Unit 3: Study Guide
Making cross national comparisons using microdata

MIMAS
The University of Manchester

3.1. Introduction

Conducting analysis that seeks to describe or explain differences between countries or groups of countries provides a more complex set of challenges that analysing data from a single survey source in a domestic context. The analyst needs to exercise additional caution to ensure that the data is suitable for the analyses intended. It is your responsibility as an analyst to ensure that the analyses you undertake are valid. Perfect data doesn't exist so you will need to apply your professional judgement as to what is "good enough" to use for a particular purpose and what is not. This will mean making an assessment of the data and the extent to which it is fit for the purpose you intend. For example, if the topic is one in which there may be an element of competition or national pride such as educational attainment, or healthy eating, then you will want the sample to be representative of the whole population and the measurement to be fairly robust. For topics where countries probably wouldn't mind being at the bottom or top of a league table you might be less concerned about coverage. Ask yourself the questions, *"How would I defend my conclusions if someone challenged them?" "Am I confident in the results of this analysis?"*

Unit 2 in this series has already dealt with issues around making international comparisons using aggregate data. This chapter will look at issues when using microdata or primary data sources to do cross-national analysis. Many of the issues are similar since they may be on the same primary data although remember that the aggregate data may have been transformed in some way.

Many of the examples are draw from the International Adult Literacy Survey www.oecd.org/document/2/0,2340,en_2649_34509_2670850_1_1_1_1,00.html. This survey aimed to measure the literacy skills of adults through a household survey using a direct assessment more commonly associated with tests carried out in schools. The survey results were controversial and as a result a methodological review was conducted to investigate whether the results were subject to some form of cultural bias. The findings from this review were important in developing an understanding of the issues in cross-national surveys and in improving the design and implementation of surveys such as the **European Social Survey.**

3.2. Learning objectives

By the end of this unit you will be able to:

- identify main areas that need to be considered before undertaking any analysis
- Be aware of steps in preparation of microdata to make cross-national comparison
- Understand the need to work efficiently to document the analysis
- Be able to discuss some of the key design features that impact on cross-national analysis
- Have considered issues in respect of a specific research question
- Understand need to set analysis in context
- Have done some basic analyses using the European Social Survey

3.3. Preparation

One of the most important steps in any cross national analysis is the preparation – expressing the research question, understanding the data being used and getting the data into the right format.

This often takes more time that the analysis itself.

3.3.1 Setting the research question

The starting point for any analysis should be the research question rather than the dataset. Having a well defined hypothesis will help in framing the analysis and in deciding which source to use and which analysis methods are appropriate. However attractive, resist the temptation to decide on the source first and to then trawl or dredge for "interesting" analysis – this rarely leads to interesting or useful insights. Your research question will also guide you as to which countries you wish to include in your analysis. For example if your research question is about the performance of educational systems you will want to make sure you include countries that have different educational structures. Similarly, if your hypothesis is something on social or political attitudes you will want to include countries that have different political systems, structures and traditions.

One pitfall that many researchers fall in to is to over complicate the question and to lose sight of the key question. It is often a good discipline to try and articulate the key question in 2 or 3 sentences. For some reason, explaining to some one verbally why it is an interesting topic or question helps crystallise what needs to be done. Imagine responding to a comment of "so what". It is very easy to lose sight of the simplicity of the question in preparing research proposals or in synthesising the theory.

3.3.2. Which countries to include in your analysis

The main constraint on which countries to include will be whether data is available. Assuming you have located a data source that has a large number of countries you might want to take different approaches at different stages of the analysis. Starting with a large number of countries can be useful in identifying the overall pattern and then you can narrow down the number of countries using the most polarised or interesting countries to illustrate the findings. Limiting the number of countries has benefits in making the analysis and presentation of

results much more manageable. Don't underestimate the difficulty of displaying data for a large number of countries in a way that is easily accessible to the reader. A handful of countries can sometimes serve to demonstrate the points more usefully than including 30 or 40 countries. For example, if all the Nordic countries or all the newer members of the EU are show a similar story then it may be worth only showing a subset in your graphics to illustrate the point rather than including all of them.

> It is important to ensure that you do not skew your analysis by excluding countries simply because they don't fit. Countries that don't fit are an interesting phenomenon so they are often worth a bit of detective work to try and understand more about why they don't behave in the expected way.

If you have a fairly well defined hypothesis about how different countries will feature, you may wish to start with a subset countries while you do the initial analysis and then add in more countries to test the hypothesis in a further set of countries and then make a final selection for the purpose of commenting or presentation of the results.

It is also important to remember that finding **no major difference across countries is as valid a research** finding as finding systematic differences, particularly if it might intuitively be expected that they would differ.

Aside from issues of substantive interest such as social or political attitudes many international surveys such as the European Social Survey offer the opportunity to analyse the effect of different research methods such as differences in mode. See link for methodological experiments using the ESS.
http://naticent02.uuhost.uk.uu.net/methodology/methodological_research.htm

3.3.3. What data to use

The first decision to be made is what data to use. There are two routes to sourcing microdata for analysis.

Unit 1 in this series has already described many of the data sources available. Using international datasets such as those available through **ESDS International** – these are datasets that are compiled for this purpose and are described in overview in unit 1 of this series. These types of datasets will usually contain data

for many countries in a single data file or in separate datafiles but with a similar structure. They will have been through a process of preparation which means that they are standardised in format and the metadata is provided to guide the user. The extent to which this is done will vary so the usual rule is "user beware". Where data is available for more than one year you will also need to check that the data structure is the same for both years. For example, the Programme for International Student Assessment www.pisa.oecd.org/pages/0,2987,en_32252351_32235731_1_1_1_1_1,00.html has the data organised into several different files and is available for different years of the survey. There is no right or wrong way to structure datasets, it is usually driven by the expected uses of the data for the majority of users so as to minimise download time. This means that for some analyses you may need to manipulate the data into a suitable structure.

A second approach is using datasets from different countries that are either from the same series on a particular topic. Such surveys are not necessarily designed to support cross-national analysis nor presented as an international dataset – however, on some topics they may be just as good as those intended as international surveys or indeed may be the only data available on a particular topic. Examples include surveys such as **Household Income and Expenditure Surveys (HIES)** or **Labour Force Surveys** which are conducted in most countries or surveys such as the **Demographic and Health Surveys (DHS)** http://www.measuredhs.com/.

