

Community
Appraisal
& Motivation
Programme



Researching Forced Migration in Pakistan:

An Introduction to Research Ethics,
Quantitative & Qualitative Methods



Researching Forced Migration in Pakistan:

An Introduction to Research Ethics, Quantitative & Qualitative Methods

Community Appraisal & Motivation Programme, Islamabad

Financial Support provided by
The Foreign Office, Federal Republic of Germany, Islamabad



Copyright © CAMP, 2011.
Community Appraisal & Motivation Programme [CAMP]

Researching Forced Migration in Pakistan: An Introduction to Research Ethics, Quantitative & Qualitative Methods

ISBN: 987-978-969-9534-02-7

Copies are available from, and feedback may be submitted to:

CAMP Head Office

House No. 217, Street 39, F-11/3

Islamabad, Pakistan

Tel: +92-51-222-5196/97

Fax: +92-51-222-5198

Email: publications@camp.org.pk & camp@camp.org.pk

Download: www.camp.org.pk

All rights reserved. The material contained in this report may be freely quoted, reprinted or translated for academic, educational or other non-commercial purposes without prior permission from CAMP, provided that the source is acknowledged.

Disclaimer: The views expressed in this report are solely those of the author and do not necessarily represent those of CAMP, or The Foreign Office of the Federal Republic of Germany.

Table of Contents

Acknowledgements	1
Glossary	2
List of Boxes, Tables and Figures	6
Module 1. Getting Started	7
Module 2. Research Ethics	13
Module 3. Quantitative Methods: Developing and Administering Questionnaires	27
Module 4. Quantitative Methods: Ensuring Scientific Accuracy	51
Module 5. Qualitative Methods of Data Collection	73
Module 6. Analysis and Presentation of Quantitative Data	91
Module 7. Analysis of Qualitative Data	109
Appendix 1. Migration Related Academic Journals	119
Appendix 2. List of International Migration Research Institutes	120
Appendix 3. Z-Score Table (or Table of Areas under the Curve)	122
Appendix 4. Student's 'T' distribution	123
Appendix 5. Bibliography on Forced Migration in Afghanistan and Pakistan	124

Acknowledgements

CAMP would like to acknowledge the contribution of Ms. Ayesha Khan and Ms. Saman Hamid Qureshi in the compilation of this Manual. In addition, several CAMP staff members contributed to the success of the project, in particular Chief Executive Mr. Naveed Ahmad Shinwari, Mr. Mohammed Bhehzad Khan, Ms. Fareeha Sultan, Mr. Muhammad Rafiq and Mr. Usama Khilji.

CAMP benefitted greatly from the expertise and feedback of our technical support partner, the Max Planck Institute for Comparative Public Law and International Law in Heidelberg, Germany, especially Dr. Tilmann Röder, Ms. Sabiha Beg, and the rest of the Max Planck team.

We also graciously acknowledge and appreciate the financial support of our funder, The Foreign Office of the Federal Republic of Germany, and particularly Ambassador Dr. Michael Koch & Second Secretary Mr. Ferdinand Jenrich.

We hope that you will find this manual to be a useful reference as you embark on research on forced migration in Pakistan, and we look forward to reading your future research publications.

Glossary

Quantitative Methodology Terms

Causal Relationship/ Causality	A relationship between two variables where one variable (the dependent variable) is shown to be the outcome/consequence of another variable (the independent variable)
Confidence Interval	Estimated interval within which the population parameter should be found. It indicates the reliability of an estimate and is calculated based on the confidence level (usually 95% or 99%).
Correlation	A measure of relationship between two variables. The coefficient of correlation describes the magnitude and direction of relationship between two or more variables. Correlation can be negative or positive.
Dependent Variable	The outcome of interest that is affected by changes in independent variables.
Dummy Variable	A dummy variable is a numerical variable used in regression analysis to represent subgroups of the sample in your study. These variables are nominal i.e. they can either have a value of 1 or 0, which means that a condition is either present (1) or absent (0).
External Validity	The degree to which the findings of a study can be generalized to similar populations in different places at different times.
Independent Variable	Variables that can be changed, manipulated in a study to change the outcome or dependent variable.
Mean	This is the average value for a variable, i.e. the sum of individual scores divided by the number of scores.
Median	This is the mid-point of the distribution of any variable, i.e. there are equal numbers of scores (observations) above and below this point.
Mode	This is value or attribute that occurs most frequently for a variable.
Non-Sampling Errors	Errors in a survey that can result from errors in measurement, data recording and/or data processing
Regression Analysis	The objective of this analysis is to explore the relationship between a dependent variable and one or more independent variables. More specifically the analysis helps one predict if and how the value of the dependent variable changes when any one of the independent variables is varied, while other independent variables are held fixed.

Sampling Errors	Sampling errors stem from the fact that data is collected from a sample rather than the whole population.
Sample Frame	A list of the units of analysis from which a sample is drawn. The term is often used interchangeably with population, however the frame is generally smaller than and a part of the total population of interest.
Standard Deviation	Measure of how much the scores in a distribution vary from the mean (average) score.
Standard Error or Standard Error of Mean	The standard deviation of the mean of a particular sample from the overall population mean. In other words the error caused by the use of the sample mean to approximate the population mean.
T-test	The t-test is the most commonly used method to evaluate the differences in means between two groups. The groups that are being tested can be independent or dependent.
Unit of Analysis	A list of the units of analysis (individuals, households, neighborhoods, etc.) from which a random sample can be drawn.
Variables	Features or properties that vary from one respondent, situation, geographical location to the other. There can be 3 types of variables; nominal, ordinal and interval or ratio.

Qualitative Methodology Terms¹

Action Research	A multi-stage type of research, in which a problem is researched, changes are made, the problem is researched again, more changes are made, and so on through a number of cycles, until the problem is solved.
Case study	A type of qualitative research which studies one or a few cases (people or organizations) in great detail.
Coding	Labelling a piece of text or a statement, to make sense of it by summarizing it. Depending on the research question, one piece of text can be coded in various different ways.
Community Profiling	Detailed social, political and economic profile for each locality selected for fieldwork.
Confidentiality	To respect the confidential nature of the information gathered during the research and preserve the anonymity of participants.
Cooperative Inquiry	A type of participatory action research in which all involved, including the participants, work as co-researchers to generate ideas, plan and emerge with conclusions from the research experience.

¹ The definition of terms is adapted from <http://www.audiencedialogue.net/gloss-qual.html> and http://www.hello.nhs.uk/documents/Qualitative_Critical_Appraisal_Glossary.pdf

Data Analysis	A systematic process of working with the data to provide an understanding of the participants' experiences.
Ethnography	A type of qualitative research which treats a group of people as an anthropologist would treat an unknown tribe, with detailed descriptions of how they live.
Facilitation	Helping a group of people come to conclusions, done by a facilitator. In group discussions, this role is called a moderator.
Field Notes	Written record of what the researcher has observed while in the field.
Field Observation	The observation of participants in their natural setting. Observation can be direct or indirect, participatory or non-participatory.
Focus Group	A common type of group discussion in which a moderator encourages a small group of people (usually 8 to 10) to gradually focus on a topic.
Grounded Theory	An approach to qualitative research where the researchers try to approach a problem with no preconceptions, and to build their theories solely from the data gathered.
In-Depth Interview	A type of qualitative research which involves long, probing interviews without the use of a formal questionnaire. Can be semi-structured or unstructured interviews.
Informal Group Discussion	Similar to a focus group, but smaller and spontaneous, without formal moderator and shorter.
Informed Consent	Participants consent as well as respect for their right to refuse involvement in the research. Participants must understand the nature of the project, the procedures that will be used and the use of the research findings.
Interaction	Spontaneous, brief conversation in the field setting which is recorded in the field notes.
Interview Guide	A list of topics to be covered in an interview. Similar to a questionnaire, but much less structured, and without multiple-response questions. Used in in-depth interviews, focus group discussions, other group discussions.
Key Informant Interview	Scheduled, formal interview based on guidelines to explore details of the community profile and initial reaction to research questions from the perspective of someone very well-versed with the community and its socio-economic context.
Mapping	Introductory exercise in a community to invite a group of participants to draw out their shared vision of a response to a research question.
Moderator	The researcher who leads a focus group.

Open-Ended Question	A type of question where it is left up to the respondent to volunteer an answer.
Oral History	A method of qualitative research that focuses on interviewing specific individuals about their lives for the purpose of creating a historical narrative.
Participant Observation	Simple non-interactive observation of phenomena in the setting of the research project.
Reflexivity	Acknowledgement by the researcher of the central role they play in the research process. A reflexive approach considers and makes explicit the effect the researcher may have had on the research findings.
Respondent Profile	Basic data collected on individuals interviewed, e.g. age, name, address, education level, number of children, employment, etc.
Short Interview	Similar to an in-depth interview but around half an hour long only. Useful for triangulation of data.
Thick Description	Thick description refers to the researcher's task of both describing and interpreting observed social action (or behavior) within its particular context. ²
Triangulation	The process by which the area under investigation is looked at from different perspectives. These can include two or more methods, sample groups or investigators. Used to ensure that the understanding of an area is as complete as possible or to confirm interpretation through the comparison of different data sources.
Visualization	A participatory research method in which participants literally illustrate in their own terms how they conceptualize a given idea or subject.

² Ponterotto (2006:543) [See References in Module 5]

List of Boxes, Tables and Figures

Box 2.1 Research Ethics Checklist
Box 2.2 Sample Written Consent Form
Box 3.1 Stages in Survey Design
Box 3.2 Summary Literature Review Table
Box 3.3 Research Designs
Box 3.4 Relating Variables to Research Questions/Hypothesis
Box 3.5 Sample Survey Budget
Box 3.6 Sample Codebook
Box 3.7 Data Entry Template
Box 3.8 Survey Inference
Box 4.1 Sample Size Table
Box 4.2 Calculation of Null Hypothesis
Box 5.1 Community Profile
Box 5.2 Respondent Profile
Box 5.3 Interview Guidelines for Study on Afghan Refugee Livelihood Strategies
Box 6.1 Interpreting Cross-Tab Tables
Box 6.2 Example of t-test in SPSS

Table 4.1 Variables and Attributes
Table 4.2 Population and Proportion of Population Represented
Table 4.3 Calculation of Weights
Table 4.4 Marital Status
Table 4.5 Descriptive Statistics (1)
Table 4.6 Descriptive Statistics (2)
Table 6.1 CNIC * Gender Cross-Tabulation
Table 6.2 CNIC * Gender * Age Groups Cross-Tabulation
Table 6.3 Correlations (1)
Table 6.4 Correlations (2)
Table 6.5 Model Summary (1)
Table 6.6 Anova (1)
Table 6.7 Coefficients (1)
Table 6.8 Model Summary (2)
Table 6.9 Anova (2)
Table 6.10 Coefficient s (2)

Figure 4.1. Percent of Normal Distribution Scores in Each Interval



Module

1

Getting Started

Module 1 – Getting Started

1.1 Why Do We Need Research?

Pakistan has had one of the longest experiences with displaced populations in the world, especially since Afghans began to enter the country in the late 1970s to escape conflict in their homeland. Much has been learned - regarding how to manage large-scale displacement - by the Government of Pakistan, UNHCR and other UN agencies, as well as international and domestic non-government organizations that have all been providing relief and assistance to the Afghan population for over three decades. Knowledge about Afghans' experiences of displacement, their ongoing needs and particular vulnerabilities, as well as challenges with respect to implementation of durable solutions exists on many levels, although there remain gaps in information as well.

Over the last six years, Pakistan has also faced the challenges of managing internally displaced populations, dislocated from their homes due to natural disasters or armed conflict. This includes survivors of the disaster-induced displacement such as the Kashmir earthquakes of October 2005, the 2010 and 2011 floods, as well as survivors of conflict-related displacement in some districts of Balochistan, and waves of similar displacement from Malakand Division and the Federally Administered Tribal Areas beginning in 2009. It appears that the Pakistani army and intelligence agencies will continue to be engaged in conflict with militant groups in the tribal and adjacent areas in Pakistan's northwest for the time being.

