Rapid Guide: Disaggregating Quantitative Data

Data disaggregation means breaking down data into detailed sub-categories, allowing you to **get a better understanding of**:
- the **needs** of specific groups of people (e.g. women, youth, persons with disabilities)
- the extent to which different groups of people (do not) **participate** in the project activities
- the extent to which they (do not) **benefit** from the project activities

As a result of using such data, the assistance your project provides can become more effective and inclusive. This brief guide was prepared to make it easier for you to disaggregate data by various factors that often influence people's vulnerability, such as age, gender, disability and wealth.

When interpreting disaggregated data, it is important to **keep in mind that there is a difference between correlation and causation**. For example, if a survey shows that children from households with a vegetable garden have more diverse diets (correlation), you should be careful about claiming that 'vegetable gardens improve dietary diversity' (causation). It is possible that vegetable gardens are kept primarily by better-off households (who have land, can afford inputs, etc.) and that the key factor is not the ownership of a vegetable garden but a household's wealth or some other factor. In such a case, you would have to compare households from the same wealth category. Always remember: **the fact that there is a correlation between two variables does not mean that there is a cause-and-effect relationship**. Consider which factors might influence your data and try to manage these during the data analysis process.[1]

Remember: <u>what matters the most is that disaggregated data is used</u> to adopt measures that enable different groups of people to equally benefit from the project and have equitable participation in its activities. Therefore, whenever you disaggregate data by any of the vulnerability factors, also report on what measures were taken as a result of the findings.

Click on a relevant link to **quickly access the guidance you need**:

Disaggregate by **AGE**	Disaggregate by **GENDER**	Disaggregate by **DISABILITY**
Disaggregate by **WEALTH**	Disaggregate by **ACCESS TO SERVICES**	Disaggregate by **OTHER CRITERIA**

Disaggregating Data by Age

WHY? The extent to which people are vulnerable to certain risks as well as the extent to which they participate in and benefit from the provided assistance is often influenced by their age. Understanding these age-specific differences can help you sharpen the focus of your intervention, making it more inclusive and increasing its overall impact. Let us take a few examples:
- disaggregating data on people's knowledge / practices / or awareness of certain services by age tells you who your project should focus on through its promotion activities
- disaggregating data on children's nutritional status (or their dietary diversity) by age can show you when they are most vulnerable and where your attention is needed the most
- disaggregating data on people's perceptions of how accessible the provided assistance was, helps you understand whether some groups of people (e.g. the elderly) experienced worse access than others

HOW? Different donors have different requirements on which age groups your reporting should use. For example, while ECHO uses age the groups 'infants and young children (0-59 months)', 'children (5-17 years)', 'adults (18-49 years)' and 'older persons (>50 years)'; USAID BHA requires organisations to use different age

---

[1] If you need strong evidence of a cause and effect relationship, it is necessary to design both the intervention and its research / M&E system in a way that allows you to gain such data (i.e. it is not enough to just disaggregate data by various variables). The most robust way of doing so is conducting a randomized controlled trial (RCT), followed by longitudinal design and different variants of quasi-experimental design.

groups. The Sphere Standards recommend completely different age groups. Therefore, the easiest way is to **record respondents' ages in completed years** (i.e. the age at last birthday), as you can then fit it into whichever age groups you need. If you know that the respondents might be reluctant to say their exact age, then you will have to use the more general age groups required by your donor.

Be aware that some indicators (especially on child health and nutrition) require a very accurate age – i.e. not only years but also months. Since people are sometimes not sure about the exact dates of their children's birth, **the enumerators need to verify the age**. This can be done by reviewing the child's birth certificate, vaccination card or other document. If such documents are not available or might have unreliable birth dates, the enumerators should assess the child's age using a **local events calendar**. Read [FAO's Guidelines for Estimating the Month and Year of Birth of Young Children](#) to learn how to prepare the calendars and how to train enumerators in using them correctly. When using local event calendars, make sure that:

- Each month of the calendar includes at least one and preferably two events.
- The calendar includes not only national holidays but also any events that have taken place at the local level and which people might remember.
- The calendar includes a combination of recurring events (e.g. holidays) and less frequent or one-off events (e.g. election, major weather events).
- The calendar is prepared in collaboration with local informants from the target communities (e.g. teachers, health staff, leaders).
- If older children or adults aren't sure of their age, use a calendar including one event for each year (it will not be completely precise but is better than having completely incorrect data).