In using this approach you will need to do quite a bit of additional work to get the data into a suitable format and make your own decision about whether they are sufficiently similar to use for the analysis intended. This of course applies also when combining sub-national surveys such as in federal systems where data may be available form separate state surveys but not compiled into a single data set. For example, in making comparisons between England, Scotland, Wales and Northern Ireland or in analysis of the different territories in Australia or provinces in Canada or other countries which have a federal structure such as India.
It is very easy to fall into simple traps by not checking that your assumptions hold true. For example, when localising a questionnaire some countries will change the order of the codes either to fit in with the norm for that country or simply so as to own the localised version. So in Country A it might be that 1= female and 2=male while in Country B the opposite applies. These subtle changes

can mean you are completely wrong footed in your analysis. Either way, not spotting the difference can waste a lot of time.

3.3.3.1. How standardised is the data?

In producing international datasets there are different approaches to harmonisation. In some surveys such as the ESS, the inputs are harmonised as much as possible – so the design in each country is as similar as possible and the questionnaire is standardised in advance. In other surveys, the international dataset set is harmonised at output level, so the specification is set as to the output variable and countries can arrive at that through their own national processes. In this instance the questions may be different by design.

For some international datasets it is useful to be aware that there are also sometimes corresponding national datasets available which may contain additional variables, slightly different versions of the international variables or indeed may contain additional sample. For example, many international data files use various international classifications such as the **International Standard Classification of Occupations (ISCO)** or the **International Standard Classification of Education (ISCED)**. The national data files may contain the usual national classification for that country which may provide more depth. For example, an international survey may contain household income variables that are based on quintiles or deciles. The national survey microdata may contain the variable either as a continuous variable showing the actual amount of household income in the local currency, it may contain banded income in the local currency and it may contain the international percentiles variable.

Surveys may also have added questions of national interest which were not required for the international survey. Similarly, the population defined as in scope for the national sample may have been broader than that required.
For example, an international survey on fertility may direct questions only at women aged 16-49 or in some countries only to women who are married or sexually active. The national samples may have used a more extended age range so additional sample may be available on a national version of the datafile. Similarly questions on drinking and smoking may be directed at different populations reflecting legal age for drinking, questions on employment may be asked of different ages reflecting the differences in the retirement age. The national data files may also contain more detail on income or education.

3.3.3.2. Assessing the data

It is essential that you fully assess any data before you start. It is your responsibility as an analyst to satisfy yourself as to the legitimacy of the analyses you intend to conduct resisting the temptation to just "plug and play" particularly when using an international dataset that is already established.

Remember that you are unlikely to be the first analyst using this dataset so a useful first step should always be to look for analyses already published from the source. For surveys such as the European Social Survey or the World Values Survey you will find a list of publications on their website.
http://naticent02.uuhost.uk.uu.net/publicity/index.htm
or http://www.worldvaluessurvey.org/

Looking at existing analyses will provide you with important clues as to what you need to take into account in analysing that particular dataset, such as population coverage issues or which weights if any need to be applied .

In particular, look for analyses that are on a **similar topic** or that use **similar types of variables**, continuous, ordinal, nominal, etc. Reading existing publications will also provide you with information on how you might display the data in tables and graphs. It is very rare that you will be conducting unique analyses so make use of what is already available to improve your analysis but obviously with a questioning mind.

3.3.3.3. Understanding the data

The first task of any analyst should be to read the detailed methodology of the survey to be used. This will give you information about how the survey was conducted and provide you with the information needed to assess whether you are happy to use this dataset for the purposes you intend. It is important that any differences evident from your analysis are real and are not simply an artefact of differences in how the surveys were conducted in the participating countries or a result of the questions having slightly different meanings.

It is important to look not only at the design specification but all at any information available about deviations from the design.

The **European Social Survey** is one of the few surveys to provide information on adherence to the design specification making it easier for the user to be aware of any non compliance or deviations that might affect the results. The detailed methodologies for the various surveys can be found at: http://www.europeansocialsurvey.org/

Essential reading on the **adherence to the design specification** and on **methodological tests** in the development of the survey are available for round 1 of the European Social Survey at: http://naticent02.uuhost.uk.uu.net/archive/end_of_grant/R1_end_of_grant.pdf

In the next section we will look at some of the things to look for in assessing microdata.

3.4. Things to look for in assessing microdata

3.4.1. Comparability of overall design

You need to satisfy yourself that the surveys you are using are of sufficiently comparable design to support the analysis you intend. For surveys that are explicitly designed to support cross-national comparisons you need only consider whether the departures from the design specification are warranted and whether the variables of interest are subject to any "difference" between the countries. For example, if the survey used different modes in different countries (telephone, face to face, self-completion with interviewer, mail, web etc) is the mode likely to have an effect on how the question was answered or whether there are any issues around the cultural equivalence of the item?

Where you are using datasets that are not explicitly designed for this purpose you need to exercise your own judgement on comparability. It is increasingly common with international surveys for countries to be allowed to use whichever mode of survey collection is "usual" in that country. This means that in some countries the survey will be by face-to-face interview while in others the survey may have been by telephone or by post. Internet surveys are still not widely used as a method although some countries are experimenting with them as an additional mode. For example, the Australian Bureau of Statistics included the option to complete the 2006 Census on line.

3.4.2. Comparability of questions

There are several elements to this. First, whether the question stem has the same meaning in all the countries, secondly whether the response categories for capturing the answer are equivalent and thirdly whether the question was asked to the same population. While the dataset will contain variables that are either simply direct responses to a question posed by the interviewer or are constructed from responses to a series of questions it is good practice to always check back to the source question to satisfy yourself that the question has been transformed into the variable in a standard way and that you fully understand the concept captured in the variable. For countries that share a language this is more straightforward than where the question has been translated or adapted but even where countries share a language this may not always mean that there are comparable. Consider for example comparisons between the US and the UK. The question will rarely be exactly the same as each country will reflect their own use of English and differences in vocabulary. Have the questions been adapted adequately to reflect the use of language in both countries.

For example – in a question in the IALS test based on a leaflet from a bank on interest rates which asked in the English version to **List all the rates** … the version used in France was Quels taux…? **(Which rates?)** … and in Switzerland Enumerez tous les taux …? **(List all the rates ...)**. In the Swiss survey 74% answered the question correctly compared with 43% in France. When this translation error was corrected in a retest of the respondents in France 75% answered this question correctly.

3.4.3. Sample Design

An important question is whether the samples in each country are sufficiently similar that you are confident in making comparisons. Has the survey used probability sampling at all stages – that is that each element of the population has a known probability of selection that is not equal to zero.