Empirically-grounded research is, therefore, an essential component of developing policies and programmes that are relevant and meet the established needs of affected populations. Research can inform policies and programmes in a variety of ways. A rapid needs assessment, for example, can gather data over a short period of time to determine the needs of displaced populations in an emergency situation. Census surveys of camps provide invaluable demographic data that can help planners determine future needs with respect to durable solutions. Well-documented interactions in the form of interviews with especially vulnerable populations, such as female-headed households, can provide valuable insight into how best to reduce their exposure to risk. And literature reviews that summarize and present the key findings of existing research on a given subject are vital to identifying research gaps and informing future research directions.

But for research to serve its purpose best, it needs to be part of an institutionalized agenda, one that systematically documents challenges and lessons learned from the field with regard to the experience of displacement, and whose findings are shared with the wider community of stakeholders. This would in turn strengthen policy and programme agendas pursued by governmental, non-governmental and multilateral institutions. As it stands today, we know more about some aspects of the displacement experience and the needs of the displaced and less about others, so that it has proven difficult to establish common standards and promote cross-sectoral best practices and collaboration with respect to policy, programming and advocacy initiatives. As Goodhand (2000) writes, there is a need for more expert programmes and interventions based on rigorous and in-depth analysis.

One obstacle to the development of research-based knowledge has been that issues of displacement in Pakistan are often linked with violent armed conflict, and the politics that gave rise to such conflict continue to pose risks for field-based researchers and participants.

Limited access to information and statistics maintained by government sources, and lack of government support in advancing social research present obstacles for independent researchers.

There is also limited capacity among stakeholders working in the field to conduct the research that is both relevant to their immediate programming, and of consistently high quality. Moreover, there has been insufficient coordination among organizations to ensure that existing research findings, challenges and tools are shared, so as to develop common standards and avoid unnecessary duplication of research efforts. Complex social problems are not always clearly defined in order to ensure that researchers investigate them using consistent criteria. Where research does take place, it receives limited publicity so that it rarely informs broader public policy development beyond the specific purpose for which the research was originally commissioned.

In the face of all these difficulties, though, if research leads to an uncovering of truth, often hidden by the complexity of conflict, and if it results in a deeper and more nuanced understanding of the participants that leads to potentially positive effects, then it is well worth it. (Goodhand, 2000:14)

In more ideal circumstances, researchers would work separately from service-provision organizations. In Pakistan, however, researchers are often individuals who work for service-oriented organizations, but they are not necessarily trained to conduct research. This is due in part to the weakness of our higher education system, which has not supported social science-related fields of inquiry.

1.2 Existing Sources of Information

There are a range of information sources regarding issues of migration, including forced, voluntary, and illicit migration i.e. human trafficking and smuggling. Such information is available on the international, regional, and national and local levels, in many forms. For example reports published by United Nations agencies – UNHCR, UNODC, ILO etc. - serve as important references for statistical data and rights-based analyses. Specialist multilateral agencies such as the International Organization on Migration, as well as research institutes based at universities located around the world have developed invaluable policy tools and provided much needed technical and analytical insight. The role of non-government organizations and networks in collecting empirical data at the grass-roots level is also extremely valuable. Unfortunately, despite Pakistan's importance as a country that has hosted one of the largest populations of refugees and IDPs in the world, scholarship on migration by Pakistani researchers is disproportionately sparse in international peer-reviewed journals.

In order to bridge this gap, Appendix 1 of the present Manual presents a list of academic journals devoted to the study of migration, and Appendix 2 presents a list of research institutes in different parts of the world engaged in migration research. Academic journals usually present the highest standard of research. Familiarity with the work of migration research institutes can inform and improve our own research here in Pakistan. Moreover, a bibliography on forced migration from Afghanistan/in Pakistan is provided in Appendix 5. It is hoped that the compilation of this information will benefit and facilitate stakeholders interested in advancing a migration research agenda within their institutions, or independently.

The exchange of research and experience must work both ways – not only do we stand to learn much from the work of these research institutes, but they also need strong research from Pakistan that can add to the body of knowledge about an issue – migration - that will continue to dominate the international and domestic political agenda.

1.3 Areas of Research in Migration

A brief overview of the questions and issues that are recommended for research by organizations that work on migration will give you a sense of what could be further studied in the context of Pakistan.

- Legal studies regarding the rights of forced, economic migrants, and undocumented migrants and extent to which they are being met under existing domestic and international legal frameworks.
- The benefits and drawbacks of community-based dispute resolution mechanisms operative among migrant/refugee communities.

- Labour migration agreements and the legalization of transnational livelihood strategies.
- Documentation of migrants' experiences, including the effects of displacement on vulnerable groups e.g. children, women, elderly and persons with disabilities.
- Impact of migrants on local communities, including the economic contribution/impact of migrants, their impact on ethnic/sectarian relations in host states, and the integration of migrants in receiving countries.
- Comparative studies between the immigration and asylum experiences of Pakistan and other developing countries.
- Best practices for the effective combination of research, programming and advocacy.
- Globalization and the emergence of migration as a legal and illicit industry, through private employment agencies and also through human trafficking and smuggling by organized crime groups.
- The interaction between non-state actors and forced migrants in armed conflict.
- The feminization of migration, in particular migration of women for marriage, illicit migration of women i.e. trafficking, smuggling and sexual exploitation, sexism and xenophobia against women migrant workers.
- The impact of internal migration on women's work in the informal sector.
- Disaster-induced migration and response strategies.
- Rural-urban migration.
- Access to citizenship and naturalization of long-term migrants.
- Statelessness.

The opportunities for empirical research are many, what is now required is more local expertise to carry out the research.

1.4 The User of This Manual

This manual is an aid in building local capacity to undertake research on migration. It will describe in some detail the two main types of research methodologies: quantitative and qualitative. The intended users of this manual will be a range of stakeholders working with the populations under consideration, including government actors such as the Commissionerate of Afghan Refugees and the Ministry of States and Frontier Regions (SAFRON). Non-government organization stakeholders have also expressed a need for training to improve their research output, and it is hoped they will benefit from this manual as well.

1.5 Content of the Manual

This Manual is intended to help overcome some of the weaknesses in research discussed above, by providing training in the basic skills that a field-based researcher in the Pakistani context would need. The Manual is divided into Modules, each of which covers an important element or methodology of the research process. The first of these is the topic of research ethics (Module 2), an area that is not often discussed in Pakistan but remains nevertheless important to developing a culture of research that meets international standards and does no harm to any actors involved in the research process. Next, the Manual will introduce quantitative research methodology, and provide an overview of the sample survey process and how to ensure scientific accuracy when using the research tools (Modules 3 and 4). The Manual will subsequently address qualitative research methodology, explain its uses and the tools to be used in the field (Module 5). Module 6 will explain the techniques for the analysis of quantitative data and Module 7 will do the same with respect to qualitative data. The Manual will offer some concluding comments and suggested reading.



Module

2

Research Ethics

Module 2 - Research Ethics

Purpose: To understand the value of research ethics and learn how to apply it to research studies conducted with migrant/displaced populations.

Research design and implementation must be done according to high ethical standards. The main purpose of having one's research project ethically reviewed is, "to protect all groups involved in research: participants, institutions, funders and researchers throughout the lifetime of the research and into the dissemination process."¹ While procedures to ensure such standards are already well developed in the area of medical/health research, they are still evolving in the social sciences. In Pakistan there are a growing number of research organizations that have begun to adhere to formal ethical standards in the conduct of their work.

When research is undertaken to study a vulnerable population, such as Afghan refugees or internally displaced people in Pakistan, special attention needs to be paid to ethics. Researchers will need to answer some difficult questions, such as:

- Have we adequately prepared for our work such that we will do no harm to participants in the field?
- To whom are we accountable if a respondent tells us, or we come to realize, that they are in breach of the law?
- Do we have the means to help those respondents who are in need of urgent medical or legal assistance? If so, to what extent?
- Will the publication of research findings put any group of respondents in danger of political violence or reprisals from other groups?
- If respondents request some compensation for participating in our research, do we deny them despite their conditions of dire poverty?
- Are the results of the research worth the risks involved?

The sections below will explain how a framework for addressing such questions has been developed by researchers internationally.

2.1 Principles of Ethical Research

The key principles of ethical research have evolved from the field of medical research. They are nonetheless adaptable and relevant to research in the social sciences as well.

The six key principles²:

- i. Research should be designed, reviewed and undertaken to ensure integrity, quality and transparency.
- ii. Research staff and participants must normally be informed fully about the purpose, methods and intended possible uses of the research, what their participation in the research entails and what risks and benefits, if any, are involved. Some variation is allowed in very specific research contexts.
- iii. The confidentiality of information supplied by research participants and the anonymity of respondents must be respected.

¹ Economic and Social Research Council. Framework for Research Ethics (Swindon, UK: 2010) at 2.

² Ibid.

- iv. Research participants must take part voluntarily, free from any coercion.
- v. Harm to research participants must be avoided in all instances.
- vi. The independence of research must be clear, and any conflicts of interest or partiality must be explicit.

These are the basic principles that underlie all ethics review processes, whether they are in medical or social science disciplines. Clearly, with a vulnerable population such as conflict-related or disaster-induced migrants, ensuring the safety of the respondents and the confidentiality of information shared must be fundamental to any research involving them. Respondents will rarely have the means at their disposal to counteract the power imbalance between the researchers, such as government staff or representatives of foreign NGOs, which already places them in a disadvantageous position. Researchers will also need to ask themselves whether the study, and the potential risks it creates for the respondents, is absolutely necessary and worth placing them at risk, and whether their security can indeed be guaranteed by the researchers.

It is helpful to study the suggestions for implementing these principles, because they help researchers to understand where the responsibility for implementing them lies. (Comments have been added in italics.)

2.2 Implementing These Principles³

The responsibility for conduct of the research in line with relevant principles rests with the principal investigator and the research/employing organization. *This means that the NGO or government agency carrying out the research will be responsible for having the research proposal ethically reviewed in Pakistan. The researcher is ultimately more responsible than the donor organization in ensuring this takes place.*

The responsibility for ensuring that research is subject to appropriate ethics review, approval and monitoring lies with the research organization receiving the grant to conduct the research, or some of the researchers when it is acting as the coordinator for collaborative research involving more than one organization.

Research organizations should have clear, transparent, appropriate and effective procedures in place for ethics review, approval and governance whenever it is necessary. *This requires some degree of preparation and guidance can be taken from sources such as the ones referenced in this Module. Research proposals should be screened by an ethics review committee.*

Risks should be minimized. *Researchers must determine whether the risks are worth the research being conducted.*

Research should be designed in a way that the dignity and autonomy of research participants is protected and respected at all times.

Ethics review should always be proportionate to the potential risk, whether this involves collecting primary or secondary data.

Whilst the secondary use of some datasets may be relatively uncontroversial, and require only light touch ethics review, novel use of existing data and especially data linkage, as well as some uses of administrative and secure data will raise issues of ethics.

Research involving primary data collection will always raise issues of ethics that must be addressed.

³ Ibid., at 3. [The italicized comments are those of the author of this training manual.]

2.3 How to Determine If Your Study Requires a Formal Ethics Review

Social science researchers in an academic setting in the United Kingdom (as well as other countries) have prepared a checklist of questions potential researchers can ask themselves while designing a proposed study (see Box 2.1). This checklist in Box 2.1 is adapted from the UK checklist and the suggestions made by Goodhand (2000) for ensuring ethics and accountability when conducting research in conflict zones.

2.4 Obtaining Informed Consent

The informed consent of respondents to participate in the research study is integral to ethical research. This means that the respondent needs to have enough information about the study, and time, to consider whether he/she would like to be involved. The information needs to include the purpose of the study, risks and benefits entailed, and the knowledge that the respondent may refuse to continue to participate at any time during the research process.

Box 2.2 is a sample written consent form. You will see that it is very formal and may not always be applicable in the context of conducting research in Pakistan with an illiterate population, and one that may be anxious about any interaction with state agencies due to the politics and insecurity of its predicament. It is very possible that by asking for written informed consent, your respondents may become suspicious about the motives for the research and refuse. Does this mean that you cannot conduct the research?

There is a way around this, which is to change the format of the informed consent. This could involve preparing a written statement covering the same information as provided in the written template, and then obtaining the respondent's verbal consent. In the project report, it would be necessary to include a copy of this statement and an explanation of how consent was obtained. Research with vulnerable populations usually entails a certain amount of social risk, and that should be covered for in all stages of the research process.