Disaggregating Data by Gender

WHY? Gender is a social construct built through cultural, political and social practices that defines the roles of women, girls, men and boys, as well as the social definitions of what it means to be masculine and feminine. It often defines the duties and responsibilities expected of women, girls, men and boys and sets some of the barriers they may face or opportunities and privileges they may enjoy. It often determines the power, which they have, and their ability to access and control resources.[2] It is different from 'sex', which refers to the biological differences between males and females. Disaggregating data by gender matters for multiple reasons:

- It helps you **refine your targeting**: For example, during the 2011 cholera outbreak in Haiti, data disaggregated by gender showed that more men than women were dying of the disease. Follow-up research showed that this was due to men being more reluctant to seek medical care. Aid agencies therefore developed specific health messages for men, which led to a decrease in their mortality rates.[3]
- It enables you to better **meet the specific needs and perceptions** of different genders: For example, girls having different needs and perceptions regarding school WASH facilities than boys do.
- It allows you to assess the **extent to which different genders participate** in the project activities: For example, men's participation in child nutrition sessions.
- It helps you to **understand the extent to which different genders benefit** from the provided assistance: For example, differences in the income earned by male and female entrepreneurs.

HOW? In the case of **self-administered questionnaires**, you can just ask people *"What is your gender?"* while offering the following options: *'male', 'female', 'non-binary'* and *'no response'*. In the case of a **questionnaire administered by an enumerator**, things get more complicated: In some contexts, sitting in front of someone and asking her/him "*What is your gender?"* might be uncomfortable for the enumerator and strange (or even offensive) for the respondent (though this depends on how the question sounds in the local language). Therefore, when introducing the interview, explain to the respondent that some of the questions might seem to have obvious answers but since they need to be answered directly by the respondent, you are kindly asking for understanding. The enumerator can then ask people "*What is your gender?"*. The main answer options included in the questionnaire (but **not read aloud**) then should be '*male', 'female', 'non-binary'* and *'no response'*.

---

[2] IASC (2018) The Gender Handbook for Humanitarian Action
[3] ECHO (2013) Gender-Age Marker Toolkit

Disaggregating Data by Disability

WHY? An estimated 15% of the global population is persons with disabilities. Compared to persons without disabilities, they are more likely to be excluded, experience discrimination, face increased barriers to accessing services and participating meaningfully in the society. Collecting data on persons with disabilities enables you to identify the prevalence of persons with disabilities, to understand their needs and to monitor the extent to which they are accessing, participating in and benefiting from your project activities. As a result, you can take measures that will make the project more inclusive, increasing its impact.

HOW? The most widely tested tool to generate comparable data about persons with disabilities is the Washington Group (WG) sets of questions.[4] While they can be used to disaggregate data by disability type, they should not be used for the identification of particular health conditions.

The short set of questions (see below) were developed to identify people (aged 5 years and older)[5] who have difficulties in functioning in six basic universal actions. The questions were designed for **individual surveys** that can give you detailed information about persons who have such difficulties.[6]

> **Q1:** *"Do you have difficulty seeing, even if wearing glasses?"*
>
> **Q2:** *"Do you have difficulty walking or climbing steps?"*
>
> **Q3:** *"Do you have difficulty with self-care such as washing all over or dressing?"*
>
> **Q4:** *"Do you have difficulty hearing, even if using a hearing aid?"*
>
> **Q5:** *"Do you have difficulty remembering or concentrating?"*
>
> **Q6:** *"Using your usual language, do you have difficulty communicating, for example understanding or being understood?"*

For all questions, the **answer options** are: *'No – no difficulty'; 'Yes – some difficulty'; 'Yes – a lot of difficulty', 'Cannot do at all'* and *'Refused to answer'* (all based on people's perceptions).