In-field sampling procedures, that is where the final stage of sample selection is conducted in the field, are particularly important – **pay attention to whether the final stage of sampling was carried out by the interviewer and the method used**. Two main types of infield sampling may occur – where the address is being selected from within an area selected at random by the centre and where a household is being selected at a multi-household address or where an individual is being selected from within a household. Where possible it is better to separate the functions – in particular, where an address is being selected in the field this should be carried out by someone other than the interviewer to avoid difficult to interview addresses being substituted. It is more difficult to apply this to selection of households within addresses or to selection of individuals within households since contact with the households before the approach for an interview may jeopardise response rates. It is more usual therefore for these types of infield sampling to be conducted by the interviewer.

Timing and reference period – Were the surveys conducted at around the **same time** and if not does that matter – for some topics you might conclude that a difference in timing of 2 or three years might not be a problem since the topic of interest is unlikely to be subject to fluctuation. For other topics, such as social or political attitudes even small differences in timing might be problematic depending on the context. For example whether there were elections held or whether there were political events that might impact on the variables of interest. Similarly, are the **reference periods** in the questions the same – where they ask questions about expenditure for example do they ask about expenditure in the

week ending last Sunday, in the past 7 days or in the past month or a longer period. This is particularly important when the question requires the respondent to recall events. The longer the time period the more difficult it is for the respondent to answer accurately.

3.4.4. Coverage

Is the population coverage similar across the range of countries both in definition and execution? For many surveys the population is defined as residents in private households rather than the entire population so residents in communal establishments such as University Halls of Residence, residential care homes, prisons and hospitals are defined as out of scope. Is this important for your research question? Have any subgroups such as speakers of minority languages been left out from the survey as a whole or excluded from particular questions? Does the within household sample selection favour the head of household? Are geographically remote areas excluded or are hard to reach groups or difficult to interview groups excluded? Many surveys set a maximum non coverage rate – have these been met?

Even if the population appears to be similarly defined, for example, if the survey is of adults of working age the population definition will vary from country to country reflecting the legislation on retirement age. Similarly, for surveys of school children, the population in scope will be defined slightly differently taking account of differences in age of compulsory schooling. It is important to pay attention to issues of population definition and scope to ensure that you are making valid comparisons. This is an area that can easily be overlooked so familiarising yourself with the detail of the survey conditions is very important.

Within the survey there may be further narrowing of scope for particular questions and this may differ with some respondents routed past particular questions or sets of question. For example, in some countries, questions to women on sexual activity, fertility and family planning are only asked of women who are married.

3.4.5. Survey response

Are the survey response rates similar across the countries you wish to include in your analysis or are there any issues around bias due to differential survey response rates? It is becoming increasingly difficult to gain co-operation from the

Although high response rates help minimise the potential bias they do not protect against it completely. It is possible to have a high response rate and still have bias if all those who did not respond are similar to each other and different from those who did participate.

public for surveys in many developed countries and response rates have fallen in many countries over the past 15 years.

3.4.6. Data collection

Have the surveys used similar methods and conditions for collecting the data? Is there anything in how the survey was introduced to the respondents that may influence their responses or anything in how the interviewers were paid or in the terms and conditions which may impact on the interview process. Were similar fieldwork procedures followed in each country in how the survey was presented to respondents, were incentives paid to respondents, were non-responding cases reissued to another interview for refusal conversion – these are all questions that may impact on respondent understanding of the questions and will vary for different types of surveys. What quality control procedures were in place? For a discussion of the potential effects see Kalton, Lyberg and Remmp (1998) Annex A p215 – 230. (http://nces.ed.gov/pubs98/98053.pdf). Reading the requirements or specifications of the survey will give you a good sense of the things the designers consider may affect comparability. The design will try to minimise variation in the surveys across the different countries but it is of course also necessary to look at the execution since sometimes departures from the strict specification may be necessary and permitted.

3.4.6.1. Questions and response categories

Are they equivalent across countries and languages and how have they been translated? Clearly, the framing of the question and the response categories are key to being able to make comparisons. Aside from looking at the question stem it is important to also look at how the response categories are set out. Different modes have different formats – for example as to whether don't know and refusal are offered explicitly as response options to the respondent or are only used in response to a spontaneous use by the respondent. Similarly, individual prompts, where the interviewer reads out each response category and the respondents' replies to each, are only possible where an interviewer is administering the questionnaire. Clearly, on self-completion questionnaires the respondent can see

the full range of response options before answering including whether don't know is a valid response. There is a vast and growing body of literature on the effect the layout and structure of the questionnaire has on the respondent so having a good look at the questionnaire for any differences in presentation is an essential starting point.

3.4.6.2. Translation

This is one of the most overlooked areas in cross cultural analysis. Some of the questions to ask are: How was the questionnaire used in each country developed? Does the translation attain equivalence of meaning? The issue of cultural equivalence also applies to the adaptation of questions within a language. For example, the adaptation of the questionnaire for use in France, Switzerland and Canada will mean that the questions are different. Translation and adaptation (or localisation) of questions can introduce important differences in meaning as seen in the IALS.

In developing a questionnaire to be used across languages or cultures it is important to try and establish linguistic equivalence. This means that the concept being measured is the same in both target languages. Being aware of the difficulties of achieving linguistic equivalence is an important first step. It is also important to recognise that questions in the source language(s) are often improved by putting them through the process of translation. In developing a questionnaire it is better to work in multiple languages from the start, in an iterative process, with all versions of the questionnaire being improved at the same time. Many questionnaires for large studies are now developed simultaneously in more than one language with several translators. In the past questionnaires were often developed in a single language, usually English and then translated or adapted. Back translation was usually the recommended method for checking the accuracy of the translation. This is no longer recommended.

If you are looking at variables which are outside the range of standard classificatory variables or are new questions in an ongoing survey you should ideally look at the questionnaire in each country that you want to include in your analysis in the original language if you have sufficient awareness of the language. This is to ensure that the questions are asking about the same concepts and about the same phenomenon. It is very easy for subtle changes to be made in translation that can change the meaning or intent of the question. In many

surveys it is difficult to get access to the various language versions of the questionnaire.

More attention is now paid to trying to achieve cultural equivalence of questionnaires and this is a quickly developing area with new guidance being developed about translation processes. The European Social Survey translation guidelines are a good example of how practice is changing in relation to survey instruments as are the PISA guidelines.
http://naticent02.uuhost.uk.uu.net/methodology/translation_strategy.htm
The questions raised by countries show some good examples of the practical problems encountered in translating survey questionnaires. For more on survey translation see Harkness, Pennell and Schoua-Glusberg (2004).

3.4.6.3. Missing data

Another area that is often overlooked in assessing data is the treatment of missing items. While good attention is paid to differences in response rates – missing cases – the treatment of missing items - how the 'don't knows' and refusal to individual questions have been treated is sometimes ignored. Sometimes differences in distribution for a particular variable can be because one country has calculated with don't knows in the base while another country has calculated with the 'don't knows' set to missing. In some instances the differences in items missing can be large with the proportion of respondent who refuse to answer the income question being much higher in one country than in another or where the mode of the survey may have impacted on the willingness to respond. Before including variables in your analysis make sure to check that it is reasonably populated and the rates are similar across countries.