Box 2.1 Research Ethics Checklist⁴

		YES	NO
1.	Does the study involve participants who are particularly vulnerable or unable to give informed consent? [e.g. children, people with learning disabilities, students, patients, people who have broken the law]		
2.	Does the study involve:		
	a. Social risk		
	b. Physical risk to participants		
	c. Psychological risks to participants, in particular the problems associated with recalling trauma during the interview.		
	d. Discomfort to subjects		
	e. Invasion of privacy		
3.	Will financial inducements (other than reasonable expenses and compensation for time) be offered to participants?		
4.	Will the study involve exposure to conflict violence?		
5.	Will the study require the co-operation of a gatekeeper for initial access to the groups or individuals to be recruited (e.g. students, members of NGOs, community-level participants, government officials)?		
6.	Will it be necessary for participants to take part in the study without their knowledge and consent at the time (e.g. covert observation of people in public places)?		
7.	Will the study involve discussion of sensitive topics (sexual activity, drug use, political rivalries, war crimes, or illegal entry into Pakistan)?		
8.	Will the study involve any invasive or potentially harmful procedures, or the administering of any substances (drugs, foods, supplements)?		
9.	Are there any security risks to the researchers themselves, if so does the research warrant the risks involved?		
10.	Will the study involve prolonged or repetitive testing?		
11.	Will the research involve administrative or secure data that requires permission from the appropriate authorities before use?		
12.	Is there a possibility that the safety of the researcher may be in question?		
13.	Will the research involve respondents interacting with the internet or other visual/vocal tools, where respondents may be identified?		

⁴ Adapted from: Ibid., at Appendix A: 33-34; and Goodhand (2000).

Box 2.2 Sample Written Consent Form

Type of Consent Form: [i.e., population being studied: e.g. Adult, Children, Teachers, etc.]

Study Title:

Principal Investigator:

[Investigators are expected to write consent forms in simple language, and convey the information in a language understood by the participant whose consent is being obtained.]

What you should know about this study

You are being asked to join a research study.

This consent form explains the research study and your part in the study.

Please read it carefully and take as much time as you need.

You are a volunteer. You can choose not to take part and if you join, you may quit at any time. There will be no penalty if you decide to quit the study.

During the study, we will tell you if we learn any new information that might affect whether you wish to continue to be in the study.

Purpose of research project

Why you are being asked to participate

Procedures

Risks/discomforts

Benefits

Payment

Protecting data confidentiality

Protecting subject privacy during data collection

Conflict of Interest

Who do I call if I have questions or problems?

Call the principal investigator, [insert name], at [telephone number] if you have questions, complaints, or get sick or injured as a result of being in this study.

What does your signature on this consent form mean?

Your signature on this form means:

You have been informed about this study's purpose, procedures, possible benefits and risks.

You have been given the chance to ask questions before you sign.

You have voluntarily agreed to be in this study.

Print name of Adult Participant

Signature of Adult Participant

Date

Include this Assent Statement below in all studies involving children.

Print name of child participant

This research study has been explained to my child in my presence in language my child can understand. He/she has been encouraged to ask questions at any time during the study.

_____	_____	_____
Print name of Parent/Legal Guardian	Signature of Parent/Legal Guardian	Date
_____	_____	_____
Print name of Person Obtaining Consent	Signature of Person Obtaining Consent	Date

2.5 Ensuring Privacy & Confidentiality

This is an extremely important part of research ethics, whatever the discipline you belong to. Respondents often will only participate in your research study if you can promise to ensure their privacy. However, difficult questions may arise, for example, how should a researcher respond to the information gleaned from an interview about ethnic killings without endangering the respondents? (Goodhand, 2000:14)

There are many stages in the research process where safeguards need to be put in place to ensure that this outcome is achieved.

Training of field researchers: Field researchers generally participate in research projects for a limited time period only, and have various levels of experience. During their training you will need to explain to them the reasons for your ethical research commitments, and show them how they can help you to meet them. They will also need to limit their own sharing of respondent identifiers to their immediate research colleagues only.

Respondent profile: On the profile sheet you can record the name and location of your respondents, along with other basic information. You cannot avoid recording this information because it establishes your sample population and allows you to return to respondents for follow up interviews if necessary.

Informed consent: If this is obtained in writing it needs a valid signature or fingerprint, and can be kept separate from the profile sheet and questionnaire.

Interview questionnaire: This is the document whose data is inputted for analysis, and should contain codes for all respondents.

Data processing: Further analysis of the data is based on codes as identifiers.

Data storage: Hard copies of the respondent profiles and informed consent forms can be locked and stored for a specified time period by the research organization.

Data handling is best kept in the hands of as limited a number of researchers as possible. This helps to avoid confusion and streamline the systems for maintaining confidentiality. There will always be some researchers who have access to the identifiers of respondents, but they should be limited to those who assume the most responsibility for the integrity of the project as a whole.

2.6 How to Get an Ethics Review Done

To date there are no formal, internationally recognized ethics review committees for the social sciences in Pakistan. There are, however, a number of such committees in the field of health research. Some of these committees have expressed an interest in incorporating the perspectives of social science research into their review processes. These include: Aga Khan University, IRD Ethics Review Board,

Pakistan Medical Council, Centre of Biomedical Ethics and Culture at the Sindh Institute of Urology and Transplantation.

Collective for Social Science Research: The ethics review committee of the Collective is for social science research only. It has members from the Collective as well as experts from outside the organization, who represent different disciplines, as well. It will review research proposals in the social sciences from outside the organization. The procedures for submitting a proposal can be found on its website: [www.researchcollective.org]

If you do wish to have your research proposal ethically reviewed by an existing committee, it will need to go through the following steps:

- a. Establish that your study, using the checklist, requires an ethics review.
- b. Fill out a form supplied by the ethics review committee to explain how your study will uphold the six basic principles and tackle any ethical questions.
- c. Submit your questionnaire and/or interview guidelines to the committee.
- d. The principal investigator, or lead researcher, on the study appears before a quorum of the committee to answer any questions committee members may have regarding the proposed study.
- e. The committee will respond in one of three ways:
 - i. Ask you to make any necessary changes to the questionnaire protocols or any other part of the study as suggested by the committee and then re-submit.
 - ii. Reject your study.
 - iii. Accept your study and request that you go ahead with the research and keep the committee informed as to its progress so as to ensure that ethics requirements are upheld throughout the research period.

2.7 Consequences of Not Conducting an Ethics Review

If you wish to publish your research findings in an academic, peer-reviewed journal, you may be asked whether your study has been through an ethical review. It is not the case with all social science research, but the number of journals that demand such a review is growing. If your study has a health angle to it, and you may wish to publish in a health-related journal, then the likelihood of being asked for an ethical review increases.

Donors that finance such research are increasingly demanding an ethics review as well, although this is not yet a consistent practice in Pakistan, it may become one.

There is a debate within the social sciences regarding just how relevant the process of ethics review is to these disciplines, where field research often depends upon using indirect methods to shed light on realities of people's lives. For example, simple observation of the activities of children in a playground may be integral to a sociology study but if individual informed consent had to be taken the chances are that the opportunity would disappear. Does this make the work of anthropologists or sociologists unethical? Proponents of this view also question whether the ethical standards that arose primarily from medical research due to its practices of testing drugs on a population that was unaware of how they were being used – should be adapted for social science research that does not do such interventions at all.

More resources on social science research ethics can be found from a variety of sources, such as:

- *Ethical Guidelines for Social Science Research in Health*, written by the National Committee for Ethics in Social Science Research in Health, India.
- ESRC Research Methods Programme. This is an informative site related to research ethics, based at the Centre for Survey Research, University of Manchester. The online research ethics guidebook can be found at www.ethicsguidebook.ac.uk

- Oral History Association offers information on ethics review and oral history. [www.oralhistory.org] On the history of the ethics debate in social science see their link: [<http://www.oralhistory.org/do-oral-history/oral-history-and-irb-review/bibliography-historians-and-institutional-review-boards/>] On their own ethics review standards see: [<http://www.oralhistory.org/do-oral-history/principles-and-practices/>]
- [www.socialsciencespace.com] A forum for bloggers and researchers to share their work and discuss issues.
- American Sociological Association [www.asanet.org]
- American Anthropological Association offers definitions, guidelines and information on conducting ethnographic research. [www.aanet.org]

Ethics review has become common practice in universities in the United States and the United Kingdom. Although they may differ in the details of their processes, they follow principles based on the ones we have discussed above. This means that university-based research is standardizing its procedures and requirement for ethics review in the social sciences. The ESRC Framework has become a template for many UK-based universities. In the South Asia the culture of ethics review in the social science is still very new, although more researchers are paying attention to ethics in bio-medical research.

2.8 An Example of Deciding Not to Share the Data

Sometimes it is not possible to minimize the potential risks to research participants even after the data collection. One such example is a study conducted on the effect of armed conflict on two groups of women: those affected by the civil strife in Karachi during the 1980s-90s and those living in Afghan refugee camps in Pakistan. The methodology was qualitative, and relied primarily on the research tool of taking life histories from respondents. There were some interviews with women respondents from Karachi whose identity could not remain completely confidential if shared because the context of their stories and the events narrated would make them identifiable to an informed reader. This was the case even if all names were changed in related publications. These women had been internally displaced within the city of Karachi and continued to move from one location to another in order to evade the police or Rangers. Those interviews, therefore, were not used for data analysis and were not included in the published research findings.

In the case of Afghan refugee women who were interviewed, some of them shared details about their experience with Afghan political organizations, such as the Taliban and former *mujahideen* groups. Some women recalled ethnic killings that had forced them to flee their homes. Yet others referred to ongoing discrimination they felt from political forces even in the camp setting. Women were often most comfortable giving these interviews not in the setting of their own homes, but in the offices of non-government organizations working with refugees because they were more neutral settings and women felt that their privacy was more maintained. While many of these interviews made their way into the analysis, but special attention was paid to concealing the names of political organizations and leaders.

2.9 Conclusion

This Module has been designed to inform you about ethics review processes, not to scare you away from researching vulnerable populations! There are conditions under which it is not possible to apply all of the recommendations discussed above, such as rapid assessments in emergency situations. However, the integrity of the research process needs to be maintained whatever the circumstances. This is a good time for more researchers in Pakistan to start developing ethics review processes, particularly as NGOs do conduct a large amount of the social science research in this country.

2.10 Practice Exercises

Case Study 1:

You are engaged in a research study with Afghan women in refugee camps, in order to determine their reproductive health needs and knowledge about how to access relevant services in their area. You have developed a questionnaire to that effect and would like to pre-test it with 20 families in one camp. The questionnaire includes a few questions on women's exposure to violence, within their families and in the public sphere as well. The purpose of these questions is to gauge how widespread this exposure is among Afghan women. During the course of the pre-testing, you discover that women respondents (a) are admitting to experiencing violence more often than you expected and (b) are exhibiting health-related consequences such as bruising and depression that may result from this violence.

Look at the Research Ethics Checklist to decide if your study needs an ethics review. List all of the ways your study needs to address ethical issues. For example, how would you go about protecting the respondents and the research study from harm? Draft an informed consent form.

Case Study 2:

You plan to do a study on Afghans in Pakistan and their income-generating activities, with a particular view to finding out about cross-border trade, including narcotics. You plan to visit their homes to interview men, women and children. The data and findings will be used to develop policy and programme interventions aimed at regularizing the transit trade, which may in the long-term decrease the income to those families that benefit from drug smuggling between Afghanistan and Pakistan.

Look at the Research Ethics Checklist to decide if your study needs an ethics review. List all of the ways your study needs to address ethical issues. For example, how would you go about protecting the respondents and the research study from harm? How would you protect yourself and fellow researchers from harm? Do policy outcomes count as a potential risk to respondents? Draft an informed consent form.

Case Study 3:

Your organization is helping internally displaced people who have fled the army action against the militants in FATA. The IDPs are being housed in makeshift camps just a few hours drive from their homes in the tribal areas, and they hope to return as soon as possible and resume their normal lives. Your organization would like to conduct a needs assessment of trauma-affected people in the camps, with a view to providing short-term mental health counseling for the worst affected individuals. In the long term your organization would like to train community members in their home settings to recognize signs of mental illness and provide referrals to medical doctors trained to provide relief.

Look at the Research Ethics Checklist to decide if your study needs an ethics review. List all of the ways your study needs to address ethical issues. For example, how would you go about protecting the respondents and the research study from harm? In this particular case, is there political risk involved for respondents and researchers alike? Draft an informed consent form.

