is likely that the notes are in very rough form. Circle and note key discussion points and responses and consolidate long narratives into summary points. Also highlight key quotes that you may want to use in your presentation of the results and keep a list of quotations that might be used to illustrate important points made by discussion or interview participants.

Step 2: Organise Key Points in Topic Areas

For each group or individual interview or discussion session organise the key discussion points, responses, and summary points by topic. Topics discussed by more than 1 group or respondent can be compared between groups or individuals. These commonly occurring topics are identified and systematically listed. It is often useful to arrange the common topics in a simple spreadsheet having each discussion group serve as a row and each topic listed as a column. This will facilitate easy comparison between groups or respondents during analysis.

Step 3: Develop Codes describing Separate Categories of Similar Responses

For brevity, you will need to 'code' common topics for each group or individual into categories, giving 'like' responses or discussion points the same code. Codes can be figures or a system of words or symbols used to describe each separate category. Determine the number of categories for each topic by looking at the varying responses or discussion points from each group discussion or individual interview. Be careful not to dilute nuances and differences in responses or discussions. If you are in doubt give responses independent codes. Sub-codes can be used to capture nuances for responses or discussion points that are similar, but not exactly the same. The coding will assist greatly in making comparisons between groups and individuals during analysis.

Use the code category 'other' only for responses or discussion points that are very infrequent and where these 'outlying' or rare responses or discussion points are not important for sub-

sequent analysis.

Use the codes in your spreadsheet and be sure to provide a description of what each code means in a key or legend that accompanies the table.

Step 4: Labelling Products from Participatory Exercises

Products from participatory exercises used to stimulate discussion such as maps, diagrams, or rankings will not fit nicely into a spreadsheet. Each of these should be separated out from other data collection notes so that they may be compared for differences and similarities between groups. The use of note cards, clearly indicating in a label the group or individual from which the product came, can be helpful.

Step 5: Listing of Discussion Points on Unique Topics

Due to the open-ended nature of qualitative inquiry, topics brought up during the discussion or interview (e.g. those not pre-planned and turned into topics and coded categories in steps 2 and 3), should be listed as bullet summary points. Many of these may not be comparable between groups due to the fact that the issue may have been raised in 1 group, but not in another. However, it is critical to separate out these points prior to analysis as they may provide valuable insights into what makes 1 group or 1 individual different from others (e.g. issues of importance to them, unique context or circumstances).

An Example of Qualitative Data Consolidation and Processing from a Proportional Piling Exercise concerning ‘Sources of Food’

Step 1: summarise Key Points and Identify Quotations

This is a transcription of the original notes. The underlining and bold notes are added during Step 1.

Village A – Discussion Notes with Women’s Group

Explained proportional piling concept, several questions were asked about how much food each of the 10 stones represented. It was explained that all 10 stones represent all of the food that the household has and that each stone represents 1/10 of that total food. It was also explained that the *reference period was the last 3 months*.

First 5 stones were used to represent food from ‘**own production**’. There was much disagreement between members of the group with 1 respondent who depends primarily on wage labour insisting that food from own source is only worth 2 stones of their total 10 stones of food. The group concluded that *very few households depend on wage labour for any income (estimated at approximately 10 out of 100)*. Therefore *the group concluded that 5 stones accurately represents food from ‘own sources’* **50% of food from own production**.

The group explained that own sources is mainly **own production = small scale agriculture (sorghum and maize)**, but also *small ruminant keeping* (mainly goats). Although goats are rarely killed for meat, there *milk makes up an important part of the diet from ‘self-production’*.

The group then used 2 stones to represent food from purchase. When asked where they get *Sources of income money to purchase food*, responses included; *cash crops, remittances from relatives in the city, sale of baskets and other non-food products*. It was pointed out by 1 respondent that *bee keeping and the sale of honey is a fairly important source of income for many households* and others agreed. *Consensus was reached that 2 stones accurately represented food from purchase*. **20% of food is purchased**.

The final 3 stones represented food from WFP and the government in the form of food aid. When asked what the 3 stones represent when food aid is not given, respondents concluded that then they are 3 stones short of meeting their food needs. **30% of food comes from food**

aid. Respondents also noted that *sometimes gifts from richer households in the form of alms provided for some of their food needs, maybe 1 stone worth, but that these decreased when food aid was available.*

It was also noted that *over the last few years fewer households are able to provide gifts/alms because “everyone is getting poor now”* **key quotation. Periodically 10% of food is from gifts.**

Participants in the group stated that almost everyone had *fewer goats now compared to 5 years ago and they explained that panic selling at poor prices and the inability to buy at higher prices once drought had subsided led to a continual spiral of decreasing household livestock assets. This was an unprompted topic brought up by the participants in the discussion.*