3.4.6.4. Data Processing

In most surveys the data is modified in some way during processing. For example, inconsistencies in the data structure will have been cleaned up and new variables will have been derived from combinations of responses to a set of questions. For example, a variable which describes the family type may have been constructed based on the response to a single question asking about family type or it may have derived from a series of questions about each individual in the household and how they are related to each other.

Other processes that may have been carried out to transform the data will be to impute for missing data. This is sometimes built into the design of the survey

such as in education assessments using direct testing methods which routinely use a methodology where scores are imputed based on the responses of others with similar characteristics. Have the same rules been applied in all the countries?

For further information on the kinds of questions to ask look at the national technical summaries questions asked in the European Social Survey.
http://naticent02.uuhost.uk.uu.net/images/filetype_pdf.gif

3.4.6.5. Measurement in specialist areas

In some areas the measurement issues are complex, for example in health and education. Ensure you familiarise yourself with the particular measurement properties of the variables you intend to use. For example, many of the educational assessment studies use an item response model which means that any one student only answers a subset of the total test. This is to ensure a broad range of topics can be covered. The design of these studies is to estimate the variable of interest at population rather than individual level although the files contain estimates for each respondent. The attainment variable is then modelled to reflect how that student would have done if they had answered all the questions in the test based on the how similar students performed on the questions not answered. Many of them also use a **Plausible Values (PV) methodology** so that different estimates of attainment variables multiple values for an individual are provided. The spread of these plausible values can be quite wide. Read the recommendations of the survey team for which method to use for selecting plausible values for your analysis.

In health there are complex measurement issues in physical and biochemical measurement. For example, the DHS takes blood spots for testing for anaemia and for HIV. Again, familiarise yourself with the methodological and process issues involved.

Activity 1 : Appraisal of microdata sets

You will need to go online to complete this activity. Conduct an appraisal of two surveys in the ESDS International microdata set and identify the strengths and weaknesses of both. Identify any particular constraints on how you might use the data. You will need your Athens username and password to complete this task.

3.5. Getting started

The previous section may seem like rather a long list of things that you need to do before you even get started. Understanding the structure of the data and all the design issues are essential as they will help guard against erroneous conclusions and will be very valuable when it comes to interpretation of the results. The consequence of ignoring these issues may not seem evident for post-graduate analysis but are of important for analyses put out by international organisation or governments where they are creating a league table of results. The lessons from the results of the International Adult Literacy Survey in the mid nineties are a good example.

3.5.1. What happens when the results are questioned: Three different examples

The growing interest in international surveys and the desire for governments to be able to make comparisons raises the importance of making sure the results are accurate. Some recent examples illustrate the political fallout from results that look implausible.

Example 1

In 1994 nine countries took part in the first International Adult Literacy Survey (IALS). The results show large differences between countries in the distribution of skills. In Sweden, only 8% of the population were found to be at the lowest literacy level compared with 14% in Germany, 17% in Canada and 21% in the United States. Both France and Poland were found to have much higher proportions of their population at the lowest level, 41% and 43% respectively.

The French Government withdrew the results from the publication in the absence of any plausible explanation as to why the rates were so high in France compared with their neighbouring countries. This resulted in front page news in France on two fronts, about whether the results were right and about the role of literacy in modern French society. The long running story led to a methodological review of the survey which found that the results for France as well as some other countries were subject to concern from a number of perspectives, most notably translation, sampling, survey procedures and scoring. France went on to develop an alternative method for measuring literacy, l'Enquête Information et Vie

Quotidienne (IVQ) (Information in Daily Life).
http://education.france5.fr/mde/W00359/7/100997.cfm
http://www.anlci.gouv.fr/documents/actes112003/jeantheau_en.pdf

Example 2

In the literacy surveys of 15 years olds conducted under the **Programme for International Student Achievement (PISA)** in 2000. In this survey German students performed at a much lower level than their counterparts in neighbouring countries. This again led to headline news stories and debate in the media. This time however the discussion focussed almost entirely on the poor standards and what to do about them rather than on whether the statistics were correct.

Example 3

The third example relates to the UK participation in PISA 2003 when the UK narrowly missed the minimum response rates required. As it was anticipated that the response rate would be difficult to achieve additional information was collected during fieldwork in order to be able to look at bias in the sample. The UK was the only country who was able to demonstrate the existence and extent of bias in the sample. Despite this, the UK results were not included in the main analyses as they were deemed to have failed the minimum quality criteria.

Events such as these, where national pride is at stake, means that international comparisons are subject to much more scrutiny than previously. It is important that your analysis is sound, particularly if comparing the relative performance of different countries.

3.5.2. Choosing your analysis platform

Once you've determined your research question and your data source the next step is to determine your preferred analysis tools.

Web analyser provided by the survey host

Many of the international data sets offer a facility to conduct analysis online using a web analyser. This is suitable for most analyses where you are happy with the classification of variables used. These sites also offer guidance on how to use the web analyser, for example on the ESS there is a helpful overview of **Nesstar**

functions http://ess.nsd.uib.no/files/IntroNesstar3.5.pdf with further links to more detailed guidance.

To access the web analyser for the European Social Survey go to
http://ess.nsd.uib.no/webview/index.jsp
To access the web analyser for the World Values Survey go to
http://margaux.grandvinum.se/SebTest/wvs/index_data_analysis

Proprietary analysis package such as SPSS, Stata, SAS, HLM

For many analyses however you will want to bring the data into a specialist package where you have more flexibility to manipulate the data. Unit 1 in this series has already covered some of the issues in downloading into specific packages.

3.6. Applying weights

In most surveys you will need to apply weights before analysis and there are different weights applicable depending on what variables you are using and the units of comparison. The weighting structures will vary depending on the design of the survey, for example if only one person was interviewed per household then people who live in smaller households have a greater probability of being selected than individuals who live in larger households. To understand the weighting structure you need to familiarise yourself with any hierarchies in the data structure – how many levels, are variables carried over from different levels of the hierarchy – for example, aggregates of individual household members income brought across to the household level or vice versa.

The technical documentation for the survey will tell you everything you need to know about how to weight the data. Where there are different weights that need to be applied at the same time these are sometimes presented as compound weights –a single weight variable that adjusts for probability of selection and to adjust for nonresponse or to gross up to population totals all of which may also be available as individual weights.

