



Tips for Collecting, Reviewing, and Analyzing Secondary Data

[email Egypt docs.doc](#)

WHAT IS SECONDARY DATA REVIEW AND ANALYSIS?

Secondary data analysis can be literally defined as “second-hand” analysis. It is the analysis of data or information that was either gathered by someone else (e.g., researchers, institutions, other NGOs, etc.) or for some other purpose than the one currently being considered, or often a combination of the two (Cnossen 1997).

If secondary research and data analysis is undertaken with care and diligence, it can provide a cost-effective way of gaining a broad understanding of research questions.

Secondary data are also helpful in designing subsequent primary research and, as well, can provide a baseline with which to compare your primary data collection results. Therefore, it is always wise to begin any research activity with a review of the secondary data (Novak 1996).

RESEARCH DESIGN AND PURPOSE

Secondary data analysis and review involves collecting and analyzing a vast array of information. To help you stay focused, your first step should be to develop a statement of purpose – a detailed definition of the purpose of your research – and a research design.

Statement of Purpose: Having a well-defined purpose – a clear understanding of why you are collecting the data and of what kind of data you want to collect, analyze, and better understand – will help you remain focused and prevent you from becoming overwhelmed with the volume of data.

Research Design: A research design is a step-by-step plan that guides data collection and analysis. In the case of secondary data reviews it might simply be an outline of what you want the final report to look like, a list of the types of data that you need to collect, and a preliminary list of data sources.

WHAT TYPES OF DATA AND/OR INFORMATION ARE NEEDED?

The specific types of information and/or data needed to conduct a secondary analysis will depend, obviously, on the focus of your study. For CARE purposes, secondary data analysis is usually conducted to gain a more in-depth understanding of the causes of poverty in the various countries and/or regions where CARE works. Secondary data review and analysis involves collecting information, statistics, and other relevant data at various levels of aggregation in order to conduct a situational analysis of the area (see Data & Indicator List in Appendix 1; refer to the LRSP Guidelines, Annex 5, July 1997). The following is a sampling of the types of secondary data and information commonly associated with poverty analysis:

- Demographic (population, population growth rate, rural/urban, gender, ethnic groups, migration trends, etc.),
- Discrimination (by gender, ethnicity, age, etc)
- Gender equality (by age, ethnicity, etc)
- Policy environment
- Economic environment (growth, debt ratio, terms of trade)
- Poverty levels (poverty and absolute),
- Employment and wages (formal and informal; access variables),
- Livelihood systems (rural, urban, on-farm, off-farm, informal, etc),
- Agricultural variables and practices (rainfall, crops, soil types, and uses, irrigation, etc.),
- Health (malnutrition, infant mortality, immunization rate, fertility rate, contraceptive prevalence rate, etc.),
- Health services (#/level, services by level, facility-to-population ratio; access by gender, ethnicity. etc.),
- Education (adult literacy rate, school enrollment, drop-out rates, male-to-female ratio, ethnic ratio, etc),
- Schools (#/level, school-to-population ratio, access by gender, ethnicity, etc.),

- Infrastructure (roads, electricity, communication, water, sanitation, etc.),
- Environmental status and problems
- Harmful cultural practices

Special attention should be given to collecting disaggregated data. That is, data that is broken down in the following ways: gender, age, ethnicity, location, etc..

Even when highly disaggregated; however, these “raw” data points alone are often only static or indirect measures of the situation or problems that exist in countries and regions – partial or imperfect reflections of reality (UNDP 1997). It is through reviewing, interpreting, and cross-analyzing the secondary that these pieces of information allow us to gain a better understanding of a specific situation, population, sector, etc. Analysis of data gives you the information that you need to make judgements, recommend areas of intervention, and/or design follow-up studies. Cross-analyzing data will also help you understand not only what is happening in a particular area but also WHY it is happening.

SOURCES OF SECONDARY DATA

Official Statistics: Official statistics are statistics collected by governments and their various agencies, bureaus, and departments. These statistics can be useful to researchers because they are an easily obtainable and comprehensive source of information that usually covers long periods of time.

However, because official statistics are often “characterized by unreliability, data gaps, over-aggregation, inaccuracies, mutual inconsistencies, and lack of timely reporting” (Gill 1993), it is important to critically analyze official statistics for accuracy and validity. There are several reasons why these problems exist:

1. The scale of official surveys generally requires large numbers of enumerators (interviewers) and, in order to reach those numbers enumerators contracted are often under-skilled;
2. The size of the survey area and research team usually prohibits adequate supervision of enumerators and the research process; and

3. Resource limitations (human and technical) often prevent timely and accurate reporting of results.

Technical Reports: Technical reports are accounts of work done on research projects. They are written to provide research results to colleagues, research institutions, governments, and other interested researchers. A report may emanate from completed research or on-going research projects.