Key quotations

A. “everyone is getting poor now” -- - explains why gifts from others are decreasing in the last few years.

Step 2: Organise Key Points in Topic Areas and

Step 3: Develop Codes describing Separate Categories of Similar Responses

Discussion Group	% of food from own production	% of food purchased	% of food from gifts	% of food from food aid	Crops as a source of own production	Type of Crops	Livestock as a source of own production	Type of livestock	Source of income for food purchase
Men in Village A	50%	20%	10%	20%	1	1, 2, 3	1	1	1, 2, 3
Women in Village A	60%	20%	0%	30%	1	1, 2, 3	1	1	1, 2
Men in Village B	40%	40%	10%	10%	2	n/a	1	1, 2	4
Women in Village B	40%	40%	0%	20%	2	n/a	1	1, 2	4

Key to Category Codes

Type of Crops		Type of Livestock		Crops as a source of own production		Livestock as a source of own production		Source of Income for food purchase	
1	Maize	1	goats/sheep	1	Yes	1	Yes	1	Cash crops
2	Sorghum	2	cattle	2	No	2	No	2	sale of non-food items
3	Vegetables	3	camels					3	beekeeping
4	Fruits	4	other					4	wage labor
5	Millet							5	Other
6	other								

Step 4: Labelling Products from Participatory Exercises

Seasonal Planting Calendar

Village A (medium agriculture potential zone)

Discussion Group with Women

Crop	jan	feb	mar	apr	may	jun	jul	aug	sep	oct	nov	dec
Maize						Harvest	Harvest					Harvest
Sorghum		Planting	Planting			Harvest	Harvest		Planting	Planting		Harvest
Millet						Harvest						
Tomatoes			Planting		Harvest	Harvest						
Chickpeas			Planting		Harvest	Harvest						

Village B (low agriculture potential zone)

Discussion Group with Men

Crop	jan	feb	mar	apr	may	jun	jul	aug	sep	oct	nov	dec
Maize							Harvest	Harvest				
Sorghum		Planting	Planting				Harvest	Harvest				
Millet							Harvest					

Key to tables Planting Harvest

Step 5: Listing of discussion points on unique topics

Village A – Discussion with Women: Key Points raised in Discussion (not universal between Groups)

- About 10 of 100 households depend to some degree on wage labour for income to purchase food.
- Goats are rarely killed for meat, but milk is an important food source.
- Droughts have induced panic selling and have led to selling goats at poor prices and the inability to buy back as many goats once droughts have subsided and prices have increased. The result is everyone has fewer goats now than they had 5 years ago.

How to analyse Qualitative Data

Introduction. This section describes some basic techniques for analysing qualitative data.

What is Qualitative Data Analysis and who is involved

Qualitative data analysis is the search for patterns and relationships in raw data. It also aims to collect explanations for those patterns and relationships. When using qualitative methods respondents or participants in group discussions should participate in the analysis and provide explanations and reasons based on their experience. This is referred to as self-analysis. During data collection the interviewer or group facilitator begins the process of self-analysis in the form of follow-up questions and requests for explanations. In this way, the raw data collected is already analysed to some extent and the relationship between data collection and data analysis is cyclical rather than linear.

The degree to which data is analysed by participants mirrors the degree of participation used in qualitative data collection methods. This self-analysis is extremely valuable because people naturally have more insights into their own lives and how various factors in their lives interact and are related, including causal relationships, than the interviewer or facilitator, who is usually not from the community.

In addition to people's own explanations for phenomena in their communities, households, and other aspects of their lives, the interviewer or facilitator can also bring a valuable 'outsiders' perspective to the topics raised. As an 'outsider', he or she is in a unique position to notice patterns and relationships between topics, both during data collection, in the form of on-the-spot analysis that informs questions asked, and after data collection when analysing the results.

The interviewer or facilitator often also brings the perspective of a wider experience, including interviews and discussions with other communities and individuals from both this and previous data collection exercises. The interviewer or facilitator may also be able to draw upon existing scientific knowledge about known causal relationships between phenomena that is available in related literature and research.

What are the Main Types of Data Analysis for Qualitative Data

Qualitative data analysis is somewhat free-flowing and subjective and comparisons between groups (or individuals) are difficult with open-ended discussion and interview data. It is difficult to distinguish between data collection and analysis as the 2 are intertwined and do not follow a distinct, linear sequence.

The following types of basic analyses can be used when analysing qualitative data, both during and following data collection:

1. **Descriptive** - Descriptive analysis is the summarisation of key topics and issues raised in discussion groups or individual interviews. The intent is to present these points as independent, un-interpreted descriptions of the condition, state, or circumstance that is stated by group discussion participants or respondents.
2. **Inferential** - Inferential analysis can be conducted by both the participants in the interview or discussion (self-analysis) or by the interviewer/facilitator. Inferential analysis looks at explanations as to why a condition, state, or circumstance exists and what might be the cause or causes.