> Remember that regardless of what weights you need to apply for your analysis you will always need to also run the same analysis without weights in order to get the unweighted cell sizes and bases for inclusion in the tables.

3.6.1. Example – the European Social Survey has two weights

The Design weight adjusts for unequal probability of selection within the household - this is simply applying the reverse of the probability of selection. The sample selection for the survey is based on households; however, only one person was interviewed per household. This means that people who lived in households where there was more than one eligible adult had a different (lower) probability of being interviewed than those who lived alone. The design weight simply makes an adjustment to compensate for the unequal probability of selection.

The Population size weight adjusts for the size of the population in a country. In most surveys the sample size is not related to population of the country so many of the samples are around the same size. Without applying a population weight the sample from small countries would contribute the same value as the sample

from larger countries. The population weight adjusts the relative contributions in proportion to the size of the country.

See example in Citizens spreadsheet available to download from LIMMD website

Which weights need to be applied depends on the analysis being conducted. When making comparisons between countries, using distributions and without reference to the combined total, then only the design weights are needed. When making comparisons with reference to the average for the combined data then both the design weight and the population weights need to be applied. In any analysis where countries are being grouped together, for example, countries from different geographical regions, or language groups, then design and population weights are required.

A guide to the use of weights on the ESS is available at http://ess.nsd.uib.no/files/WeightingESS.pdf

3.6.2. Working with weights – practical issues

Working with weighted data means you need to go through additional processes. It is good practice to check how the weights are affecting the data by comparing weighted and unweighted distributions. Develop the habit of routinely checking the percentage distribution and the mean and standard deviation. This will give you a sense of where the main corrections are being applied and whether they are different in different countries. See the worked example in the citizens spreadsheet.

It is also good practice to check your bases to make sure you have accounted for all the cases, to alert you to whether you have weights on or off. Checking bases is a really good discipline to establish to check that your analysis is on the correct population. For example, on a whole sample table is the total number of cases the same or close to the original sample? How many cases have been processed. If you are using the population weight then the sample size will be much larger than the number of people interviewed unless the weights have been scaled back. It is useful to keep on a post-it note or somewhere near your computer the numbers you would expect to have in the main categories for weighted and unweighted – e.g. how many women and men, how many in each of the main age-groups used in analysis, how many are in full-time work etc.

3.7. Preparing the data ready for analysis

Many of the international datasets offer a facility to analyse the data online. For straightforward descriptive analysis or for initial exploration or assessment of the data this is probably sufficient for your needs. For more complex analysis you will want to download the microdata for analysis with a statistical software package. While it is possible to analyse many datasets online using It is often the case that the dataset does not contain the variables you wish to use in the right format. For example, you may be interested in a particular cohort but the variable on age is in 5 years bands. You need to create a new age variable with the bands of interest for your analysis. Similarly, you may intend to run a multivariate analysis where you need to set a reference group or create dummy variables so expect to spend some time getting the data in the format necessary for the type of analysis you want to do. For example, you may wish to create a dichotomous variable as to whether someone is religious or not from a question or set of questions on church/place of worship attendance, self categorisation of religiosity or frequency of private prayer. As the norms of collective worship or frequency of prayer differ between religions you will want to try and determine a best fit definition. As the distribution of affiliation to particular religions varies been countries you will need to be sure that where you identify differences between countries you are not simply finding differences between religions and vice versa.

Download and study the ESS Round 2 Gender weighting spreadsheet

ESS2 sample dist gender.xls

These tables show the effect of weighting on the gender composition of the sample and illustrate what the population weight does. This reduces the contribution of the sample for any given country pro rata to its size.

Download and study the spreadsheet on interview length.

interview length.xls

These tables show the effect of weighting on interview length - the first table is unweighted, the second has design weight and the third has the design and population weight.

Before conducting any analysis stop and ask yourself if the variable meets the requirements for that analysis technique. For example, does the analysis assume

that the data is normally distributed? Does the data need to be continuous? Is there a requirement of independence?

3.7.1. Organising and documenting your analysis

It is important to develop good habits in organising and saving your work. This is particularly true when working on large datasets.

The need to be well organised when handling large datasets is particularly important. The difference in time to run analysis on a dataset containing 250000 cases and 3000 variables requires a different approach to analysing a file with 2000 cases and 350 variables. Not being organised will mean that you spend a lot of time waiting for the process to complete.
An important element of any analysis is reproducibility – can you reproduce the analysis and get the same result. With many analysis packages offering browser type interfaces where you choose actions from drop down menus this can be difficult, particularly if the analysis involves a lot of transformations of variable.

Most analysts only use menu driven options when they want to quickly answer a very simple question such as what is the basic distribution of the variable in the population or to check cell sizes. Developing the habit of working with syntax (a programme specification for the analysis) rather than menu clicks will make your analysis much more efficient in the long run. For example, if you need to rerun the analysis for any reason, the steps in the analysis are already set out in the programme rather than trying to redo all the steps in sequence.

Example: SPSS Syntax

```
freq vars=cntry freehms prtyban euftf.

* you can insert comments into your syntax to document what you are
doing.
* an asterisk at beginning means it will be recognised as a comment
not a command

CROSSTABS
/TABLES=cntry BY freehms prtyban
/CELLS= COUNT ROW.
FREQUENCIES
VARIABLES=freehms.
EXECUTE.
```

```
* create new variable grouping into 3 categories rather than 5

RECODE
freehms
(1 thru 2=1) (3=2) (4 thru 5=3) (SYSMIS=Copy) INTO freehms3 .
VARIABLE LABELS freehms3 'Gays and lesbians free to live as they
wish 3 category'.
VALUE LABELS freehms3
1 'Agree or agree strongly'
2 'neither'
3 'disagree or stronger' .freq vars=freehms3.
EXECUTE .
FREQUENCIES
VARIABLES=freehms freehms3.
EXECUTE.
FREQUENCIES
VARIABLES=freehms freehms3.

* apply the design weight, run the tables and then remove the
weight.

WEIGHT
BY dweight .

CROSSTABS
/TABLES=cntry BY freehms freehms3
/CELLS= COUNT ROW.

WEIGHT
OFF.
```

3.7.2. Some do's and don't in managing your analysis

- Save the syntax for any derived variables
- **Always save any derived variables** to the dataset once you are happy with them so that you don't have to keep creating them.
- Develop a set of conventions for naming variables such as Age5 – age in 5 bands, Age7 – age in 7 bands etc.
- Develop a set of conventions for naming your files so that you can quickly find the syntax for the analysis you are looking for – for example adding a number at the end to show which the latest version is.
- **Always switch the weights** off before saving the data file.
- **Always give the datafile a new name** so as not to permanently overwrite the previous file.
- In handling very large dataset you may want to break the file up into smaller units – for this you need to think about what variables will be pertinent for the analysis and what classificatory variables you need.