Scholarly Journals: Scholarly journals generally contain reports of original research or experimentation written by experts in specific fields. Articles in scholarly journals usually undergo a peer review where other experts in the same field review the content of the article for accuracy, originality, and relevance.

Literature Review Articles: Literature review articles assemble and review original research dealing with a specific topic. Reviews are usually written by experts in the field and may be the first written overview of a topic area. Review articles discuss and list all the relevant publications from which the information is derived.

Trade Journals: Trade journals contain articles that discuss practical information concerning various fields. These journals provide people in these fields with information pertaining to that field or trade.

Reference Books: Reference books provide secondary source material. In many cases, specific facts or a summary of a topic is all that is included. Handbooks, manuals, encyclopedias, and dictionaries are considered reference books (University of Cincinnati Library 1996; Pritchard and Scott 1996).

WHERE TO FIND SECONDARY DATA

There are numerous sources of secondary data and information. The first step in collecting secondary data is to determine which institutions conduct research on the topic area or country in question.

Large surveys and country-wide studies are expensive and time-consuming to conduct; therefore, they are usually done by governments or large institutions with a research orientation.

Thus, government documents and official statistics are a good starting place for gathering secondary data; however, as previously stated, the quality of the documents will vary depending on the country of study and the amount of resources dedicated to data collection.

Make Use of Local Experts

When searching for secondary data or questioning the quality of a source that you have already collected, seek advice from sector specialists and other experts in your country office. Your colleagues are valuable sources of information and expertise.

Other major sources of international development data are the World Bank, the United States Agency for International Development (USAID), the United Nations Development Programme (UNDP), the Food and Agriculture Organization of the United Nations (FAO), the International Fund for Agricultural Development (IFAD), the World Health Organization (WHO), International Center for Research on Women (ICRW), the Chronic Poverty Research Center (CPRC), the Center for Research on Poverty (CROP), Overseas Development Institute (ODI), and Institute of Development Studies (IDS) to name a few.

International development institutes commonly share information sources and have libraries for archiving these materials. Thus, a data-gathering visit to one office might yield numerous sources of information on the topic area of interest.

University libraries are good sources of information and should be consulted. Also, it would be beneficial to establish contact with experts at local university departments that are dedicated to research on the topic areas that you are interested in (e.g., Departments of Agricultural Sciences, Public Health, Economics, Anthropology, Sociology). These experts can be important sources of information on on-going research projects as well as for guiding you toward other sources of topic area information or individuals that can be contacted.

Local NGOs also often conduct empirical research and can be valuable sources of information. This is particularly true when you are searching for local-level information and data. In some cases, NGOs might also have small libraries that provide additional information.

Secondary Data Sources

**Government Documents
Official Statistics
Technical Reports
Scholarly Journals
Trade Journals
Review Articles
Reference Books
Research Institutions
Universities
Libraries, Library Search Engines
Computerized Databases
The World Wide Web
(Shell 1997)**

EVALUATING THE QUALITY OF YOUR INFORMATION SOURCES

One of the advantages of secondary data review and analysis is that individuals with limited research training or technical expertise can be trained to conduct this type of analysis. Key to the process, however, is the ability to judge the quality of the data or information that has been gathered. The following tips will help you assess the quality of the data.

Determine the Original Purpose of the Data Collection: Consider the purpose of the data or publication. Is it a government document or statistic, data collected for corporate and/or marketing purposes, or the output of a source whose business is to publish secondary data (e.g., research institutions). Knowing the purpose of data collection will help to evaluate the quality of the data and discern the potential level of bias (Novak 1996).

Attempt to Ascertain the Credentials of the Source(s) or Author(s) of the Information
What are the author's or source's credentials -- educational background, past works/writings, or

experience -- in this area? For example, the following sources are generally considered reliable sources of data and information: research reports documenting findings from agricultural research published by the FAO or IFAD; socioeconomic data reported by the World Bank; and survey health data reported in USAID's Demographic Health Surveys.

Does it include a methods section and are the methods sound? Does the article have a section that discusses the methods used to conduct the study? If it does not, you can assume that it is a popular audience publication and should look for additional supporting information or data. If the research methods are discussed, review them to ascertain the quality of the study. If you are not a research methods expert, have someone else in your County Office review the methods section with you.