2.11 Examples of Ethics Violation

See the examples below and identify ways in which the ethics violation could be avoided.

Example 1: You are conducting a needs assessment of women IDPs and are using the informed consent template as above. Since the women are illiterate, they cannot read the form themselves, but they do need to give consent. You obtain their fingerprint on the consent form without reading to them its contents.

Example 2: You are conducting a study with Afghan refugee men to determine how often they travel back and forth across the Pak-Afghan border, and what means they use. You have told your respondents that all interview material will be confidential. On the way back from the field you point out to an official from a government agency the neighborhood where you did your interviews, and name a male respondent who told you how he sneaks across the border by bribing officials regularly.

Example 3: You need to access women and girl IDPs from Pakistan's tribal areas and male relatives are reluctant to allow you and your field team into their tents to talk to women. They are suspicious of you. You promise them the interviews will remain confidential but they do not trust you ever since you mentioned that the purpose of your study is to determine education levels among female IDPs. In a last ditch effort, you offer to pay each male household head Rs. 300 if he allows for the interviews to take place.

Discuss what alternatives could work in difficult situations in the field.

References and Additional Reading

- Birman, Dina. "Ethical Issues in Research with Immigrants and Refugees," in Joseph, E. T., and Celia, B. F. (eds.) *The Handbook of Ethical Research with Ethnocultural Populations & Communities* (Thousand Oaks, CA: Sage Publications, 2006).
- Dahinden, Janine & Efonayi-Mader, Denise. "Challenges and Strategies in Empirical Fieldwork with Asylum Seekers and Migrant Sex Workers," in Van Liempt, Ilse & Bilger, Veronika (eds.). *The Ethics of Migration Research Methodology* (Brighton: Sussex Academic, 2009) at 98-117.
- Duvell, Franck, Triandafyllidou, Anna & Vollmer, Bastian. *Ethical Issues in Irregular Migration Research* (European Commission/Clandestino: 2008).
- Economic and Social Research Council. *Framework for Research Ethics* (Swindon, UK: 2010). [Accessed online at: www.esrc.ac.uk and <http://www.esrc.ac.uk/about-esrc/information/research-ethics.aspx>]
- Goodhand, Jonathan. "Research in Conflict Zones: Ethics and Accountability," (2000) *Forced Migration Review* 8 at 12-15.
- Guerin, P., & Guerin, B. "Research with Refugee Communities: Going Around in Circles with Methodology," (2007) *The Australian Community Psychologist*, 19(1) at 150-162.
- Hart, J., & Tyrer, B. *Research with Children Living in Situations of Armed Conflict: Concepts, Ethics and Methods*. RSC Working Paper No. 30 (Oxford: Refugee Studies Centre, 2006).
- Hyder, A. A. et al. "Ethical Review of Health Research: A Perspective from Developing Country Researchers," (2004) *Journal of Medical Ethics* 30 at 68-72.
- Jacobsen, K., & Landau, L. *Researching Refugees: Some Methodological and Ethical Considerations in Social Science and Forced Migration*. New Issues In Refugee Research, Working Paper No. 90 (Geneva: UNHCR, 2003).
- Jacobsen, K., & Landau, L.B. "The Dual Imperative in Refugee Research: Some Methodological and Ethical Considerations in Social Science Research on Forced Migration," *Disasters* 27 (2003) at 185-206.
- Joseph, E. T., and Celia, B. F. *The Handbook of Ethical Research with Ethnocultural Populations & Communities* (Thousand Oaks, CA: Sage Publications, 2006).
- Mackenzie, C., McDowell, C., & Pittaway, E. "Beyond 'Do No Harm': The Challenge of Constructing Ethical Relationships in Refugee Research," (2007) *Journal of Refugee Studies* 20 at 299-319.
- Medicines Sans Frontieres. "Research Ethics Review in Humanitarian Contexts: The Experience of the Independent Ethics Review Board of Médecins Sans Frontières," (2009) *PLoS Medicine* 6(7).
- Robinson, Julie. *Resources Relevant to Ethical Issues in Research Involving Migrants and Refugees* (Adelaide: Flinders University, School of Psychology, 2010). [Accessed online at: <http://marrc.flinders.edu.au/downloads/EthicsResourceKit.pdf>]
- Schopper, D., et. al. *The Belmont Report* (2009). [Accessed online at: <http://ohsr.od.nih.gov/guidelines/belmont.html>]
- Thomas, S., & Byford, S. "Research with Unaccompanied Children Seeking Asylum," (2003) *British Medical Journal* 327 at 1400-1402.
- Thompson, J. "Exploring Gender and Culture with Khmer Refugee Women: Reflections on Participatory Feminist Research," (1991) *Advances in Nursing Science* 13(3) at 30-48.

United Kingdom Research Integrity Office. *UKRIO Research Integrity Checklist*. [<http://www.ukrio.org/home/>]. [Accessed through ESRC, *Framework for Research Ethics*, at 37].

Wiles, Rose, Charles, Vikki, Crow, Graham & Heath, Sue. "Researching Researchers: Lessons for Research Ethics," (2006) *Qualitative Research* 6(3) at 283-299.

Wiles, Rose, Heath, Sue, Crow, Graham & Charles, Vikki. *Informed Consent in Social Research: A Literature Review*. NCRM Methods Paper Series. (Southampton, UK: ESRC National Centre for Research Methods, 2005).

World Medical Association, Declaration of Helsinki: Ethical Principles for Medical Research Involving Human Subjects (Adopted by the 18th WMA General Assembly, Helsinki, Finland, June 1964 and last amended in October 2008). [Accessed online at: <http://www.bioethics.nih.gov/international/declarat/helsinki.rtf>]

Zwi, A.B. et. al. "Placing Ethics in the Centre: Negotiating New Spaces for Ethical Research in Conflict Situations," (2006) *Global Public Health* 1 at 264-277.

Quantitative Methods: Developing and Administering Questionnaires

Module 3 - Quantitative Methods: Developing and Administering Questionnaires

Purpose: To take the reader through the steps for developing and conducting effective surveys.

Surveys are a common data collection instrument in social science research. The goal of a survey is to provide us with reliable quantitative data⁵ that can be used to answer our research questions, explore our hypotheses and provide hard data to support our arguments. However the findings of a survey will only be as good as the effort that has gone into preparing the survey and the rigor with which data analysis is conducted. Modules 4 and 6 will further cover the fundamentals of sampling and quantitative data analysis. After going through these Modules, it is hoped that the reader will be familiar with the mechanics of the survey research cycle, from identifying research questions and hypotheses to data analysis and presenting results. One only learns survey design and the analysis of data through practice. In consequence, in these Modules we focus on principles and on the practical application of principles.

We will begin with an overview of the various stages in survey research design. We then move on to the actual design and development of a survey followed by principles of pre-testing. Towards the end of this Module we will further discuss some of the processes we should keep in mind when conducting surveys in the field.

3.1 Overview of the Survey Process

A survey can be described as “a systematic method for gathering information from (a sample of) entities for the purpose of constructing quantitative descriptors of the attributes of the larger population of which these entities are members” (Groves 2004:2). For instance, the Demographic Health Surveys⁶ provide health, nutrition and demographic characteristics (such as the male/female ratio, the fertility rate, awareness of diseases like tuberculosis, HIV/AIDS etc.) of the entire population of a country based on interviews conducted with participants of a nationally representative sample.

Based on their timeframe, surveys can be categorized as follows:

Cross-Sectional: These studies gather data from a section of the population at only one point in time. For instance what are the views of refugees/IDPs on the housing conditions in their camp?

Repeated Cross-Sectional: In such studies different samples of the population are studied over time but the variables that are observed remain constant in each cycle. For example examining the status of women’s reproductive health in migrant communities over several years conducting the survey with fresh migrants each year.

⁵ Although surveys/questionnaires are also used to gather qualitative data, for the purpose of this module and related discussion in following modules, we have limited ourselves to quantitative surveys.

⁶ Measure DHS. Demographic and Health Surveys [Accessed online at: <http://www.measuredhs.com/>]

Panel or Longitudinal: This involves repeated measures of the same person, household, community, etc. over time. For instance a study to measure the changes in the incomes of refugee or IDP households over time after the introduction of a livelihoods strategy targeted towards migrants.

Cohort: Cohort studies, like panel studies, focus on a particular section of the population examined over time. However the difference is that whereas the cohorts are the same, the sample members will be different each time. An example is if we were to study the nutrition status of migrant children between the ages of 5 and 12 years, every two years, we would be studying the same group of subjects but not the same individuals.

The survey itself can be in the form of a written questionnaire, either administered by the respondent herself or administered in a face-to-face interview by a surveyor. Other questionnaires could be administered through a telephone interview or a web-based survey. In Pakistan, especially in the development sector, surveys are mostly conducted face-to-face with the respondent where the surveyor records the responses in the questionnaire.

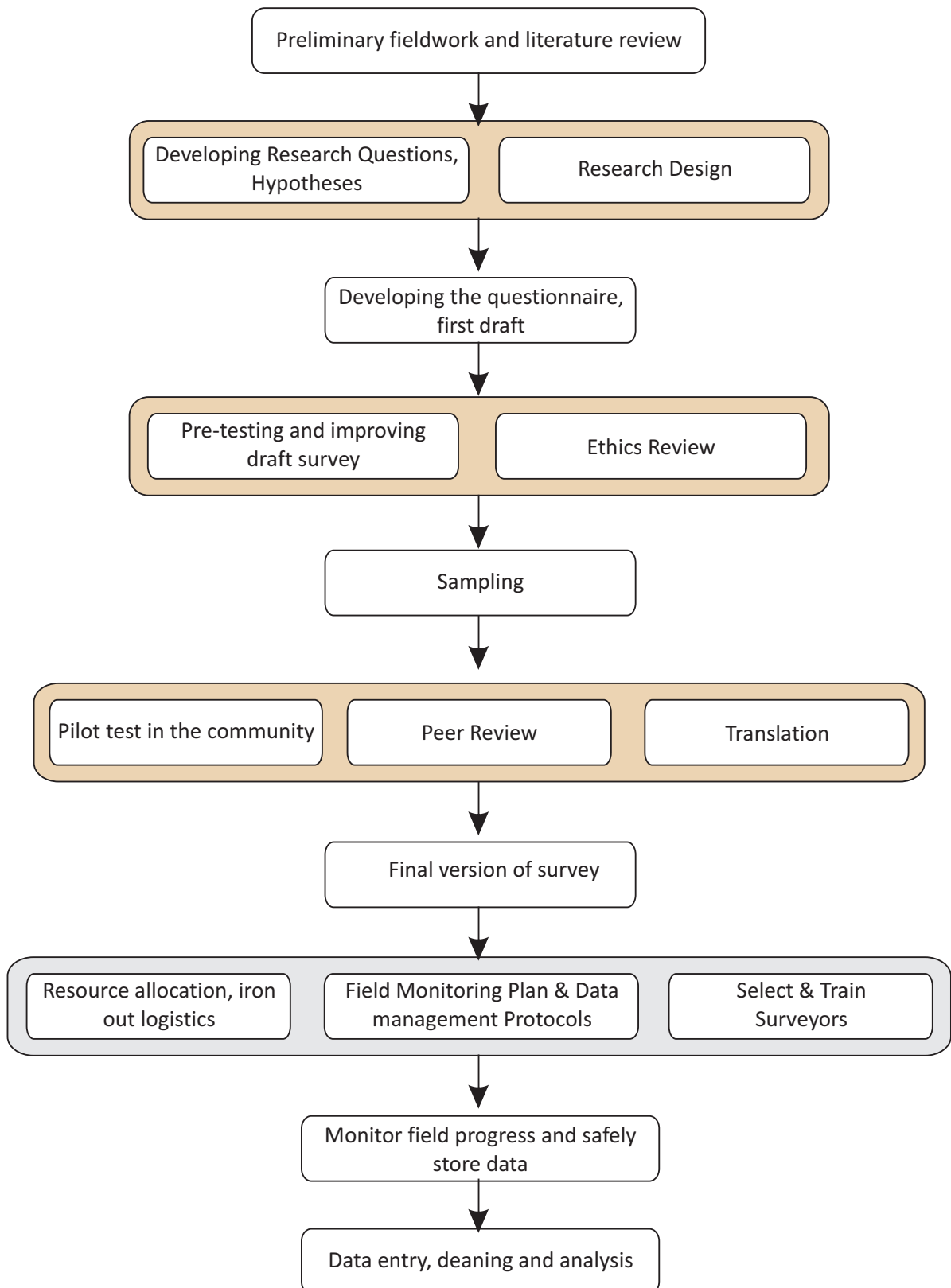
Irrespective of the type of survey and its form, to be effective it must go through a series of stages (at least once for recurrent surveys) before and after its execution. These stages are accepted globally as standard practices in survey research design. Box 3.1 presents these stages in a suggested order; although the sequence can be altered to suit the survey requirements and constraints. What is essential is to understand what each of these stages entails and how they can be applied practically. We will go through each of these stages now.