4.8. Documenting your analysis

All syntax files should be saved in a way that is meaningful and that you can quickly find the files you need. It is amazing how quickly files can proliferate as you start to try different pieces of analysis and how quickly you can lose track of which file ran which bit of analysis. It is helpful to develop a routine for how you approach your use of software. A suggested format for an analysis log is below

Log of programmes – ESS Year 1 data

Date	**Programme File name** (*.sps)	**Data file** (*.sys)	**Saved file** (*.sys)	**Function**	**Analyst**
20 July 06	Data1.sps	Spss version of ESS *.sys	ESSmain 1.sys	Gets issued core datafile for all countries from ESRC, extracts the data for the countries to be included in the analysis and extracts only the variables to used including any classificatory variables.	
21 July 06	PolAtt1.sps	ESSmain1.sys	ESSmain 2.sys	Creates additional derived variables needed for analysis for political attitudes chapter	
21 July 06	PolAtt2.sps	ESSMain2.sys		Applies weight (neweight1), table run for main political attitudes and correlations	
22 July 06	PolAtt3.sps	ESSMain2.sys		Applies weight, Multivariate analysis for political attitudes chapter	
22 July 06	SocAtt1.sps	ESSMain2.sys	ESSMain 3.sys	Creates and saves derived variables for social attitudes	

3.8.1 Data Analysis

Once you understand the data and have organised it ready for analysis the rest is straightforward and not that much different to analysis in a single country context. The same tools and methods you would use in analysis of a domestic survey can be applied. The ESS has a helpful analysis support **ESS EduNet** which provides examples and a glossary of statistical terms. (click data then click on Methodological glossary) It has two worked examples on "Social and Political Trust" and "Human Values" both of which take you through the analytical process.

3.8.2. Setting the context

> It is important to remember that the reader may not have the same understanding of the topic as you or the same knowledge about the countries being analysed.

Setting the context and bringing the reader with you are useful first steps in addressing your research question. Providing simple descriptive analysis of the situation relevant to the topic of interest helps the reader follow where you are going.

Before using any variable in analysis you should look at how it is distributed across the countries. Applying a "common sense" test is a good starting point for investigating any difficulties with the data. Does the distribution fit with your expectations? Are there similar patterns in countries that share a common language or that have similar characteristics? For example, does the distribution in the Nordic countries look similar or the Club Med countries? If there is an outlier, is this an outlier or an error?

> The usual rule of thumb is that if it looks wrong then it probably is wrong! Satisfy yourself that any unexpected variation is real and not due to the survey process.

Aside from applying your common sense what external data can you use to check that the data is plausible? For example, if you think the variable of interest might be associated with age or with education then you can check that the distribution observed is not simply a result of the differences in the age and education profile of the country. Check the correlates in each country and compare with aggregate data available.

Example **International Adult Literacy Survey** – expectations based on distribution of related characteristics in the population.

Age and education are known to be correlated as access to education has changed over recent decades and both age and education are independently correlated with performance on the direct assessments of literacy. How countries perform relative to each other will partly be a function of how age and education and other characteristics are distributed in the population. The distribution of the

population in terms of age is not the same in each country. For example, Germany has a higher percentage of its population in the oldest age-group than other countries while the US distribution tends toward the younger end of the age range. If all other things were equal you would expect Germany to have a greater proportion at the lower literacy levels than the US, just because of their age distribution. Similarly in comparing the distribution of educational attainment the Netherlands had a much higher proportion educated to second level, 1st stage (ISCED 2) than the other countries. Canada and the United States had higher proportions of their population who had completed third level education at that time. Almost one in four Americans (24%) aged 25-64 had completed a university programme compared with just 12% of the UK population in the same age range and 13% of Germans.

These differences were not simply due to the different age profiles of the countries, there were differences in educational attainment within age-groups or cohorts. Table 6.4 shows the percentage who had attained at least upper secondary education (ISCED 3) in each country for different age-groups. Among the older age-groups, 84% of those aged 45-54 in Germany and 85% of the same age group in United States had completed at least upper secondary education compared with about 70% of the population of Canadian, Swedish and British population. Apart from the Netherlands the proportion of those aged 25-34 who attained at least upper secondary education in each country is fairly similar ranging from 82% of Canadians to 90% of Germans in that age-group.

It is evident from the distribution of literacy skills as assessed in the survey that the differing age and education distributions alone do not explain differences in the profile of literacy skills between countries. Although the Swedish population was very similar to that of Britain both in terms of the age profile of the labour force and the distribution of educational attainment, even across age-groups, the distribution of literacy skills in the two countries is quite different. Sweden had larger proportions of people performing at the higher literacy levels in each age-group and in each level of education compared with Britain.

The opposite phenomenon was also evident when comparing Britain and the US. Both countries had similar age profiles and in all but the youngest age-group the US had higher proportions of the population who had third level education yet the literacy profiles of the two countries were very similar.

When education was held constant there was no consistent pattern across countries other than Sweden who consistently outperformed those in other countries.

This type of external benchmarking and setting the context in which the observed measures can be set is useful in understanding the observed differences. Without it the reader doesn't know whether the measures are simply a reflection of the characteristics of the population in those countries.

You may want to go on line and download and study the file **IALS_tables 6.2_6.3.xls** which contains IALS tables 6.2, 6.3 and 6.4 from Carey, S et al, (1998) Adult Literacy in Britain. London: The Stationery Office in MS Excel worksheet format.

In many international surveys the analysis of the correlates of the distribution are often more illuminating than the comparison of the distribution alone. For example it may be more interesting to look at whether the relationships between education and literacy is of the same power across countries rather than focusing on which countries have higher scores than others.

Example: How to produce a table showing - attitudes to citizens cheating on their taxes by country

This is a worked example using the ESS to produce a table showing the distribution of key characteristics for a selection of countries of interest with and without weights for attitudes to the statement **"Citizens should not cheat on their taxes"**. You will need to go online to follow this activity – You will need to register on line for ESS Data to complete this task.

1. Go to ESS Round 2 tab - Online browsing and analysis http://ess.nsd.uib.no/. Two datasets are available in left hand panel
2. Click on **data icon** for file ESS2- 2004, ed. 2.0
3. Expand **variable description > Economic morality > Citizens should not cheat on taxes**
4. In right hand panel, click on **Tabulation** tab
5. Left hand panel click on **variable** to tabulate
6. Left click **Add to column** or **Add to row** (if wanting to include a

large number of countries best to choose **Add to column**). You will be prompted to login if you haven't already done so

7. In left hand panel expand variable country
   Country (right click **Add to row**)

This will produce a table with the countries down the left-hand stub and the responses to the questions across the top. The default for percentage calculation is column percent. In this instance we want row percent.