When assessing disability during **household surveys**, it is necessary to collect data about all the household members. It is recommended to use the following approach:

1) At the beginning of the questionnaire, the enumerator records the total number of household members aged 5 years or more (i.e. those whose situation will be assessed).

2) When administering the questionnaire, the enumerator asks whether all the household members who are at home could come to the same room for a few minutes (except children under 5 years and those who cannot respond to the questions due to their disability).

3) When everyone is gathered, the enumerator explains in front of everyone: *"In the next questions I will ask about difficulties you or other household members, who are absent, may have doing certain activities because of a health problem. These questions do not concern children younger than 5 years. When answering the questions, please tell me if some of you or any absent household members have such a difficulty."* The enumerator then asks the household members who are present (together, not individually) the following amended version of the WG questions. Their advantage is that they 1) account for the possibility of some household members not being at home; and 2) the information about various members can be provided not only by the head of household but also by other present members.

> **Q1:** *"Do any of you or other household members have difficulty seeing, even if wearing glasses?"*
>
> **Q2:** *"Do any of you or other household members have difficulty walking or climbing steps?"*
>
> **Q3:** *"Do any of you or other household members have difficulty with self-care such as washing all over or dressing?"*
>
> **Q4:** *"Do any of you or other household members have difficulty hearing, even if using a hearing aid?"*
>
> **Q5:** *"Do any of you or other household members have difficulty remembering or concentrating?"*
>
> **Q6:** *"Using your usual language, do any of you or other household members have difficulty communicating, for example understanding or being understood?"*

---

[4] Read more about WG question at www.washingtongroup-disability.com/about/

[5] The questions listed above were not designed to assess disabilities among young children. If you need to assess the prevalence of disabilities among children under five, use the Child Functioning Module available at this website.

[6] The questions listed in the box above are part of the Washington Group Short Set on Functioning (WG-SS). For others sets of questions that focus on specific groups (e.g. children, labour force) or that collect more detailed data, please refer to this website.

4) If any member of the household has such a difficulty, the enumerator needs to ask:
- who has the difficulty (gender, age, other data)[7]; and
- the degree of difficulty (as listed above – some difficulty, a lot of difficulty, cannot do at all)

For all questions, there needs to be an answer option 'refused to answer'. An example of such a questionnaire is included here: Enketo version; XLSForm version.

An alternative approach is to interview the head of households only, asking them about their situation and the situation of their household members. This approach is easier but is less precise, as the person answers on behalf of other household members. **Additional options** are presented in HI's factsheet; however, they do not deal with the (very likely) situation where some household members are absent.

During the **data analysis**, 'a person with disability' can be considered anyone for whom the option *'a lot of difficulty'* or *'cannot do at all'* applied (for any of the six questions). Disaggregating data by disability then means breaking down (some of) the overall survey data into:
- In the case of **individual level surveys**: Data that concern persons with (any type of) disability and data that concern persons without any disability.
- In the case of **household level surveys**: Data that concern households with one or more members having (any type of) disability and data that concern households without any members with a disability.

When using the questions, **take advantage of the following tips**:
- Do not try to simplify the work by asking the head of the household whether there are any persons with disabilities living in the household. Such an approach is less precise than asking people directly.
- Ask enumerators to avoid using the word "disability" (use "difficulty"). It can create stigma and induce bias. Furthermore, ensure that the enumerators ask the questions exactly as they are written, never skip a question or guess the answer (for example, based on observations or assumptions). For example, a person sitting in a wheelchair does not necessarily mean that s/he cannot walk. Make sure that all enumerators are familiar with the additional guidance summarized in this two-pager.
- The enumerators must be trained in concluding correctly about the extent of difficulties people experience (some / a lot …). Use the guidance provided in the 'Training Pack' at HI's website.
- Translations of the Washington Group Questions into different languages are available at this website. If you need to use them during a household survey, amend them slightly as shown above.
- If you cannot include all six questions in the questionnaire, it is possible (though not recommended) to skip Q3 and Q6. This will miss some people but testing suggests they are likely to have already been identified in one of the other four questions.[8]

Disaggregating Data by Wealth

WHY? Disaggregating data by the respondent households' wealth helps you understand the extent to which people from poorer as well as slightly wealthier **households participate in** and **benefit from the project activities** (including any services it promotes). For example, a survey might show that 65% of respondents have accessed the promoted agricultural extension services, which might seem to be a good result. However, by disaggregating the data by wealth, you might find that only 39% of the poorest respondents have used the services as opposed to 82% of the most affluent respondents. Such a result would put the data in a different perspective, requiring the organisation to work towards ensuring much better equity.