What's the Date of Publication? When was the source published? Is the source current or out-of-date? Topic areas of continuing or rapid development, such as the sciences, demand more current information.

Who is the Intended Audience? Is the publication aimed at a specialized or a general audience? Is the source too elementary -- aimed at the general public?

What is the Coverage of the Report or Document? Does the work update other sources, substantiate other materials/reports that you have read, or add new information to the topic area?

Is it a Primary or Secondary Source? Primary sources are the raw material of the research process, they represent the records of research or events as first described. Secondary sources are based on primary sources. These sources analyze, describe, and synthesize the primary or original source. If the source is secondary, does it accurately relate information from primary sources?

Importantly, Is the Document or Report Well-Referenced? When data and/or figures are given, are they followed by a footnote, endnote - which provides a full reference for the information at the end of the page or document - or the name and date of the source (e.g., Burke 1997)? Without proper reference to the source

of the information, it is impossible to judge the quality and validity of the information reported.
DO THE NUMBERS DO NOT MAKE SENSE?

Data reporting characteristics vary according to what the data is being collected for and the stage of reporting. For example, health clinics might report quarterly the number of cases of diarrhea, upper respiratory infection, or malnutrition that they have been treated at a clinic.

This information is useful for healthcare professionals who will later analyze the information to ascertain the percentage of the population in a municipality or province that were diagnosed with these problems over a given period of time. For the purpose of secondary data analysis, the aggregated percentage figure, rather than the number of "cases" reported, should be used.

Another area of data analysis that requires a skeptical eye is employment-related data. It is difficult to count the employed accurately, especially in developing countries. Employment data often do not take into account the number of people involved in informal or unrecorded activities, seasonal agricultural laborers, women's agricultural labor, or child labor.

Thus, official employment statistics should be viewed in light of these inadequacies. Labor force data that provides a list of the categories used (e.g., employed, unemployed, underemployed, own-account workers, unpaid family workers) will help you determine the quality of the measure (Worldbank 1997).

When you feel that the employment data is unreliable, looking at other economic indicators will help you develop a clearer understanding of the situation. For example, if your employment data state that only 25 percent of the population is economically active. However, data from a recent poverty survey state that only 5 percent of the population live below the absolute poverty line, you can conclude that the employment data is not a good measure to use.

Questions for Evaluating Data Quality

- **What are your source's credentials?**
- **What methods were used?**
- **Is the information current or out-of-date?**
- **Is the intended audience other researchers or the general public?**
- **Is the document's coverage of the topic area broad or too narrow?**
- **Is it a primary or secondary source? If it is a secondary source, does it accurately cover and report on the primary sources?**
- **Does the author provide references for the data and information reported?**
- **Do the numbers make sense? Are they the numbers you want – cases versus percentages? When compared to related data are the measures somewhat consistent?**

WHAT DO YOU DO WHEN DATA SOURCES DISAGREE?

When conducting secondary data analysis, it is not uncommon to come across data sources that disagree or conflict with each other. To help overcome this problem you should:

1. Decide if the source of the data is a primary or a secondary source. In other words, look for a citation. If the source is simply quoting a number or statistic, it may not be accurate, and should be taken cautiously.
2. If you cannot find the original source of the data in question, look for more data sources covering the topic and determine the most widely held conclusion. If two independent secondary data sources agree, the information is probably more believable.

3. Consult a local expert in the topic area. Make use of the valuable resources around you. More than likely, there are colleagues at your country office, in local government offices, or other institutions that can easily help resolve an issue, answer your questions, or direct you to the answers.

THE IMPORTANCE OF DATA DISAGGREGATION

The level of data aggregation or disaggregation simply refers the extent to which the information or data is broken down.

Aggregate Data: Aggregate data are data that describe a group of observations, with the grouping made on a defined criterion. For example, geographic data are often grouped by spatial units such as region, state, census tract, etc. Aggregate data can also be defined by time interval, for example: the number of persons that migrated to urban areas in the last five years.

Disaggregated Data: These are data on individuals or single entities, for example: age, sex, level of education, income, occupation, etc. These data are generally more informative and useful than aggregate data.

There have been increasing efforts over the last couple of decades to encourage more data disaggregation, particularly in the international development arena. Strong emphasis has been placed, for example, on data disaggregation by gender, ethnicity, age, and location. This type of data enables researchers and development practitioners to obtain a more comprehensive understanding of how groups within a society react differently to and/or are effected differently by various conditions or events (e.g., structural adjustment, other socioeconomic conditions, environmental degradation, government policies, development projects, etc.).