The participants are well situated to provide explanations based on their own understand-

ing and perspective of their lives. Relationships, causal and otherwise, that are obvious to participants are often not readily perceivable by the interviewer or facilitator. Even where relationships are perceived by the interviewer or facilitator, giving the appearance of some degree of ignorance on his or her part may be useful for soliciting detailed descriptions or further analysis by the participants.

The interviewer as the 'outsider' provides a different view and may be able to notice patterns and relationships between factors that are not readily apparent to participants or respondents. This is especially true when comparisons are made between groups and individuals.

3. **Comparative** - Comparative analysis between groups or between respondents is often done by the interviewer or facilitator and can be either descriptive or inferential. However, in many cases discussion group participants or individual respondents may compare themselves to others. These differences or similarities provide important insights into how people perceive themselves in relation to others with whom they interact. >

Comparisons can be made between the: a) products of participatory exercises (proportional piling, mapping, ranking, scaling, etc.) or b) responses and discussions related to the various checklist items introduced by the facilitator or interviewer.

The unsolicited topics and discussion points that come up during a discussion or interview can also be used for comparative purposes (e.g. unsolicited topics raised in more than 1 interview or discussion, unique or 'outlying' issues raised). These topics are often useful for identifying shared and divergent priorities and concerns between groups and individuals.

Examples of 3 Types of Qualitative Data Analysis during and following Data Collection

Descriptive

Men's group village A - Almost every household has fewer livestock in comparison to 5 years ago.

Women's group village B - Although food aid is targeted, the commodities are re-distributed in the community so that each household gets an approximately equal share. If richer households want to then give to poorer households they can.

Inferential

Self-Analysis

Men's group village A - Households have fewer livestock because of the livestock trade ban in Saudi Arabia and panic selling of livestock at poor prices during droughts and the inability to re-purchase livestock once the drought has subsided and prices increase.

Women's group village B - Food is redistributed to all members of the community because it is a cultural norm that all households have an entitlement to free goods coming into the community. While this may appear to be negative to the outsider, women in the group stated that this is a critical cultural practice that existed long before food aid and will exist long after. Women also pointed out that existing social safety nets, in the form of alms or gifts from richer households, ensure that no household suffers when others have means.

Outsider Analysis

Men's group village A - In addition to the livestock ban, in a separate study of livestock markets traders indicated poor animal quality for livestock from this region as a reason for poor prices both during and following droughts. Poor animal quality appears to be a result of tsetse fly infestation and poor dry season fodder.

Women's group village B - Redistribution seems to suggest that further attempts to refine the

targeting process are not money and energy well spent. Given that redistribution is institutionalised, perhaps efforts should be made to work within this system rather than try and overcome it.

Comparative

While men in groups A and B both cited decreasing livestock numbers as a major sign and cause of food insecurity, the causes appear to be different. Men in group A, which is close to the border with Kenya, said that although the livestock ban has hurt prices they can still fetch good prices in Nairobi. The major problem they cite is transport to high demand markets. By contrast men in Group B, which is closer to the Somali port of Kismayo, do not normally trade and sell livestock in Kenya and are therefore more affected by the Saudi livestock ban. It also appears that the livestock in Group B's village are of poorer health and quality than the livestock in group A.

Interestingly, men in Group A raised planting crops as a potential source of food and income, though they have never done it before. The men in Group A did not raise the issue.

What do I need to know about Variables in order to conduct Data Analysis?

Introduction. This section defines key variable characteristics and the relationship between multiple variables during analysis of quantitative data. Some illustrative examples of variable characteristics are given. Some of the concepts are also applicable to qualitative data analysis.

What is a Variable

Data is a term given to raw facts or figures, which alone are of little value. Data represent something, like body weight, the name of a village, the age of a child, the temperature outside, etc. These can be anything from a date or number, to a name or event. The 'thing' that the data represents is called a variable. For example, the labels male and female are values for the variable sex or gender.

What are Variable Characteristics

The following important variable characteristics are used when monitoring and evaluating WFP operations and should be understood prior to analysing quantitative data.

- **Numeric** - Variables with numbers as values are called numeric.
- **Nominal** - Variables with names or labels as values are called nominal. Nominal variables are often converted into numeric variables through coding in both quantitative and qualitative analysis. This is especially true where computer software will be used to analyse the data, but is helpful in manual analysis as well.
- **Continuous** - Numeric variables are those that can have an infinite or a large number of values, including those for which numerous decimal places can be used.
- **Discrete or Categorical** - Discrete or categorical are those numeric or nominal variables that take a finite, limited, or small number of values. Note that a continuous variable can be transformed into categories and become a categorical variable. A categorical variable with 2 'categories' is known as a dichotomous variable.
- **Ordinal** - Ordinal are those numeric or nominal categorical variables for which there is an obvious and known order (e.g. low to high, light to heavy, young to old).