3.2 The Survey Cycle

3.2.1 Preliminary Fieldwork and Literature Review

Once you have identified the problem/issue that you want to study through your survey, your next step should be identifying and reviewing existing literature on the subject. Module 1 has already familiarized you with the concept and importance of a literature review. For instance if the aim of the research is to study cross-border human trafficking, relevant reference materials could include studies on trafficking in conflict zones, studies on trends in human trafficking in the region, studies on causes and consequences of human trafficking and government reports (or reports by NGOs) on trafficking in the country/region. At this stage it will also be beneficial to take guidance from studies conducted in similar/relevant areas by reputable local or international organizations such as the International Organization for Migration (IOM), the United Nations Office on Drugs and Crime (UNODC), the International Labour Organization (ILO), the Human Rights Commission of Pakistan (HRCP) and so on. The information that you gather through this comprehensive review will not only give you existing data, it will also guide you in the formulation of your research questions and in the identification of variables you need to consider in your survey. If you create a summary table of the studies that are most relevant to your research it will give you the relevant information in one document, reducing the need to refer to individual studies again and again. It also facilitates referencing when writing the research report. A suggested template is provided in Box 3.2.

Box 3.1 Stages in Survey Design



Box 3.2. Summary Literature Review Table

	Author & Year of Publication	Title of Study	Country/ Region	Sample Population (if survey based)	Method of data collection & analysis	Definition of the problem	Research Questions and Findings	Significance of results (statistical significance for quantitative studies)
Study 1								
Study 2								
Study 3								
Study 4								

While a comprehensive literature review provides a strong theoretical foundation to your survey to ensure your research is timely and relevant, the literature review also needs to be complemented by first hand information from the field through some targeted preliminary field work. This would include interviews (key informant interviews, as they are also known) with experts and practitioners in the field. In order to maximize the benefits from such interactions it is advisable to have some key questions ready beforehand to direct the flow of the discussion and if possible you should take along a preliminary version of your survey (even if it is a basic check list of questions for your survey respondents) for the key informant to comment on.

It will be extremely valuable at this stage to hold discussions with people from your target population. These can be focus group discussions with six to eight people. The idea is to ensure that the survey instrument that is ultimately designed is realistic and captures the target population’s perceptions and experiences. The number of such discussions will vary but it is good practice to hold at least two or three. (Fowler 2009)

The preliminary fieldwork and the literature review are the foundations of your survey and it is important to give this initial phase sufficient time and attention.

3.2.2 Research Questions and Research Design

Research Questions/Hypotheses

Going back to our example of cross-border human trafficking, we can see that this is a broad area of study and that covering all possible aspects would be beyond the scope of one research project. Specific research questions need to be formulated that will narrow the focus of the study and allow the researchers to use their time and resources effectively. The groundwork you do in Stage 1 will greatly assist you in narrowing the focus. Your aim in studying human trafficking might be to inform policy decisions that will reduce the incidence of trafficking in women, during and after conflict. You might choose to study people who are displaced due to military action in the Khyber Pakhtunkhwa. Honing in on the problem, your question could be as follows:

“What are the conditions in IDP camps that make women and children vulnerable to trafficking?” or

“What changes need to be made to IDP relocation strategies/plans that will reduce opportunities for trafficking?”

While not necessary, often quantitative research studies build on the research questions to develop more specific statements called hypotheses. “Hypotheses... are predictions the researcher holds about the relationship between variables” and they are typically used in studies that compare groups. (Creswell 2003:108) In our example, the possible hypotheses could be:

“Lack of appropriate employment opportunities in IDP camps significantly increases the likelihood of trafficking”; or

“Improvements in IDP registration processes will significantly reduce opportunities for trafficking”; or

“Improving law enforcement mechanisms/security in camps for IDPs will significantly curb trafficking”.

Once data has been collected, it is used to test whether or not our hypotheses hold true. We will discuss this concept in more detail in one of the later Modules.

Research Design

You will notice that at this stage we have started talking about the type of respondents you need for your survey. Furthermore it is important at this stage to decide the **research design** your study will follow. Box 3.3 provides a brief introduction to the typology of research design. When choosing your research design you will have to decide whether or not you will have more than one group of respondents. Continuing with our example, you might decide you want to interview two groups of respondents in two different IDP camps. One group might belong to a camp where several instances of trafficking have been recorded and the other from a camp where the phenomenon is absent or the incidence is lower.

At this stage you should have also decided on the **unit of analysis** for your study, i.e. whether your variables will be measured at the level of the individual, household, institution or community. It is quite possible that you require data at several different levels, for instance you might want to know whether a woman suffers from domestic violence but you might also want to relate that to the characteristics of the household in which the woman lives. Similarly you might want to know how individuals seek medical help and you might want to relate that to the medical facilities available in the community in which case you will also need data at the community level. While it is perfectly fine to collect data at different levels of analysis, the information should be collected through different instruments, i.e. through a household survey for household information and an individual survey for individual characteristics and so on. As a rule, “always collect data on the lowest unit of analysis possible” since individual data can always be aggregated during analysis but grouped data cannot be disaggregated. (Bernard 2000:46)

Variables to be Studied

Before you go on to design the actual questionnaire, you need to identify the most important **variables** that you will be observing through this instrument. Variables are features or properties that vary from one respondent, situation, geographical location to the other. For instance age, gender, marital status, wealth status, and so on. All variables have attributes or values, for instance the age has several age groups like 1 to 5 years, 6 to 10 years and so on. Variables are defined at three different levels of measurement; *nominal, ordinal and interval or ratio*. (Bernard 2000:41) Values of a **nominal variable** comprise of a list, for instance, caste, religion, occupation, etc. The list is mutually exclusive and exhaustive.

Ordinal variables, like nominal variables, comprise of exhaustive and mutually exclusive lists but their values can also be ranked, for instance socio-economic class --which is commonly described in terms of three categories: ‘lower’, ‘middle’ and ‘upper’.

Box 3.3 Research Design

Research design determines how you will organize your research, i.e. it is the structure of your research. It outlines the instruments you will use for your observations (e.g. questionnaires, interviews, etc.), whether or not there is a treatment/intervention, how many groups, if at all, will be part of your study (test, control group) and how will subjects be assigned to those groups. The main categories of research design are as follows:

Randomized Experiment: in which subjects (of research) are randomly assigned to intervention and control groups, the treatment/intervention is introduced and the outcomes or dependent variables (refer to Glossary) are measured. This design is best suited to studying the causal relationship between dependent and independent variables. Common experimental design classifications are:

The two-group pretest-posttest design: where two randomly assigned groups are tested before introducing an intervention to ensure comparability, a treatment is given to the one group (treatment group) and withheld from the other group (comparison/control group) and outcomes (dependent variables) measured.

The two-group posttest only design: where treatment is given to randomly assigned treatment group and not the control group (also randomly assigned) and outcomes are measured but no pre-test is conducted.

Quasi-Experimental: in such studies participants are not randomly assigned to treatment and control groups by the researcher; pre-existing groups/groupings are used.

There are studies in which there are no control and treatment groups and no intervention is induced. This can be called non-experimental research design. Some variations of this design are:

The one-shot survey design: where a single group of participants is studied, after an intervention/event has taken place, to examine the dependent variable.

The one-group pretest-posttest design: this is like the one-shot survey study but some variables are measured and then intervention takes place and the same variables are measured again.

Interval variables have the properties of nominal and ordinal variables. In addition, the distances between attributes are meaningful and they have ratio properties, for instance age, education, income, years of employment, distance in kilometers from school or health facility, and so on. The rule here is to try and measure attributes at the highest level of measurement possible (Bernard, 2000), i.e. at the interval level rather than nominal or ordinal levels; interval level data can always be used to generate lists and ranks later on during analysis. For instance to determine the household wealth status or socio-economic status you might also have to collect data on not only household income (i.e. how much the household members earn in wages) but also the household expenditure, food consumption and household assets. Once you have this information it can be collated to determine household wealth status.

You also need to distinguish between your **independent** and **dependent variables**. The dependent variable (often denoted by Y) is the outcome of interest and independent variables are what influence this outcome (usually denoted by X). In our example the outcome of interest is the incidence of trafficking. Possible variables that will affect the outcome could include the demographic composition of the displaced groups (age, gender etc.), the location of the camps, prevalence of unemployment, access to financial resources, strength/weakness of law enforcement apparatus, and so on. It is important that you relate these variables with your research questions/hypotheses to ensure that you get the answers you are looking for and to give structure to your questionnaire. Using a table to connect

the research questions/hypotheses and variables can be a useful exercise. A template is provided in Box 3.4 and our hypothetical research situation has been used to fill it in.

3.2.3 Developing the Questionnaire

Once you have clearly laid out your research questions and variables you are ready to develop the actual questionnaire. Now that you know what you want to ask, you should know how to ask it. While writing out a questionnaire, some key issues you should consider include: its length, type of questions, wording of the questions and progression of questions.

Length of the Questionnaire

There is no limit on the number of questions that can be included in a questionnaire. However every effort should be made to keep the questionnaire as concise as possible. This is important for two reasons: (a) we have to respect our respondents and their time and (b) long and complicated questionnaires are more likely to lead to enumerator fatigue.

Box 3.4 Relating Variables to Research Questions/Hypotheses

Research Question/Hypotheses	Variables
<p><i>1. Lack of appropriate employment opportunities in IDP camps significantly increases the likelihood of trafficking</i></p>	<ul style="list-style-type: none"> - Availability of temporary livelihood earning opportunities - Recreational facilities within the camp - Demographic composition (male/female, age groups, dependents) - Socio-economic/wealth status - Education levels of IDPs - Access to financial resources
<p><i>2. Improving law enforcement mechanisms/security in camps for IDPs will significantly curb trafficking</i></p>	<ul style="list-style-type: none"> - Existing laws and anti-trafficking regulations - Geographical location of the camp - Safety/security arrangements in and around the camp - Knowledge/attitude of security forces regarding trafficking - Knowledge/attitude of displaced persons regarding trafficking - Perceived comfort & security of IDPs - Ease of access to security forces/police

It is therefore extremely important to carefully identify your most important variables; the more variables you want to observe the more questions you'll have to ask. It helps to think of yourself as a respondent while designing the questions (it might help you identify annoying or offensive questions.) To avoid repetitive questions make sure you know exactly what data you are getting out of the questions; you might be asking the same thing three times. However there are certain variables for which you might need more than one question. This is generally for variables that you suspect will be misreported, such as the age of illiterate respondents, year/month of migration, serious health conditions, for example.

Finally, every question has a price tag; most of your resources will go towards conducting the questionnaire so make sure you spend them wisely.

Types of Questions

Your questionnaire will usually be a mix of different types of questions. These are:

Open- Ended: these questions allow for the respondent to answer the question in whichever way they like without restricting their response options. In quantitative surveys, these responses are coded after data collection. If you are looking for very detailed responses these questions are useful otherwise their use should be restricted since you are unlikely to get clear, concise answers. An example of an open-ended question could be:

What are the conditions that need to be met for you to be repatriated?

Closed-Ended: these limit responses to the suggested answers accompanying the question, which have already been coded. Chances of getting responses that you are looking for are greater using closed ended questions. One example of a closed-ended question could be:

In order for you to be repatriated, which of the following conditions would have to be met?

1. *Availability of employment opportunities back home*
2. *Assurance of security*
3. *Availability of land or housing*
4. *Availability of schooling for children*

Single/Multiple Responses: When you provide the respondent with a list of response options, you have to decide whether you want a single response or whether the respondent can choose more than one option. In the example above there are four different options and as there could be multiple responses to this question more than one can be selected. In contrast, the following question can only have one response:

Are you willing to be repatriated?