HOW? To disaggregate data by the respondent households' wealth, you first need to know which wealth categories the survey respondents (respectively, their households) belong to. IndiKit does not recommend determining people's wealth by asking them about their **income**, as many people 1) experience significant seasonal differences in the income they generate; and 2) are reluctant to report openly the amount of money they earn. Furthermore, people who have a higher income but have to pay rent, buy most of their food, etc. might be less wealthy than those who earn less but have much lower financial expenditures.

---

[7] If you need more data about the person who has a disability (e.g. whether s/he is a head of household), include an additional question.

[8] HI, E-learning Collecting Data for the Inclusion of Persons with Disabilities in Humanitarian Action, Module 3

A more reliable method to determine the wealth category is to ask people about those aspects of their **household's characteristics and asset ownership** that are closely correlated with (not) being poor. For example, *"How many members live in the household?"* or *"What is the main source of cooking fuel in your household?"*. These questions must be specific to the given country and must be proven to indicate reliably the respondent household's wealth (i.e. you should **never try to design the questions on your own** – it will not be reliable). The EquityTool and the Poverty Probability Index (PPI) tools offer sets of questions for over 60 countries (these sets were tested and you **cannot change them**). Each set consists of 10-14 simple questions that take only very little time to answer. Their main function is to compare the wealth of the respondents' households to the wealth of the national or urban population, enabling you to see the extent to which you are reaching poorer households.

However, they can also be used to **divide the respondents into several categories according to the wealth of their households**. This allows you to disaggregate data by wealth, using one of these approaches:

1) Comparing the respondents to the **wealth of the national population**. For this, respondents are divided into national wealth quintiles based on whether they belong to the poorest 20% of a country's population, second poorest 20%, third poorest 20%, and so on. You can then disaggregate the data based on the national wealth quintile a respondent's household belongs to. For example, you can assess the dietary diversity of children from households that fall in the poorest quintile, second poorest quintile, etc. The disadvantage of this approach is that if your programme targets poor households, there will be many respondents in the poorest quintile and fewer respondents in the better-off quintiles. As a result, the data for the better-off quintiles will have a higher margin of error. Guidance on assessing respondents' wealth quintiles is provided on the PPI and EquityTool websites.

2) Comparing the respondents to the **wealth of other respondents that were interviewed** during the same survey. Here you can divide the respondents into categories with an equal number of respondents (e.g. thirds) and then compare the data for respondents from the lowest category (e.g. the lowest third, the poorest) with the data for the respondents from the highest category (e.g. the top third, the relatively better-off). For example, you can compare the dietary diversity of children from the poorest and from relatively better-off households. Using categories with the same number of respondents (e.g. thirds) ensures that the disaggregated data has a very similar margin of error. The disadvantage of this approach is that it might create a misleading impression that people from the highest category (e.g. the top third) are wealthy, even though they are relatively poor. This is especially the case if your programme targets poorer households. When presenting the disaggregated data, if you clearly acknowledge that even people from the top third can be relatively poor, then using this approach is **IndiKit's recommended way of disaggregating data by wealth**.

To use this approach, you need to divide respondents into several wealth categories. This can be done in the following way:

- **EquityTool:** Follow the provided guidance until you have calculated a 'total national score' for each respondent (see point 4 in the guidance). As the next step, divide the respondents into several categories (e.g. thirds) according to the value of their total national scores. Each category must include the same number of respondents. For example, if the survey used a sample of 300 respondents, one third would contain 100 respondents with the lowest national scores (= the poorest), another third would contain 100 respondents with the highest national scores (= the relatively better-off) and the remaining third would contain the other 100 respondents that are in-between. This means that the respondents will not be compared to the wealth of the national population but to the wealth of other respondents (the same applies to PPI below).