The next 2 definitions have to do with the relationship between variables when they are considered together during an analysis.

- **Dependent** - The variable under study in any analysis is known as the dependent variable. This is the variable of primary interest and is known as 'dependent' because it might be dependent on, or affected by, another variable that you've measured. This is what your analysis will be testing. While we are interested in these other variables, in this analysis we are primarily interested in them in terms of how they affect the dependent variable. They may become dependent variables in a separate analysis.
- **Independent** - The variable used to stratify the dependent variable during analysis for comparative purposes. This is the variable that the dependent variable might be dependent on, or be affected by.

Variables taking on the characteristics listed in the top of this section can be used as both dependent and independent variables. However using continuous variables as independent variables requires more advanced statistical knowledge. Continuous variables should be transformed into categorical variables before using them as an independent variable in an analysis.

Some Examples of Variable Characteristics

These examples highlight the fact that the variable types are not distinct from one another, but rather a variable may take on many of the variable characteristics.

- **Numeric** - weight, height, age, income, number of female children under 5, number of cattle.
- **Nominal** - village, sex or gender, tribe, source of water, type of livestock.
- **Continuous** - weight, height, income, age, distance.
- **Categorical** - age group, sex or gender, class in school, source of water, type of crops.
- **Ordinal** - age or age group, class in school, level of village leadership, height, weight.

In the following analysis example a child’s nutritional status is the dependent variable (e.g. the variable under study) and the sex or gender is the independent variable (e.g. the variable we are using to stratify the dependent variable). The dependent variable is the variable under study. We are interested in how the dependent variable nutritional status in this case, varies, is affected by, or is dependent on the independent variable, sex or gender (a dichotomous categorical variable). Here we are interested in how malnutrition is related to sex or gender.

% of Children Moderately and Severely Malnourished by Sex/Gender

	Male	Female	Total
% of children under 5 moderately and severely malnourished	19%	27%	46%

Note that in this example there was not an equal number of male and female children in the study such that the overall % of moderately and severely malnourished is not the mean of the female estimate and male estimate.

If we reverse the 2 variables and sex or gender becomes the dependent variable and child nutrition the independent, then the analysis changes slightly and would be stated as the % of malnourished that are males and the % of a malnourished that are female and the % of malnourished that are male. In this case we are only considering the malnourished and the total % of malnourished that are male plus the % of malnourished that are females would be equal to 100%. Here we are interested in how each sex or gender is affected by malnutrition.

	Male	Female	Total
Moderately or Severely Malnourished	65%	35%	100%

How to consolidate and process Quantitative Data

Introduction. This section describes how to consolidate and process quantitative data prior to analysis. An example is given that clarifies the application of each step outlined.

Why is Data Consolidation and Processing necessary

As a critical first step, following data collection and prior to data analysis, raw quantitative data from questionnaires (or other data collection instruments) must be processed and consolidated in order to be usable. This will require some form of data cleaning, organising, and coding so that the data is ready to be entered into a database or spreadsheet, analysed and compared.

Quantitative data is usually collected using a data collection instrument such as a questionnaire. The number of questionnaires or cases, is usually fairly large, especially where probability sampling strategies are used. Due to the nature of quantitative inquiry, most of the questions are closed ended and solicit short 'responses' from respondents that are easy to process and code. It is almost always necessary to use computer software to analyse the data due to this relatively large number of cases (in comparison to qualitative data) as well as variables (e.g. questions on the questionnaire). Microsoft excel and access provide basic spreadsheet and database functions, whereas more specialised statistical software such as SPSS and EpilInfo can be used where available and where expertise exists.

Ideally consolidation and processing is conducted by the team of interviewers who completed the data collection (WFP or implementing partner staff or consultants), however, in many cases additional staff are specifically tasked with the work of entering data into pre-formatted spreadsheets or databases. Data processing and consolidation needs to be well supervised and conducted as it can significantly affect the quality of subsequent analysis.

Steps to follow for consolidating and processing Quantitative Data

The following 6 steps outline the main tasks related to consolidating and processing quantitative data, prior to analysis.

Step 1: Nominate a Person and set a Procedure to ensure the Quality of Data Entry

When entering quantitative data into the database or spreadsheet, set up a quality check procedure such as having someone who is not entering data check every 10th case to make sure it was entered correctly.