1. *Yes*
2. *No*
3. *Not sure*

Ranked Responses: In these questions you ask the respondents to rank the suggested responses. For instance:

Please rank the following in order of importance to you when considering repatriation:

<i>Factor</i>	<i>Rank (1 to 4, with 1 being most important and 4 being least important)</i>
1. <i>Availability of employment opportunities back home</i>	
2. <i>Assurance of security</i>	
3. <i>Availability of land or housing</i>	
4. <i>Availability of schooling for children</i>	

Wording of Questions

Paying careful attention to the wording of questions is extremely important. What you want to ask should come out clearly, unambiguously and in as few words as possible. Given below are some examples of ill-worded questions and suggestions on how to improve them.

Example 1. Do you agree or disagree that the government’s action against militants was necessary?

1. Agree 2. Disagree

By putting ‘agree’, ‘disagree’ and ‘was necessary’ in the same sentence we have made the question ambiguous and confused the respondent. The question could be rephrased to read like this:

It was necessary for the government to conduct military action against militants.

1. Agree 2. Disagree

Example 2. Where do you get most of your information about current events?

1. Radio 2. Newspapers 3. Magazines

This question can have multiple responses and people might use sources other than those listed above. The rephrased question could be:

How much do you rely on each of the following sources for information on current events?

	1. All the time	2. Sometimes	3. Never
1. Radio			
2. Newspapers			
3. Magazines			

Example 3. Do you agree with the general sentiment that the government has failed in delivering basic services like food, water and security in camps?

1. Yes 2. No

This is an example of a leading question; it directs the respondent to a particular answer. Such wording should be avoided and instead the question can be phrased as follows:

Do you think that the government has failed in delivering basic services like food, water and security in camps?

1. Yes, to a large extent
2. Yes, to some extent
3. No, it has not failed

Example 4. Do you think health facilities in the city are limited and that the government should set up separate services for refugees to reduce the burden on existing facilities?

These are actually two questions (double-barreled) posed as one and the response you get will be misleading. It is better to break this up into two questions:

Do you think the use of health facilities by refugees increases the burden on existing health facilities in the city?

Do you think the government should set up separate health services for refugees?

Example 5. How much food is consumed in your house in one week?

This question does not only test the respondent’s recollection or memory but it is also vague in what it asks. Most people don’t remember how much food they’ve consumed in a week and moreover food consists of many items. To get specific answers you need to ask specific questions. The same question can be asked as follows:

Please tell us how many times the following items were consumed by your household in the past seven days:

Frequency	Food Items			
	Meat	Eggs	Oil	Sugar
Everyday				
Every other day				
Twice a week				
Once a week				
Never				

To recap, good survey questions:

- Should not be ambiguous
- Should not be leading
- Should not load two questions into one
- Should not be too long
- Should have an exhaustive list of responses and appropriate response options (i.e. whether they can select more than one)
- Should ask for specific responses
- Should not overwhelm the respondent

Progression of Questions and Format

As a general practice straightforward and less intrusive questions, related to the topic of research, are placed before more personal or intrusive ones. For instance if you were to ask someone about the use of contraception, it should not be part of your initial line of questioning. You can start instead by taking background information on more general variables such as age, years of marriage, education, employment, etc. Once the respondent is comfortable with the presence of the enumerator and gets into the flow of answering questions, it will be easier to elicit more sensitive information.

Make sure the instructions for the interviewer are clear and can be differentiated from what the interviewer has to read out to the respondent. You can use a different color, larger font, capital letters or italics for instance. All the instructions, transitions and required definitions should be written clearly in the questionnaire. Instructions to skip questions that might not apply to particular respondents should also be clearly written next to such question.

Once you have compiled the first draft of your questionnaire and have critically reviewed it with your team, it is advisable to test the instruments with actual respondents. While you are working on the questionnaire you should simultaneously identify respondents for your pre-test from among your target population. This need not be an extensive sample; usually ten (or fewer) interviews suffice. (Fowler, 2009)

3.2.4 Survey Pre-Testing and Ethics Review

Prior to conducting formal field tests a technique known as cognitive testing can be used to assess the consistency, validity and strength of your questionnaire. (Fowler, 2009) These interviews are conducted, not necessarily in the environment in which the actual survey will take place, using the draft survey by a senior interviewer/researcher who knows the exact objective of each question. The respondents should be informed of the purpose of the exercise. The purpose of these tests is to know whether the questions are being understood in the way they should be, whether the response options appropriately capture what respondents want to say and do the responses reflect what the questions

are trying to measure. (Fowler, 2009) Ultimately this exercise will help you in proactively addressing some of the issues you might face in the field and allow you to make timely changes in your questionnaire to avoid costly mistakes.

If your study requires an ethics review, you will need to follow the appropriate steps. The instructions for doing so are given in Section 2.6 of Module 2. Since ethics reviews can take time it is suggested that the process be initiated as soon as the first draft of your questionnaire is ready. The survey pre-test might have to be put on hold till the review is complete. Two aspects of research ethics that are of critical importance to surveys, and should be done irrespective of whether a formal ethics review is required, are obtaining informed consent (Section 2.4 of Module 2) and ensuring privacy and confidentiality of the respondent (Section 2.5 of Module 2). You can choose to follow the template provided in Module 2 (Box 2.2), with required alterations, as your consent form. Alternatively you could place it before the questionnaire as part of the identification page, as is also commonly practiced. Either way the respondent's consent, written or verbal, has to be obtained before the interview can begin.

3.2.5 Sampling

Please refer to Section 4.1 of Module 4.

3.2.6 Finalizing the Questionnaire

Before you can say that your questionnaire is ready to be used in the field for data collection you should have carried out a final field pre-test, the questionnaire should ideally have been reviewed by peers and it should have been translated (if needed) into the language that is understood by the respondents.

Peer Review

It would be helpful to have your questionnaire reviewed by a small group of peers- academics, other researchers, field experts, and research sponsors, when all your questions are in a coherent order. [Peer review is not the same as ethics review, which should be done well before this]. Review by an outsider/third person might identify mistakes or inconsistencies that you might have overlooked on account of being so closely involved in designing the instrument. Concerns about confidentiality are understandable, therefore the survey should only be shared with people you know will not use or share your survey for any purpose other than review.

Final Pre-Test

This pre-test differs from the one conducted earlier on in the survey cycle in that these interviews are conducted under conditions in which the actual survey will be conducted, thus the purpose is to determine how the survey actually works in the field. Care should be taken not to interview respondents from the survey sample. This also prevents the survey respondents from talking about the survey before it actually starts. It is important that you (the lead/senior researcher) conduct some of the interviews yourself since you have designed the questionnaire and would be in the best position to address any identified problems.

While some experts say 10 interviews should suffice for the purpose of pretesting (Bernard, 2000: 254), others recommend conducting 20 to 50. (Fowler, 2009:122) One could take a range and consider doing at least 10 interviews if not more. Depending on your resources and time schedule you should decide on the number of interviews and the cutoff date for the pretesting phase; if this phase is prolonged more resources will be diverted away from the actual survey and there will be time delays. Bear in mind, all that you learn from this phase will have to be incorporated into your questionnaire before it is finalized.

One technique that can be used during pretesting is to rate the questions on the following criteria: "(a) it is easy to read the question as worded; (b) respondents understand the question in a consistent way

and; (c) respondents can answer the question accurately". (Fowler, 2009:123) Another method is to use audio-recordings of the interview to check whether the questions are being read properly, whether the respondent asks for clarification and whether questions require further probing from the interviewer. (Fowler, 2009:123)

Having done two rounds of pretests and a peer review you should be confident enough to take your survey into the field. What is left now is to make sure that the questionnaire is translated, if needed in another language, and that it has been properly formatted. Some general formatting tips are:

- Make sure all the instructions for the interviewer/enumerator are in the questionnaire. The interviewer should easily be able to distinguish between instructions specifically for him/her and those that the interviewer has to read out to the respondent. Use upper case letters or different colors for instructions.
- Do not try to cram too many questions on one page. Avoid using small font size and give appropriate space for answers instead of trying to make the questionnaire appear smaller than it actually is!
- Do not break questions at the end of a page, start on a new page.
- Make skip patterns very clear; if a question does not apply to a particular respondent they should not be made to answer it and the interviewer should have written instructions explaining which question should be next.
- Try to line answers (options) vertically rather than horizontally.

3.2.7 Preparing for the Field

Resource Allocation & Logistics

Like every project, a survey should have a budget of its own which ought to be followed strictly to prevent overspending. The budgeting process should be initiated at the beginning of the survey cycle and at this stage only final adjustments should be made. The budget should demonstrate clearly how many resources- days/times, staff/personnel, money, material, etc. - will be required for each identified activity. Based on our survey cycle, the budget heads one can expect to find in the budget (these are not exhaustive) and a suggested budget format are presented in Box 3.5.

Details regarding travel and accommodation of the research team should be sorted out at this stage; there should be no logistics-related surprises in the field for the team. When the teams go into the field they (the supervisor) should have enough cash with them to cover all their requirements.

Recruiting and Training Surveyors

The success of your survey depends heavily on how well the interviewers/surveyors understand the survey and how motivated they are. If interviewers fail to do their job properly the survey data will be fraught with errors. The way to prevent this from happening is to select the right enumerators, train them properly and supervise their work.

Interviewers might be from your own organization or you might need to hire external help, which is very common. When choosing enumerators it is advisable to select people who, in addition to having good reading and communication skills, have some prior experience in data collection, not necessarily in the same field though. Generally the age, gender or ethnicity of the interviewers does not impede data collection; however in certain circumstances demographic characteristics of interviewers might play an important role. For instance, if the survey is about reproductive health and the respondents are women then the interviewers should also be female. If your respondents do not understand Urdu then it is useless to send enumerators who can only communicate in Urdu.

According to Fowler (2009), the purpose of the training should be to familiarize interviewers with the specific objectives of the study and to teach them to be *standardized interviewers*. In order to be standardized interviewers the enumerators need to:

Box 3.5 Sample Survey Budget

	Activity	Required Units	Amount (Rs) per unit	Total (Rs)
1 Preliminary research				
a	Literature review	-1 senior researcher -2 research associate for 30 days	Rs xyz for senior researcher Rs xyz for RAs	Rs xyz x 1 x 30 + Rs xyz x 2 x 30
b	Preliminary interviews	Travel to place to interview for 5 interviews	Cost per visit (cab fare/rent a car)	Cost per visit x 5
c	Preliminary research report	Copies of report	Printing cost/copy	No of copies x cost/copy
	Sub-total			xxx
2 Developing the Questionnaire				
a	Questionnaire design and development	-2 senior researcher -2 research associates for 60 days	Rs xyz for senior researcher Rs xyz for RAs	Rs xyz x 2 x 60 + Rs xyz x 2 x 60
b	Ethics review		Rs xxx (if there is a review fee)	
c	Translation	1 translator to translate 150 questions	Rs yy per question	150 xRs yy
d	Printing	200 copies	Printing cost/copy	No of copies x cost/copy
	Sub-total			xxx
3 Enumerator Training				
a	Reading material for trainees	10 folders	Rs xy per folder	10 x Rsxy
b	Training venue charges	1 hall for 2 days	Rs xyz/day	1x2 x Rs xyz
c	Travel and accommodation			
	Sub-total			xxx
4 Conducting the Survey				
a	Travel to survey sites	Air travel/road travel (number of visits)		
b	Enumerator fees	10 people for 30 days	Rs xyz/day	10x30 x Rs xyz
c	Food & Accommodation			
d	Other daily expenses			
	Sub-total			xxx
5 Data Entry & cleaning				
a	Data entry	2 data entry personnel	Rs xx/questionnaire	2xRs xx
b	Supervision of data entry	1 supervisor for 15 days	Rs xx per day	1x15 x Rs xx
	Sub-total			xxx
Total for all survey activities xxx				xxx

- i. Present the study in a way that the respondents have a common understanding of the objectives of the survey;
- ii. Ask the questions in the way they are in the questionnaire without any variations;
- iii. Probe with follow-up questions where the response is unclear or incomplete without suggesting any answers or trying to shape the response in any one way;
- iv. Record the answers as they are reported by the respondent; i.e. record answers verbatim for open-ended questions and record only the answer selected by the respondent for fixed-response questions;
- v. Enumerators should also be trained to verify or cross-check information which is provided by respondents (such as age through NIC cards, distance from public facilities such as schools/hospitals by asking neighbors or other community members); and
- vi. Maintain a professional relationship with the respondents and refrain from expressing their opinions or judgments. However the enumerator should not appear to be hostile, rude or uninterested.