For example, although women have been found to be the primary agriculturists in many societies, until recently the majority of data related to farming was gathered from male farmers and often did not represent the same reality experienced by women farmers. Thus the lack of gender-disaggregated agricultural data has been a major constraint for effective integration

of women in the planning and implementation of agriculture and rural development programs (FAO 1997).

When conducting analysis of poverty determinants it is always advisable to gather your data at the lowest possible level of geographic/political unit aggregation (e.g., municipality, county). This allows you to not only ascertain what is happening at the department or state level but also within these areas. In some situations, you might find that the level of data disaggregation varies across or between political/geographic unit.

Aggregate versus Disaggregate

In a nutshell, you can say that the more aggregated the data, the more invisible the people.

You might have nutritional data that are disaggregated only at the department level for 10 out of 14 departments in a country, with only four departments having both department and municipal level data. In this case you would present data at the department level because it represents the broadest level of consistent data collection (i.e., you have it for all departments in the country).

If time permits, it is also advisable to briefly discuss the municipal level data for the four departments as an example of what is happening at the municipal level. While this information is not generalizable at the country level, it will enhance your knowledge of the nutritional situation in the specific geographic areas/municipalities. When complemented with other data (livelihood systems, levels of poverty, agroecological zone, etc.), these data can help us further characterize and understand the local situation and make inferences to higher levels of data aggregation.

GETTING FROM WHAT AND WHERE TO WHY

Secondary data is generally referred to as *outcome* data. This is because secondary data generally describe the condition or status of phenomena or a group; however, these data

alone do not tell us *why* the condition or status exists. This limitation can be overcome in two ways.

First, it can be overcome by using information from case studies and other research to fill in the gaps. For example, data on child malnutrition rates and women's level of education provide information relevant for understanding why some children are more likely to be malnourished than others. The child health research literature tells us that children whose mothers have low levels of education will likely exhibit higher malnutrition rates than children of mothers with higher levels of education. Thus, consulting relevant literature can help illuminate causal relationships.

Second, analysis of additional key data and indicators can help us acquire more explanation as to why a problem exists. For example, if low farm income has been identified as a problem, data on land tenure, land size, types of crops, production value, cost of inputs, and so on, can be compared to help identify who has this problem and possible causes and solutions.

Therefore, cross-analyzing key indicators and using additional information sources help us understand or make reasonably sound inferences about unmeasured conditions or situations; thus allowing us to better understand not only *what is happening* and *where it is happening* but also *why it is happening*.

USING SECONDARY INFORMATION TO STRENGTHEN PRIMARY RESEARCH

Secondary information is also valuable for generating hypotheses and identifying critical areas of interest that can be investigated during primary data gathering activities. For example, secondary data analysis conducted prior to the Tanzania Urban Food and Household Livelihood Security Assessment, identified key research areas that should be closely studied during the assessment (e.g., the influence of seasonality on urban livelihoods; the impact of increased privatization; gender differentiation in urban land tenure policies; and social cohesion and locality). Questions were then developed and included in the household, community group, and key informant interview questionnaires to allow analysis of these key areas of interest.

Without thorough analysis of secondary data, these key constraints to urban food and livelihood security possibly would not have been identified, thus the problem analysis – the *why* of the primary research exercise – was strengthened through secondary data and information analysis.

THE ADVANTAGES AND DISADVANTAGES OF SECONDARY DATA REVIEW AND ANALYSIS

Advantages:

- Secondary data analysis can be carried out rather quickly when compared to formal primary data gathering and analysis exercises.
- Where *good* secondary data is available, researchers save time and money by making good use of available data rather than collecting primary data, thus avoiding duplication of effort.
- Using secondary data provides a relatively low-cost means of comparing the level of well-being of different political units (e.g., states, departments, provinces, counties). However, keep in mind that data collection methods vary (between researchers, countries, departments, etc.), which may impair the comparability of the data.
- Depending on the level of data disaggregation, secondary data analysis lends itself to trend analysis as it offers a relatively easy way to monitor change over time.
- It informs and complements primary data collection, saving time and resources often associated with over-collecting primary data.
- Persons with limited research training or technical expertise can be trained to conduct a secondary data review (Beaulieu 1992; University of Cincinnati 1996).

Disadvantages:

- Secondary data helps us understand the condition or status of a group, but compared to primary data they are imperfect reflections of reality. Without proper

interpretation and analysis they do not help us understand *why* something is happening.