Step 2: Entering Numeric Variables on Spreadsheets

Numeric variables should be entered into the spreadsheet or database with each variable on the questionnaire making up a column and each case or questionnaire making up a row. The type of 'case' will depend on the unit of study (e.g. individual, households, school, or other).

Step 3: Entering Continuous Variable Data on Spreadsheets

Enter raw numeric values for continuous variables (e.g. age, weight, height, anthropometric Z-scores, income). A new categorical variable can be created from the continuous variable later to assist in analysis. For 2 or more variables that will be combined to make a third variable, be sure and enter each separately. (For example, the number of children born and the number of children died should be entered as separate variables and the proportion of children who have

died could be created as a third variable). The intent is to ensure that the detail is not lost during data entry so that categories and variable calculations can be adjusted later if need be.

Step 4: Coding and Labelling Variables

Code categorical nominal variables numerically (e.g. give each option in the variable a number). Where the variable is ordinal (e.g. defining a thing's position in a series), be sure to order the codes in a logical sequence (e.g. 1 equals lowest and 5 equals the highest). In SPSS and some other software applications it is possible to give each numeric variable a value label (e.g. the nominal label that corresponds with the numeric code). For excel and other software that do not have this function, create a key for each nominal variable that lists the numeric codes and the corresponding nominal label.

Step 5: Dealing with a Missing Value

Be sure to enter 0 for cases in which the answer given is 0, do not leave the cell blank. A blank cell indicates a missing value (e.g. the respondent did not answer the question, the interviewer skipped the question by mistake, the question was not applicable to the respondent, or the answer was illegible). It is best practice to code missing values as 99, 999, or 9999. Make sure the number of 9's make the value an impossible value for the variable (e.g. for a variable that is 'number of cattle', use 9999 since 99 cattle may be a plausible number in some areas). It is important to code missing values so that they can be excluded during analysis on a case by case basis (e.g. by setting the missing value outside the range of plausible values you can selectively exclude it from analysis in any of the computer software packages described above).

Step 6: Data Cleaning Methods

Even with quality controls it will be necessary to 'clean the data', especially for large data sets with many variables and cases. This allows for obvious errors in data entry to be corrected as well as for excluding responses that simply do not make sense. (Note that the majority of these should be caught in data collection, but even the best quality control procedures miss some mistakes.) To clean the data run simple tests on each variable in the dataset. For example a variable denoting the sex or gender of the respondent (1 = male, 2 = female) should only take values 1 or 2. If a value such as 3 exists, then you know a data entry mistake has occurred. Also look for impossible values (outside the range of plausibility) such as a child weighing 100 kg, a mother being 10 years old, a mother being a male, etc.

An Example of Quantitative Data Consolidation and Processing through the Application of the 6 Steps outlined above

In this example, each household is the unit of study for the survey and is considered a case.

Step 1: Nominate a Person and set a Procedure to ensure the Quality of Data Entry

Every 4th case will be checked by a non-data entry person (WFP Field Monitor) to ensure quality in data entry.

Step 2: Entering Numeric Variables on Spreadsheets and

Step 3: Entering Continuous Variable Data on Spreadsheets

Q1: The estimated expenditure on food in the last 6 months

Responses on Questionnaires:

Case 1 \$30

Case 2 \$23

Case 3 \$112

Case 4 \$40

Q2: The estimated total expenditure in the last 3 months

Responses on Questionnaires:

Case 1 \$50

Case 2 \$35

Case 3 \$140

Case 4 \$35

Enter into database or spreadsheet and create a third variable that is food expenditure as a percentage of total expenditure.

	Food Expenditure in Last 3 Months	Total Expenditure in Last 3 months	Food Exp as a % of total Exp
Case 1	\$30.00	\$50.00	60.00%
Case 2	\$23.00	\$35.00	65.71%
Case 3	\$112.00	\$140.00	80.00%
Case 4	\$40.00	\$35.00	114.29%

Step 4: Coding and Labelling Variables

Code the nominal variables using numeric values. For ordinal variables make sure the order or sequence of numeric values makes sense.

Q3: Name of Village (with corresponding numeric code added)

Case 1 Hagadera = 1

Case 2 Hagadera = 1

Case 3 Kulan = 2

Case 4 Bardera = 3

Q4: Highest level of education completion of the head of household (with corresponding ordinal numeric codes that reflect least education to most)

Case 1 some primary, did not complete = 2

Case 2 no formal schooling = 1

Case 3 completed primary, some secondary = 4

Case 4 completed primary = 3

Enter into database:

	Food Expenditure in Last 3 Months	Total Expenditure in Last 3 months	Food Exp as a % of total Exp	Village	Highest ed. level completion by HofHH
Case 1	\$30.00	\$50.00	60.00%	1	2
Case 2	\$23.00	\$35.00	65.71%	1	1
Case 3	\$112.00	\$140.00	80.00%	2	4
Case 4	\$40.00	\$35.00	114.29%	3	3

Step 5: Dealing with a Missing Value

Coding missing values

Q5: Number of children under 5 in household

Case 1 = 2

Case 2 = 0

Case 3 = no answer given (missing value)

Case 4 = 3

Enter into database giving missing value a value of 99 (we use 99 because with multiple wives 9 children under 5 within a household is a possibility, even though it is a remote 1 for this area).