8. Click on the third dropdown menu in the right hand panel **TYPE** and change to row percent.

This produces a table showing the distribution of the variable for each country on **unweighted data**. The N in the right hand column is the number of cases in each country. Comparisons made using this table would be invalid.

9. To weight the data click on the **balance icon**. In the left hand box click on design weight so that it is highlighted, select by clicking the **>** so that it appears in the right hand box, then click **OK**.

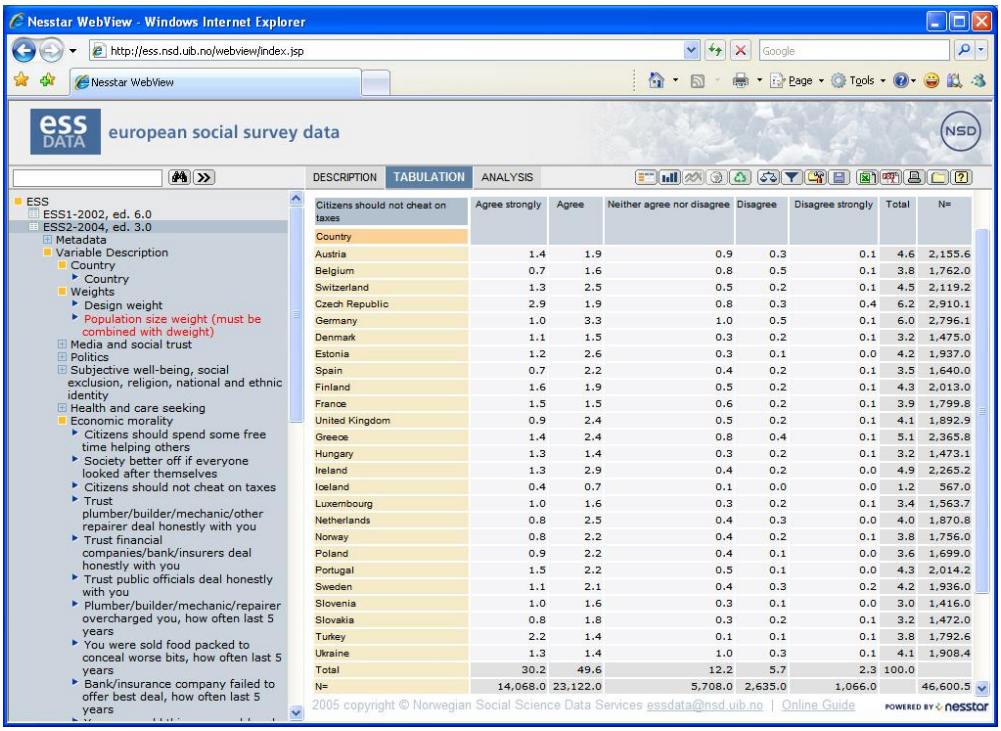


Figure 1: Generating a table with ESS Data

You can also limit the number of categories displayed using the drop

down menu for the variable, selecting Choose categories and ticking which categories you want to include, for example, Agree strongly and agree and click Update, or choosing a smaller number of countries. The table now is more readable but pay attention to how the percentages are calculated.

Deciding how to display your data will depend on what the data shows.

Activity 2: Attitudes to whether healthy people should use medicines to improve their sex life - by country.
Using the same steps outlined above produce a table showing response to the question **"do you approve if healthy people use medicines to improve their sex life…"** .

To find this data … look in **variable description > Health and Care Seeking > Approve if healthy people use medicines to improve sex life.** You will need to register on line for ESS Data to complete this task.

Activity 3: Produce graphs showing - attitudes to citizens cheating on their taxes by country, weighted and unweighted
Using the examples above produce **graphs** showing responses to the question *"Citizens should not cheat on their taxes"* before looking at our results. You will need to register on line for ESS Data to complete this task.

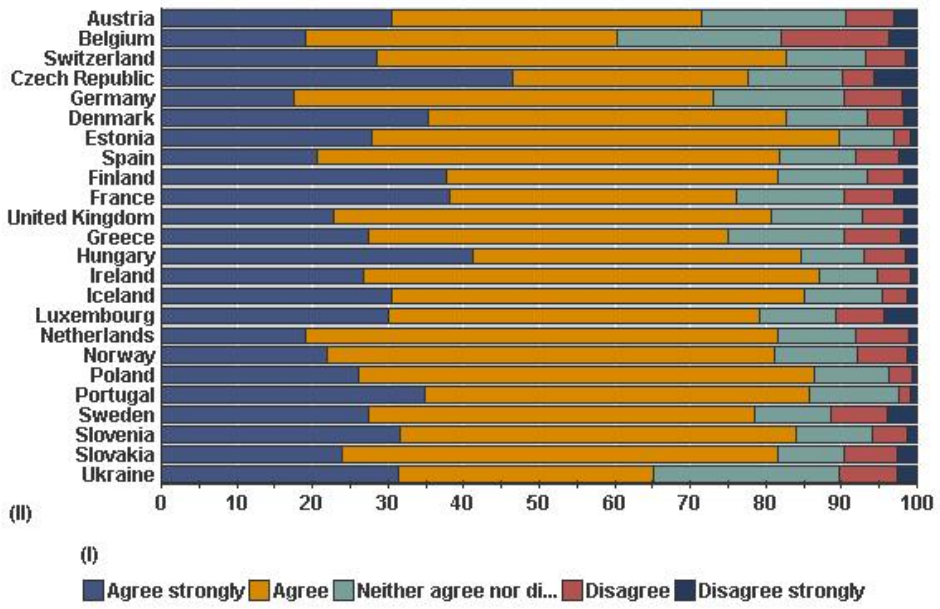
Task 1
Produce graphs showing the weighted and unweighted distributions of the proportions who agree or strongly agree with the statement *"Citizens should not cheat on their taxes"* Graph showing proportions who agree or strongly agree combined

Task 2
Display same graph with the countries presented in ascending order of support for the statement rather than alphabetically.

Task 3
Using the same data graph the distribution who agree or strongly disagree with the statement. Do you get the same story?

Results

Your chart with weights on should look something like this.



Comment on graphs: In this example both Belgium and the Ukraine have lower proportions who support the statement about paying taxes. Looking at the corollary however, it is clear that Ukraine is not so different from other countries. Both Belgium and the Ukraine had higher proportions of respondents who said they neither agreed nor disagreed.