- The person reviewing the secondary data can easily become overwhelmed by the volume of secondary data available, if selectivity is not exercised.
- It is often difficult to determine the quality of some of the data in question.
- Sources may conflict with each other.
- Because secondary data is usually not collected for the same purpose as the original researcher had, the goals and purposes of the original researcher can potentially bias the study.

Secondary versus Primary Data

Secondary data complements, but *does not replace*, primary data collection and should be the starting place for any research activity.

- Because the data were collected by other researchers, and they decide what to collect and what to omit, all of the information desired may not be available (Israel 1993).
- Much of the data available are only indirect measures of problems that exist in countries and regions (University of Cincinnati 1996).
- Secondary data can not reveal individual or group values, beliefs, or reasons that may be underlying current trends (Beaulieu 1992).

THE IMPORTANCE OF PROPERLY REFERENCING YOUR SECONDARY DATA REVIEW

Secondary data review and analysis is a form of research and data compilation that is demanding and time-consuming; however, without proper citation (i.e., author, date, title) of materials that you used, your work will often be disregarded as it will only have limited use by those who wish to follow in your footsteps.

- A well-documented secondary data review and analysis allows for easier use of the material by other interested parties.
- Properly citing the publication date of the sources you used will allow subsequent researchers to use your work to make comparisons over time and between countries, communities, towns, regions, etc.
- Proper citation allow subsequent researchers to use your work, thus preventing unnecessary duplication of research efforts (Qualidata 1997)

SUMMARY

Secondary data can be a valuable source of information for gaining knowledge and insight into a broad range of issues and phenomena. Review and analysis of secondary data can provide a cost-effective way of addressing issues, conducting cross-national comparisons, understanding country-specific and local conditions, determining the direction and magnitude of change -- trends, and describing the current situation. It complements, but does not replace, primary data collection and should be the starting place for any research

Originally Prepared by: M. Katherine McCaston, Food Security Advisor & Deputy Household Livelihood Security Coordinator, PHLS Unit, CARE, February 1998. Updated by same June 2005

References Cited

- Beaulieu, Lionel J.
1992 Identifying Needs Using Secondary Data Sources, Institute of Food and Agricultural Services, University of Florida.
- Cnossen, Christine
1997 Secondary Reserach: Learning Paper 7, School of Public Administration and Law, the Robert Gordon University, January 1997. Available online (telnet): jura2.eee.rgu.ac.uk/dsk5/research/material/resmeth
- FAO
1997 User/Producer Workshop on Gender-Disaggregated Agricultural Statistical Data, FAO Women in Development Service and FAO Regional Office for Africa, Harare, Zimbabwe, September 1997.
- Gill, Gerard J.
1993 O.K., The Data's Lousy, But It's All We Got (Being a Critique of Conventional Methods). London: International Institute for Environment and Development.
- Israel, Glenn D.
1993 Using Secondary Data for Needs Assessment, Fact Sheet PEOD-10, Program Evaluation and Organizational Development Series, Institute of Food and Agricultural Services, University of Florida.
- Novak, Thomas P.
1996 Secondary Data Analysis Lecture Notes. Marketing Research, Vanderbilt University. Available online (telnet):www2000.ogsm.vanderbilt.edu/marketing.research.spring.1996.
- Pritchard, Eileen and Paula R. Scott
1996 Literature Searching in Science, Technology, and Agriculture. Westport, CT: Greenwood Press.
- Qualidata
1997 Guidelines for Depositing Qualitative Data, ERSC Qualitative Data Archival Resource Centre. Available online (telnet): www.essex.ac.uk/qualidata
- Shell, L.W.
1997 Secondary Data Sources: Library Search Engines, Nicholls State University.
- Trochim, William
1997 The Knowledge Base Homepage: Research Methods, Cornell University. Available online (telnet): trochim.human.cornell.edu/kb.

M. Katherine McCaston, HLS Advisor June 2005

Updated from M.Katherine McCaston (1998) -Partnership & Household Livelihood Security Unit

University of Cincinnati

1996 Critically Analyzing Information, the
Reference Library, University of
Cincinnati. Available online (telnet):
www.libraries.uc.edu/libinfo

UNDP

1997 Sustainable Livelihoods: Concepts,
Principles and Approaches to Indicator
Development (Draft Discussion Paper).
Available online (telnet):
www.undp.org/seped/sl/ind2.htm.

Worldbank

1997 World Development Indicators.
Washington, D.C.: Worldbank.