	Food Expenditure in Last 3 Months	Total Expenditure in Last 3 months	Food Exp as a % of total Exp	Village	Highest ed. level completion by HofHH	Number of children U5 in HH
Case 1	\$30.00	\$50.00	60.00%	1	2	2
Case 2	\$23.00	\$35.00	65.71%	1	1	0
Case 3	\$112.00	\$140.00	80.00%	2	4	99
Case 4	\$40.00	\$35.00	114.29%	3	3	3

Step 6: Data Cleaning Methods

Run data validity checks to ‘clean the data’. Try to find impossible values for each variable. If they are found and reverting to the questionnaire does not clarify the mistake, then set the value to missing (step 5).

In this case the third variable in case 4 (refer to the table under step 5) suggests either an entry error or a mistake on the questionnaire. Food cannot be 114% of total expenditure since food is a portion of expenditure and the maximum value it could take is 100% (food expenditure represents all expenditure).

After reverting to the questionnaire, it is confirmed that data was entered correctly and that the error lies in the respondent’s understanding of the question or in the interviewer’s recording of the response. It is decided that the best course of action is to set variables 1, 2, and 3 for Case 4 to ‘missing’ so that the analysis is not misleading.

	Food Expenditure in Last 3 Months	Total Expenditure in Last 3 months	Food Exp as a % of total Exp	Village	Highest ed. level completion by HofHH	Number of children U5 in HH
Case 1	\$30.00	\$50.00	60.00%	1	2	2
Case 2	\$23.00	\$35.00	65.71%	1	1	0
Case 3	\$112.00	\$140.00	80.00%	2	4	99
Case 4	9999	9999	999	3	3	3

How to analyse Quantitative Data

Introduction. This section describes some basic principles for analysing quantitative data. It is particularly useful for readers who have some professional training in statistics.

What is Quantitative Data Analysis

Data analysis is the search for patterns in raw data and for explanations relating to those patterns. Quantitative data analysis can be broken into 2 categories: descriptive analysis and inferential analysis.

Descriptive analysis is simply the presentation of numeric results for the study population. Two examples of descriptive analysis are: 28% of children under 5 have diarrhoea and the mean income in the study area is \$213 per year. Descriptive analysis also encompasses the presentation of results for various sub-groups within a study. An example is as follows: the mean income in District A is \$289 per year, whereas the mean income in District B is \$344 per year.

By contrast inferential analysis seeks to establish relationships between variables that may explain why differences exist. If descriptive analysis tells us that the rate of malnutrition is higher in 1 district than in another, inferential analysis would look for relationships with other variables to help explain this difference.

For both descriptive and inferential analysis, differences between sub-groups (descriptive) and associations between variables (inferential) can be analysed statistically if a probability sampling strategy has been used. This type of data analysis draws on statistical theory to support or reject associations between variables.

For quantitative data analysis from non-probability studies the same type of descriptive and inferential analyses apply, though conclusions about differences and associations are not statistically supported.

Guidelines for Analysing Quantitative Data

The following progression from simple to more complex analysis should be used for each variable being analysed. The variable under investigation is known as the dependent variable. Other variables by which the variable is being compared (or stratified) are known as independent variables. In many cases a variable that serves as an independent variable for 1 analysis becomes the dependent variable in another analysis and vice versus.

A.

Simple descriptive

This is the descriptive presentation of a variable for the study population. The variable is not disaggregated in this simplest form of analysis. In this type of analysis percentages are calculated for dichotomous and categorical variables, and means and medians are calculated for continuous variables.

B.

Stratified Descriptive – Geographic and Gender Disaggregated Analysis

This is the descriptive comparison of a variable between sub-groups in the study population defined by 1 or more other variables (if more than 2 seek expert guidance for regression analysis). Geographic comparisons (e.g. by districts or another geographic aggregate such as village or province) are commonly used, as is gender. This type of analysis may also be referred to as disaggregated analysis. In stratified descriptive analysis create sub-groups

for which you hypothesize differences that are likely to exist for your variable of interest. The intent at this point is to describe differences between major sub-groups, not to infer why these differences exist.