- **PPI**: On PPI's website, select the country where you plan to collect the data. Download the related documents and guidance from the relevant country page. In the Excel document 'PPI Scorecard and Look-Up Tables', click on 'How to use' sheet. Follow Step 1 and Step 2 in the guidance, until you calculate the 'PPI Scores' of all the survey respondents. As the next step, divide the respondents into several categories (e.g. thirds) according to the value of their PPI scores. The same example as above applies.

When disaggregating the data by wealth, you should then **compare the data for respondents from the lowest category** (e.g. the lowest third, the poorest) **with the data for the respondents from the highest category** (e.g. the top third, the relatively better-off). Ensure that the categories you use include a sufficient number of respondents, so that the disaggregated data has an acceptable margin of error. For example, using thirds might be safer than using quarters or quintiles.

Disaggregating Data by Access to Services

WHY? The extent to which people access the promoted services (health care, agronomic training, saving groups, etc.) often influences how much they benefit from a project. Being able to see the relationship between people's use of services and the desired results (of these services and of the project in general) can provide useful insights into their effectiveness.

HOW? You need two key types of data:

- Data on people's use of the promoted services, focusing not only on whether they used them but also how many times they have used them (if the frequency of use matters) or why they did not use them.
- Data on the results the services (or the project in general) were supposed to achieve, such as increased use of promoted agronomic practices.

Disaggregating Data by Other Criteria

There are dozens of additional factors according to which you can disaggregate data, better understand the results of your work and get ideas for ways forward. At the same time, each new disaggregation requirement brings extra data analysis and reporting work, so it is important that you **focus only on what you need the most**. Before you design a survey, discuss with your colleagues what disaggregated data would help in your work (i.e. do not automatically disaggregate all the collected data by the same factors). Some of the additional data disaggregation options include:

- **Disaggregating Data by Participation:** The level of people's participation in the project activities often impacts on the benefits they receive. For example, women who regularly attend mother-to-mother support group meetings might be more likely to adopt the promoted practices. Understanding the extent to which (lack of) participation influences the desired results can help you sharpen the focus of your activities. Measure not only frequency but also duration of people's participation in the project activities.

- **Disaggregating Data by Location:** The geographical area where people live often influences their access to services and other outcomes that projects aim to achieve. Therefore, consider disaggregating your data by meaningful geographic factors, such as distance from a larger market / health facility / school (in minutes – not kilometres); urban vs. rural areas; camp vs. non-camp; and districts.

- **Disaggregating Data by Ethnicity:** In many contexts, ethnicity is a sensitive topic, mainly due to people being discriminated against based on their ethnicity. Always consider 1) how sensitive it is to ask about ethnicity; and 2) whether your project must have such data. Furthermore, keep in mind that:
  - often there is no shared definition of ethnicity and there are many factors influencing how a respondent will respond to a question about her/his ethnicity
  - if only a small portion of respondents self-identify themselves as belonging to a certain ethnicity, due to the small number of respondents, the data might have an unacceptably high margin of error
  - ethnicity (along with health and religion) comes under General Data Protection Regulation's (GDPR) definition of special category data and can be collected and processed only following an organisation's rules on data protection and consent

- **Disaggregating data by displacement status**, including internally displaced people (IDP), refugees, residents, and returnees. Consider also unaccompanied minors.

- **Other factors include:** number of household members / children; single headed households; marital status (single, married, separated, widowed, no response); pregnant and lactating women; minority groups (linguistic, religious); nationality; caste; language; education level; economic activity; etc.

- **Do you have a suggestion for improving this Rapid Guide's content? Send it to us please!**
- **Would you like this Rapid Guide to be available in a different language? Get in touch with us!**

Published by: **People in Need (2021)** | Author: Petr Schmied | indikit@peopleinneed.cz | www.indikit.net