In addition the trainees should be oriented with more specific aspects of the study itself such as the study research objectives and how the study findings will be used. It is important for the interviewers to be able to explain the purpose of the study to the respondents as clearly as possible. Steps to safeguard confidentiality of respondents should also be communicated to the respondents.

Trainees should also be informed about the sampling methodology and the reasons for selecting respondents that have been chosen for the study. Not only will this give them confidence in answering respondent queries in the field but will also help them in identifying other respondents in case of non-response, especially where non-probability sampling techniques are used.

The duration of the training sessions will depend on the length and complexity of the questionnaire as well as the time and resources that are available at hand. The format of every training session will be different but a rough outline is as follows:

- i. Provide the trainees with a copy of your final questionnaire and an outline of the training session.
- ii. After introductions you can discuss the research objectives of the study, their importance and how the data will ultimately be used. You can also discuss here the research design and sample selection methodology. It will be important for the interviewers to know what type of respondents they will be interviewing.
- iii. Discuss the roles and responsibilities of the interviewer and importance of this role.
- iv. Go over remuneration, logistics and team coordination. The enumerators and the team supervisors should have each other's contact information. If the team is going to work in a high risk area, security and safety arrangements should be discussed to everyone's satisfaction.
- v. The field strategy should be clear to all the interviewers, i.e. how many interviews to be conducted per day per person, how many survey sites are to be covered, how many interviews in one site, how to proceed once an interview is complete.
- vi. Discuss interpersonal aspects, i.e. how should the interviewers interact with the respondents, how to motivate respondents to participate in survey, what to do in case a respondent declines to be interviewed, how to handle a situation where the respondent gets upset or becomes hostile, and so on.
- vii. Go through the questionnaire, question by question, discussing the purpose of the question. The trainer should ensure that the enumerators understand each question as it is written. Make sure the interviewers understand skip patterns and identify questions where further probing might be required, possibly in open-ended questions.
- viii. After having gone through the questionnaire it might be useful to have the trainees take turns as interviewers and respondents in practice exercises. Not only will this give the interviewers a better idea of how to conduct the interview but will also identify the weak and strong interviewers as well as areas where further work might be required.

The training sessions are extremely important but no matter how extensive these sessions are the interviewers will, inevitably, learn more once they are in the field. The role of the supervisor becomes extremely important here. The supervisor should try and review completed questionnaires while the team is in the field. Through this review the supervisor can identify whether skip patterns are being followed, whether answers to open-ended questions are being recorded verbatim and whether the interviewer needs to revisit the respondent. It will give the supervisor a general sense of the quality of data recording. However, not all aspects of interviewing can be gauged through the review of completed questionnaires therefore the supervisor should also observe actual interview sessions. While some interviewers might require a bit of retraining, there would be a few who will not improve even with further training; in case of the latter it is best to take them off the study.

Field Monitoring Plan & Data Management

Once the research team is in the field their performance should be monitored against a time schedule. This schedule (a timeline or chart) should map out the time to be spent in each survey site based on the total time available to conduct the survey. If there are any delays in actual fieldwork appropriate adjustments should be made to the schedule after discussing with the research team.

In addition to time management monitoring field progress entails ensuring the research team has the resources and support needed on ground to facilitate data collection. This includes ensuring sufficient financial resources are available, that the field contacts are helping the research team (for instance by taking surveyors to the respondents), that the right respondents are being interviewed and that the data being collected is of the desired quality. A great deal of the field monitoring responsibility falls on the shoulders of supervisors who should be in regular contact with the central research unit, office or senior research team.

Storage and safety of questionnaire hard copies are important considerations in data collection. What happens to the questionnaires once they are filled? If the research site is far away from the central research unit/office, how will the questionnaires be sent back? How do we ensure safety of questionnaires while they are on the research site? Once all questionnaires are in one place how will they be sorted and stored? How do we restrict access to the questionnaire hard copies? These are some of the questions that the researchers should consider before data collection begins.

3.2.8 Data Coding, Entry and Cleaning

It is one thing to collect data and another to use it for analysis; data needs to be 'prepared' for statistical analysis. The aim of this phase is to sort, clean and organize data to make it usable not just for you but also for others who might want to utilize it for further exploration. As soon as data from the field starts coming in you can begin the process of data entry. However, before you can start entering the data, you need to ensure it is properly coded. **Coding** is the process of transforming variables and responses in your questionnaire into numbers. The purpose of **data entry** is organizing the coded variables and responses in a format that can be processed by a computer program such as SPSS.

The variable names and codes for survey data are collated in a **codebook**. Every question represents a variable- such as age, education, income, marital status etc. - that needs to be given an appropriate name. Some examples are: AGE, EDU (for education), and MARSTAT (for marital status). Variable names should generally be short (computer programs usually restrict it to 8-10 characters) and intuitive. For ease of reference variables should appear in the codebook in the order that the questions appear in the survey. A large part of the coding process is undertaken while preparing the questionnaire with pre-numbered response options. For instance for the following question, the response codes will be 1 for Radio, 2 for Newspaper and 3 for Magazines:

What is your main source of information about current events?

1. Radio
2. Newspapers
3. Magazines

The variable name in this case could be SRINFO.

The information that one should be able to find in a codebook includes at least: the variable names; variable descriptions; variable format (numeric or text); response codes (i.e. the possible values the variable can assume); and reference to the specific question in the questionnaire from which the variable is derived. You can use Excel to create the codebook. An example of how this information can be presented is as follows:

Box 3.6 Sample Codebook

Variable Name	Variable Description	Format	Response Codes
RESID	Area of residence of respondent	Numeric	1. Urban 2. Semi-urban 3. Rural
LOCAL	Name of locality where respondent lives	String/text	As reported/observed
GENDER	Gender of respondent	Numeric	1. Male 2. Female
AGE	Age of respondent	Numeric	Age reported in years by respondent. For 'Don't know' use 999 For 'Unascertainable' use 888
EDU	Highest level of education achieved by respondent	Numeric	0. No education 1. Less than primary 2. Up to primary 3. Less than secondary 4. Up to secondary 5. College 6. Intermediate 7. Graduate 8. Other 999. Don't know

In Box 3.6 you will notice the terms 'Don't know' and 'Unascertainable'. These are categories of missing data, i.e. for which you do not have any response or have an ambiguous response. There can be many reasons for missing data; the respondent might not want to give an answer, the respondent might not know, or the respondent might give an answer that is not understood by the interviewer. Different codes should be used for different categories of missing data and they should be distinct from the regular response codes. In addition the missing data codes should be used consistently for ease of entry and to reduce chances of error in data entry.

For fixed-response questions the coder only has to assimilate the pre-numbered options into the code list and assign values to missing data. You can start entering codes for such questions in the codebook even before you get data from the field. However where 'Other' is given as a possible response option and where the question elicits a response from the respondents in their own words you will have to wait until you get the filled questionnaires before coding since you cannot predict the range of all possible answers in advance. When coding such responses one has to try and consolidate the responses by grouping answers that are analytically similar and distinguishing responses that are different. (Fowler, 2009) Care should be taken not to make the categories too fine, which is likely to complicate the analysis, or too broad, in which case you might miss out important variations. For instance if you ask respondents to report any severe illnesses they might have suffered from during the

past twelve months you will get a wide range of answers from cough to tuberculosis; from minor injuries to cancers and so on. Once you get all the responses you can list them all down and then categorize them as (1) mild; (2) moderate; and (3) severe, for instance, according to medical conventions.

When data is entered, in Excel or directly into the data analysis program such as SPSS, each row represents one record or one case/respondent. The columns contain information on the different variables. A typical data entry template would be similar to Box 3.7. [Please refer to Box 3.6 for variable names and codes].

Box 3.7 Data Entry Template

Serial Number or Respondent ID (should be cross referenced with the actual questionnaires)	Respondent Name (makes reference to questionnaires easier)	Variable 1: RESID	Variable 2: GENDER	Variable 3: AGE	Variable 4: EDU	Variable 5: x
1001	Xyz	1	1	45	5	
1002	Pqr	2	2	32	2	
1003						

When data is being entered, the likelihood of making mistakes is quite high but at the same time this is the phase of the survey cycle where you have maximum control over error reduction. Fowler (2009) suggests some approaches to data entry that can minimize errors:

- Minimize the need for interviewers to make coding decisions during the interview process; you increase the chances of incorrect codes being used by doing so. What the respondents report should be recorded verbatim and coded later by the coder.
- All coders should be trained to ensure that the same responses are not being coded differently. This can be done by asking all coders to code the same questionnaire and then comparing the codes.
- The data entry process should be supervised by a senior coder or researcher. They can run checks by picking up a few questionnaires randomly and checking how they have been entered in the system/database.
- Having two people enter the same data independently (or double entry) is a commonly used practice to identify data discrepancies. There are some data entry programs that can check for inaccuracies as soon as data is entered, especially in computer aided data collection, but such systems are not widely available in Pakistan.

Once data has been entered another round of checks should be performed to 'clean' the data as far as possible. This task should be undertaken by the coding supervisor and/or senior researcher. The main objectives of this exercise are to ensure that:

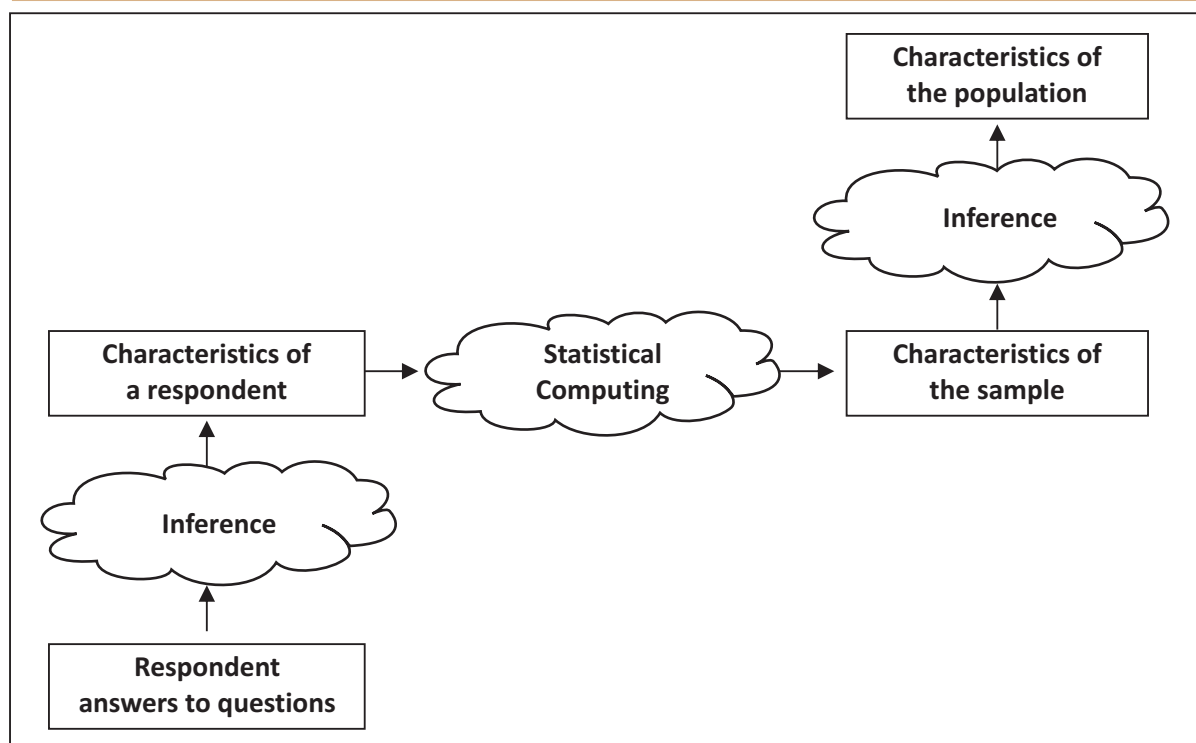
- Entries are consistent with each other;
- The skip pattern is not violated (questions meant for a subsection of the population should only have answers from the sub-group); and
- Data is within the acceptable limits.