For sub-groups defined by 1 independent variable use a table to compare the dependent variable between sub-groups. The number of rows in the table will vary depending on whether the dependent variable is continuous (1 row with mean), categorical (number of rows depends on number of categories), or dichotomous (2 rows with percentage). The number of columns in the table will depend on whether the independent variable is categorical (number of columns depends on number of categories) or dichotomous (2 columns with percentage). If you wish to do so you must first convert the continuous variable into a dichotomous or categorical variable.

Also use a table for sub-groups defined by 2 independent variables. One independent variable forms the columns and the other the rows. The content of the cells is the dependent variable in the form of a mean, median, or percentage.

C.

Inferential Analysis

Inferential analysis moves beyond description and attempts to look for associations between variables to explain differences and variations between groups. The analysis follows the same procedure as for stratified descriptive analysis, except that the intent is to identify not only difference, but to infer why they exist either through known causal relationships (e.g. water source and diarrhoea) or through testing relationships with other independent variables. Choosing which variables to use as independent variables should be based on hypotheses about potential factors related, causal or otherwise, to the dependent variable.

Use tables to conduct the analysis (either manually or using a software package such as excel or SPSS). For analyses with 1 independent variable and 1 dependent variable, use rows to express the dependent variable and columns to express the independent variable. For analyses with 2 independent variables, use rows for 1, columns for the other and put the dependent variable in the cells of the table. If you wish to do so you must first convert the continuous variable into a dichotomous or categorical variable. For analyses with more than 2 independent variables or using continuous variables as independent variables seek expert guidance in regression analysis.

Although inferential analysis attempts to explain differences and variation, it is important to note that causality should not be concluded. When variables are found to be in association and a known causal relationship exists (e.g. physical causal relationships are well grounded in the literature relevant to the operation's activities) some degree of causation can be inferred, but not confirmed. For associations between 2 variables, even those with statistically valid associations, the conclusion should be that 1 is a potential cause of the other, recognizing that the direction of cause and the proof of causality are often difficult to determine. This is important because outside of clinical trials too many confounding factors (e.g. factors that have a relationship with both the dependent and independent variables) exist and could potentially explain the association found.

Take care in interpreting differences: It is also important to realise that when a sample (probability or non-probability) is being used to generalise about the larger population that the sample results are estimates, more accurately reflected by a range of values, and not exact figures. Therefore, use extreme care when concluding differences between groups, especially where the sample size is small.

If a probability sample is used, we can statistically test to see whether the difference between groups is statistically significant (meaning the estimates for each sub-group, expressed as a range of values, do not overlap at 95% confidence). Seek expert guidance in statistical analysis and interpretation of quantitative data gathered using probability sampling techniques.

For non-probability samples concluding differences is more subjective and therefore should only be inferred. In general, do not infer that a difference exists if the means, medians, or percentages are reasonably close to 1 another. It is difficult to give guidance on 'reasonably close' because this will vary with sample size, variables, and variation within the sample population.

An Example of Quantitative Data Analysis Progression from Simple to Complex using Analysis Tables

Simple Descriptive

- % of children under 5 in the study area have diarrhoea
- The mean household income for the study area is \$312
- The median age of mothers in years at 'first birth' is 17
- 33% of children under 5 are moderately or severely malnourished = wasted

Stratified Descriptive

With 1 independent variable

Table A – Moderate and severe malnutrition among children under 5 (dependent variable) by district (independent variable)

	District A	District B
% of children moderately and severely malnourished	32%	26%

With 2 independent variables

Table B – Moderate and severe malnutrition among children under 5 (dependent variable expressed in each cell) by district and gender (independent variables as columns and rows respectively).

	District A	District B
Male	33%	18%
Female	31%	33%

Although a statistical conclusion about the association between malnutrition and district/gender cannot be concluded due to the small sample size (and non-probability sampling) it is noted that the potential difference in malnutrition prevalence by gender in District B that does not exist in District A is very striking and requires further investigation.

Inferential Analysis

With 1 independent variable

It is thought that a potential relationship exists between the high prevalence of diarrhoea and the high prevalence of moderate and severe malnutrition, noting that the literature suggests a mutually causal, feedback relationship between the 2.

Table C – Moderate and severe malnutrition among children under 5 by whether or not child has had diarrhoea in the last 2 weeks.

	Children with diarrhoea	Children without diarrhoea
% of children moderately and severely malnourished	44%	21%

Although it is interesting to note the much higher prevalence of malnutrition among children who have had diarrhoea in the last 2 weeks, it is too early to make any conclusions. We must first examine other factors that may explain the relationship between diarrhoea and malnutrition

(e.g. are related to both variables independently = confounding factors).