3.8.3 Other resources

There are a large number of other international datasets that might be of interest, particularly in the field of education and health.

The **International Adult Literacy Survey (IALS)** is a direct assessment of the literacy and numeracy skills of adults in more than 30 countries, many of them OECD members.

The **Programme of International Student Achievement (PISA)** is a direct assessment of the reading, maths and science skills of students in the last year of compulsory education.

Other international studies of educational attainment are those conducted by the **IEA** such as **TIMSS**, **PIRLS** which look at the literacy skills of younger cohorts of students as well as science and mathematics.

In the field of Health there is the **World Health Surveys**, the **World Mental Health Surveys** and the **Demographic and Health Surveys** is your focus is on developing countries.

3.9. Presenting your results

Presenting analysis of international datasets offers additional complexities over single country analysis. Tables and graphs that display anything other than simple distributions for a range of countries can quickly get very busy and difficult to read. It is often useful to look to publications produced by international organisations such as the **OECD**, the **World Bank** or the **UN** who regularly have to summarise complex information on a large number of countries in a visual way.

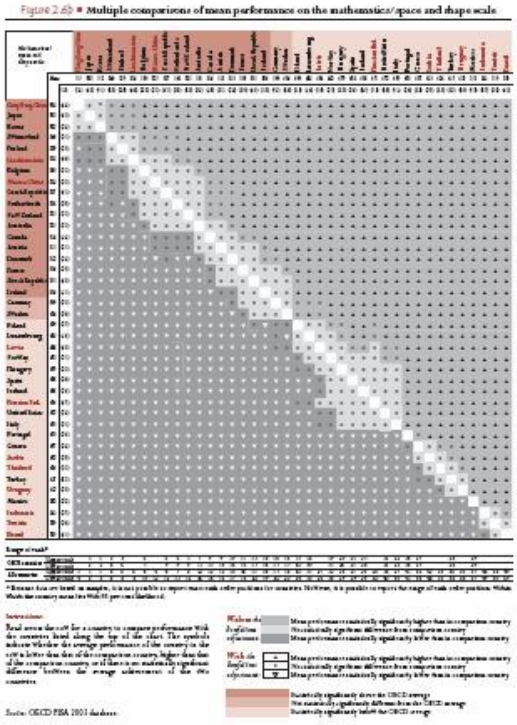
For example, the results from the PISA study convey both the overall score for a country but also show whether it is significantly different from all the other countries presented. Similarly, the UN agencies such as the UN Statistics Division, UNESCO Institute of Statistics, the World Health Organisation, the International Labour Organisation and the Food and Agriculture Organisation all provide examples of how to present data comparisons across countries.

http://www.oecd.org/statsportal/0,2639,en_2825_293564_1_1_1_1_1,00.html

3.9.1 Multiple comparisons

Making multiple comparisons of country means that the likelihood of overestimating significant results is increased. You might want to make an additional adjustment such as those used in PISA -

http://en.wikipedia.org/wiki/Multiple_comparisons
http://www.pisa.oecd.org/dataoecd/58/41/33917867.pdf



3.10 Further exploration

Activity 4: further exploration – victims of crime

Using the **ESS** looking at the victims of crime and feeling safe walking in neighbourhood.

Does it fit with your expectation?

Do the Nordics look high – if so why might that be?

You will need to go online to follow this activity – You may need your Athens username and password to complete this task.

Activity 5: further exploration – attitude to migration

Suggested area to explore using the **Eurobarometer** – look at the survey of accession countries questions on **attitude to migration**.

You will need to go online to follow this activity – You may need your Athens username and password to complete this task.

Activity 6: further exploration – confidence in different institutions

Suggested area to explore using the **World Values Survey** –

1. Look at survey information about confidence in different institutions – examine the correlates of questions in some developing countries to see if there are differences.
2. Look at Ireland, France, Spain, and Portugal for 1990 and 1999 to see if the values have changed over time and whether correlates have changed over time.
    a. **Is the change driven by demographic change?**
    b. **Are there gender or rural/urban differences?**
    c. A001 to A007 – examine difference and suggest analysis for investigating further.
3. Look at Church, police, and civil service.
    a. Is the self-positioning in political scale (EO69) consistent with views expressed elsewhere?

You will need to go online to follow this activity – You may need your Athens username and password to complete this task.

An example of analysis which uses both the ESS and the Eurobarometer is

http://eumc.europa.eu/eumc/material/pub/eurobarometer/EB2005/EB2005-summary.pdf

3.11. Suggested reading and resources

http://www.statsoft.com/textbook/glosfra.html

http://stats.oecd.org/glossary/

http://www.gesis.org/en/research/eccs/csdi//

Dillman, D., J. Eltinge, Robert M. Groves, R. Little. "Survey Nonresponse in Design, Data Collection, and Analysis." In Survey Nonresponse edited by R. Groves, D. Dillman, J. Eltinge, and R. Little. Pp. 3-26. New York: Wiley. 2002.

Groves, Robert M., F.J. Fowler, Mick P. Couper, James M. Lepkowski, Eleanor Singer, R. Tourangeau. Survey Methodology. New York: Wiley. 2004.

Groves, Robert M., Mick P. Couper. Nonresponse in Household Interview Surveys. New York: Wiley. 1998.

Tourngeau, R. Rips, L.J., and Rasinski, K (2000). The Psychology of Survey Response. Cambridge: Cambridge University Press.

ZUMA-Nachrichten Spezial, Vol 11 Methodological Aspects in Cross-National Research Jürgen H.P. Hoffmeyer-Zlotnik & Janet A. Harkness (Eds.) Mannheim: ZUMA 2005, 305 Seiten, ISBN 3-924220-29-8, 13 Euro

Harkness, J. (2004) "Overview of Problems in Establishing Conceptually Equivalent Health Definitions across Multiple Cultural Groups", in: Cohen, S.B. and Lepkowski, J.M. (eds.), Eighth Conference on Health Survey Research Methods,Hyattsville: US Department of Health and HumanServices: 85-90.

Harkness, J. (2004) Guidelines for the ESS translation in Round two (ESS web).

Harkness, J., Pennell, B.-E. and Schoua-Glusberg, A. (2004) "Questionnaire Translation and Assessment", in: Presser, S., Rothgeb, J., Couper, M., Lessler, J., Martin, J. and Singer, E. (eds.), Methods for Testing and Evaluating Survey Questionnaires, New Jersey: John Wiley and Sons: 453-473.

Harkness, Janet A., Van de Vijver, F.J.R. and Mohler, P.Ph.(eds.), Cross-Cultural Survey Methods,New Jersey: John Wiley and Sons Inc: 35-56.