We next look at age groups and moderate and severe malnutrition

	0	1	2	3	4	5
% of children moderately and severely malnourished	55%	46%	30%	30%	21%	12%

Although we cannot statistically conclude the relationship due to our small sample size and the inability to apply statistical tests, the linear relationship (e.g. decreasing malnutrition with increasing age) suggests that a causal relationship is possible.

With 2 independent variables

It is hypothesised that the true cause of high levels of moderate and severe malnutrition is the lack of breastfeeding and that this also has an independent relationship with diarrhoea as well as age (e.g. breastfeeding is only applicable to children under 2 and has both high energy density and properties that protect against disease). This suggests that perhaps it is not diarrhoea that has a relationship with malnutrition, but rather breastfeeding. This hypothesis is tested using the following table only looking at the high malnutrition age groups of children under 2.

Table D - Prevalence of moderate and severe malnutrition among children under 2 by whether or not the child has had diarrhoea in the last 2 weeks and whether or not the child is being breastfed (malnutrition prevalence in cells)

	Breastfed	Not breastfed
Diarrhoea in last 2 weeks	12%	65%
No Diarrhoea in last 2 weeks	11%	63%

Although the sample size is too small to conclude a statistically valid association, it is clear by these results that the relationship between diarrhoea and malnutrition disappears when breastfeeding is controlled for (malnutrition prevalence is the same for each 'breastfeeding group' regardless of diarrhoea prevalence). Therefore it is likely that the causal relationship is between breastfeeding and malnutrition as hypothesised. This also means there is a relationship between diarrhoea and breastfeeding with children who are not breastfed having a higher prevalence of diarrhoea. A simple 1 row by 2 column (with diarrhoea as the dependent variable or row and breastfeeding as the independent variable or column) table can be used to estimate this relationship

Module Summary

What has been covered in this module?

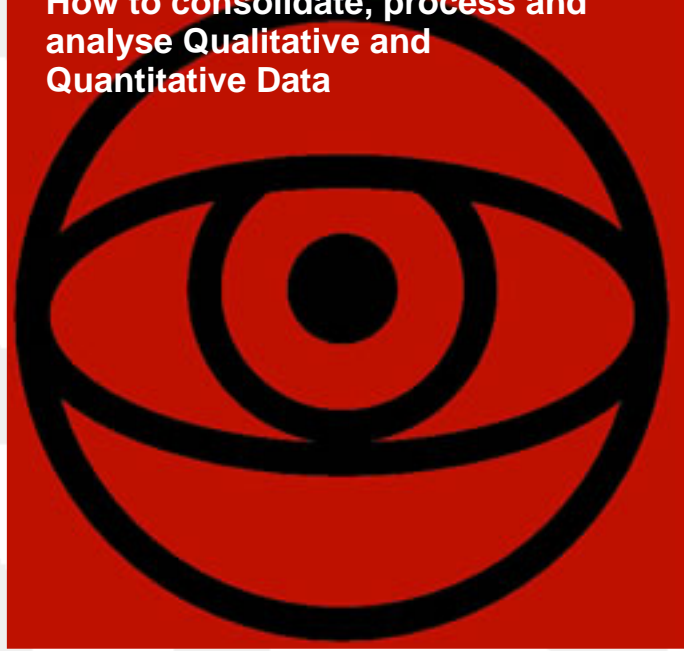
This module has described the steps to take, following data collection, to prepare the raw data collected for the analysis stage. It gave guidance on ensuring that the data is systematically consolidated and appropriately screened and checked for completeness and accuracy. The module also demonstrated that significantly different methods are employed for consolidating and processing qualitative and quantitative data although some aspects are overlapping. It is important to review these steps prior to embarking on data collection in order that sufficient human and cash resources are allocated to consolidation and processing tasks from the outset. Effective completion of these tasks is important as the quality of the data consolidation and processing work significantly affects the quality of subsequent analysis.

What additional resources are available?

For further information the following modules and resources might be useful:

- How to design a Results-Oriented M&E Strategy for EMOPs and PRROs
- How to design a Results-Oriented M&E Strategy for Development Programmes
- Choosing Methods and Tools for Data Collection
- Going to the Field to collect Monitoring and Evaluation Data
- How to plan a Baseline Study
- How to Plan an Evaluation
- How to plan and undertake a Self-evaluation
- Reporting on M&E Data and Information for EMOPs and PRROs
- WFP Participation Tool Kit

**How to consolidate, process and
analyse Qualitative and
Quantitative Data**



United Nations
World Food Programme
Office of Evaluation and Monitoring

Via Cesare Giulio Viola, 68/70 - 00148
Rome, Italy

Web Site: www.wfp.org
E-mail: wfpinfo@wfp.org
Tel: +39 06 65131