A common way of checking data accuracy is to do some preliminary data analysis using descriptive statistics such as summaries, frequencies, mean and standard deviation. Frequency, or how often a value appears for a variable helps you spot unequal distribution. Too many zeros or too many '999's should alert you to the possibility of errors, and so should concentration of respondents in one value,

for instance more than 60% of the population being in the same age group when you know from your sample selection methodology that that cannot be true. Another example could be where you know your sample should have an equal number of males and females but summary or frequency of the GENDER variable shows that 55% of your sample is male, there is likely to be an error in recording the gender of respondents. Simple logical analysis of descriptive statics will reveal if recorded responses make sense. For instance you will know instantly that there has been a recording error if 5% of your male respondents are pregnant! Measuring standard deviations is also another check. If for instance the standard deviation (refer to glossary) is very high, higher than the mean of the variable, it could indicate some extreme values in the data which should be cross-checked with the source questionnaires.

3.3 Survey Errors

Box 3.8 Survey Inference⁷



The diagram above (Groves, 2004) represents the process through which a survey collects data from individuals in a sample to draw inferences regarding the characteristics of a larger population. In doing so, the survey process relies on two basic principles:

- (a) that the selected sample represents the target population; and
- (b) that the answers of respondents in the sample can be used to define characteristics of those sampled

What if the characteristics of those sampled are different from those of the larger population, or, what if the characteristics of the larger population that we hope to capture through our survey are not distributed in the same way in our sample as they are in the target population? What if the answers are not accurate?

⁷ Groves (2004:39)

There are two fundamental sources of error in the survey process; i.e. “deviations of what is desired in the survey process from what is attained.” (Groves, 2004: 40) For both types of errors, i.e. the sample not being representative of the target population (Error A) and answers inaccurately describing the sample (Error B), Fowler (2009) states there are two sources:

Error A: Target Population

(i) **Random Variability or Sampling Error** “stems from the fact that data are collected from a sample rather than from every single member of the population” (Fowler, 2009: 13) It is quite possible that the sample selected is not a perfect representation of the target population, for instance the proportion of educated people or younger people or females might be higher in the sample than the proportions of the same in the target population. Research design and sample selection processes will determine how well the sample represents the target population.

(ii) **Systematic bias** means that the sample members are different from the target population in some systematic way. This could be due to the sampling frame used to select the sample, which could deprive a section of the population from the chance of being selected in the sample. For instance if you only take the population of relief camps as your target population, you will automatically exclude other IDPs/refugees who do not find space in camps or who opt not to stay in camps. Another way in which bias can be introduced into the sample is through the process of sample selection, i.e. the way members of the sample are selected and this is of particular concern if the sample is not selected randomly. If you, for example, select your respondents through the *snowballing* method (relying on information gathered from key informants to select your respondents through referrals), you will exclude all those who are not acquaintances of your key informants and those who are selected are likely to be a different group from those who are not selected; you are likely to find correlations in the results that might not be applicable to those individuals not selected. Finally, there could be instances where some of the selected respondents refuse to be part of the survey or are unavailable to answer questions, which could potentially bias the results of the study.

Error B: Sample

(i) **Random variability or Invalidity** of answers could result from several factors such as misunderstanding the question, not having sufficient information to answer the question, poor recall by the interviewee, misunderstanding on the part of the interviewer or deliberate under/overstatement of facts. It is assumed that there will always be an element of error in the answer that will distort the actual value of a given variable. How close the answer is to the true value of the variable is known as validity. (Fowler, 2009) These variations in responses can be random, i.e. sometimes the error is in one direction and sometimes in the other and it averages out over many responses. An example of random variation would be one in which some respondents overstate their age or height while others under-report but in the average of the overall sample the over and under reporting even out.

(ii) **Systematic bias** is the other possible variation in answers that is possible. It is introduced into answers when in some systematic way the answers differ from the true values of variables and the error is more likely to be in one direction than the other. For instance, if women are asked to report instances of domestic abuse they are likely to under-report it for fear of backlash. Similarly if men are asked to report their expenditure on cigarettes, for instance, they are likely to under-report the amount for fear of reproach.

3.4 Practice Exercises

(These exercises can be done in groups)

1. An international aid agency plans to implement a programme to improve the food security and nutritional status of IDPs housed in relief camps in different parts of Khyber Pakhtunkhwa. Your organization/department has been selected to assist the agency in the design and execution of the

programme in your district. As part of the groundwork you are required to conduct a situational analysis survey of the existing food security situation in the camps in your district. You have been given two months in which you will conduct the survey and present your findings. Following the survey cycle discussed in Section 3.2 of this Module:

- i. Identify possible sources you will refer to when conducting the literature review;
 - ii. Develop your specific research questions;
 - iii. Develop an appropriate research design (including identification of the target population from which you will draw your sample; identification of important variables of your study and matching them with your research questions);
 - iv. Develop some questions (not less than 15) that will go into your questionnaire, explaining what information you will get out of each question, and place them in the order you think they should appear in the questionnaire. [Note these questions should be less general and more specific to the topic]; and
 - v. Identify some of the possible errors or biases that could be associated with your study.
2. You want to study the impact of displacement on refugee women (not IDPs) particularly in terms of gender-based violence and human rights abuses. The aim is to assess whether displacement has made these women more vulnerable to violence and abuses. For this study:
- i. Develop specific research questions and identify the important variables you would want to measure through your survey. Link the variables to your research questions;
 - ii. Develop some questions (not less than 15) that will go into your questionnaire, explaining what information you will get out of each question, and prepare a codebook for these questions (this will include all the fixed-response questions that you develop);
 - iii. Discuss pre-testing options;
 - iv. Discuss some ethical concerns that might be associated with the study and explain how you will address them; and
 - v. Prepare a brief enumerator training plan.

3.5 Conclusion

The purpose of an effective survey design process is to reduce these errors to the minimum, and the systematic approach laid out in this Module is geared towards helping you achieve that goal. At every stage we have discussed ways to reduce errors and increase confidence in our survey findings. These methods include the following:

- Conducting targeted preliminary fieldwork and in-depth literature review to ensure that the foundations of your survey instrument are strong and that it is grounded in reality. This way you can get insights into designing realistic research questions and be sure that the views of field experts have been accounted for.
- Defining specific research questions/hypotheses sharpens the focus of your study and prevents the research from derailing. When you know exactly what you want out of your research/survey, you will be able to identify clearly the variables that your study should measure and target the appropriate population, thereby enhancing the validity of your survey and improving chances of your sample being more representative of the target population.
- Much rests on adopting the appropriate research design. It defines the research instruments that will be used, the type of sample that will be selected, and how it will be selected. It also provides a map of how variables are related to your research objective and how they will be measured. Needless to say, the sample selection methodology will have a direct impact on how well your sample represents the target population and quality of your survey instrument(s) will determine how well the responses capture the characteristics of your sample.
- By wording your questions carefully, using the appropriate question formats (e.g. more closed-

ended than open-ended questions), providing suitable response options and organizing the questionnaire in a logical manner you reduce chances of questions being misunderstood and responses being misreported.

- Pretesting enables early detection of errors in the survey by testing whether or not questions are being understood and whether the responses accurately capture what is being measured. By making adjustments at this stage you reduce errors of random variability in responses.
- Sampling methodology will determine how well your sample represents the target population and thus has a direct impact on the sampling error and systematic sample bias. Detailed discussion on sampling in the next Module will demonstrate how this is achieved.
- Thorough training of interviewers ensures that they understand the questions clearly and record responses in the way you want them to; if interviewers fail to record data correctly the survey will be fraught with all kinds of errors.
- There are various steps in preparing data for analysis, i.e. coding, entering and cleaning and each step is designed to minimize errors and discrepancies in the collected data.

Further Reading

On Survey Methodology:

Measure DHS, Surveys and Methodology [Accessed online at: <http://www.measuredhs.com/aboutsurveys/>]

The World Bank, Living Standards Measurement Study (LSMS), Survey & Analysis Tools [Accessed online at: <http://econ.worldbank.org/WBSITE/EXTERNAL/EXTDEC/EXTRESEARCH/EXTLSMS/0,,contentMDK:21555895~menuPK:4196884~pagePK:64168445~piPK:64168309~theSitePK:3358997,00.html>]

United Nations Statistics Division, Designing Household Survey Samples: Practical Guidelines [Accessed online at: <http://unstats.un.org/unsd/demographic/sources/surveys/Handbook23June05.pdf>]

On Survey Design:

Bernard, H. Russell. *Social Research Methods: Qualitative and Quantitative Approaches* (Thousand Oaks, CA: Sage Publications, 2000) at Chapter 4: Research Design.

Creswell, John W. *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches* (Thousand Oaks, CA: Sage Publications, 2003) at Chapter 9, Quantitative Methods.

Examples of Questionnaires:

National Institute of Population Studies & Macro International Inc. *Pakistan Demographic and Health Survey 2006-07 Report* (Islamabad/Calverton, Maryland: June 2008) at 221-379. See: Appendix F, Questionnaires. [<http://www.measuredhs.com/pubs/pdf/FR200/FR200.pdf>]

The World Bank, Living Standards Measurement Study (LSMS) [<http://econ.worldbank.org/WBSITE/EXTERNAL/EXTDEC/EXTRESEARCH/EXTLSMS/0,,contentMDK:21610833~pagePK:64168427~piPK:64168435~theSitePK:3358997,00.html>]

United Nations Statistics Division, Household Sample Surveys in Developing and Transitional Countries [<http://unstats.un.org/unsd/hhsurveys/>]

References

Bernard, H. Russell. *Social Research Methods: Qualitative and Quantitative Approaches* (Thousand Oaks, CA: Sage Publications, 2000).

Creswell, John W. *Research Design, Qualitative, Quantitative and Mixed Method Approaches* (Thousand Oaks, CA: Sage Publications, 2003).

Fowler, J.Floyd. *Survey Research Methods* (Thousand Oaks, CA: Sage Publications, 2009).

Groves M. Robert, et al. *Survey Methodology* (Hoboken, NJ: John Wiley & Sons Inc., 2004).

Iarossi, Giuseppe. *The Power of Survey Design; A User's Guide for Managing Surveys, Interpreting Results, and Influencing Respondents* (Washington, DC: The World Bank, 2006). [Accessed online on August 3, 2011 at <http://issuu.com/world.bank.publications/docs/9780821363928>].

National Institute of Population Studies & Macro International Inc. *Pakistan Demographic and Health Survey 2006-7* (Islamabad, 2008).

Trochim, William M. *The Research Methods Knowledge Base*, 2nd Edition. [Accessed online at: <http://www.socialresearchmethods.net/kb/>] (version current as of October 20, 2006).

Quantitative Methods: Ensuring Scientific Accuracy

Module 4 – Quantitative Methods: Ensuring Scientific Accuracy

Purpose: To introduce the fundamentals of sampling.

This Module will introduce the reader to the concept of sampling, as well as descriptive statistics and descriptive analysis, which form the basis of any statistical analysis. This helps us to understand our variables before we can infer any associations or relationships from our data.

4.1 Theoretical Basis of Sampling

How well your sample represents the population and how confidently you can generalize your findings, depends on the precision of your sample selection process. This process includes identifying the appropriate sampling frame, determining sample size, reducing non-response and minimizing sampling errors.

The concept of generalization, i.e. the degree to which you can apply the findings of your study to similar populations in different places is known as **external validity** of the study and is often subject to a lot of scrutiny and critique. You could err in making generalizations and that is a **threat to external validity**. These threats emanate from the people you choose for your study. For example, they might be different from the general population, or the place where the study is done. The latter might be an unusual place or setting. Further, the timing of the study is important; the study might have been conducted at an extraordinary time (Trochim, 2006). For migration studies or studies concerning refugees, the people you interview will be different from the rest of the population (i.e. those not affected by the persecution/conflict/disaster causing displacement). Their situation and the timing will also be uncommon, due to the post-conflict, post-disaster context. That is only natural. However the findings of the study should be useful for future studies/programmes regarding other refugees or migrants displaced due to similar circumstances in different regions.

Threats to external validity can be mitigated by drawing an appropriate sample, where possible using random selection techniques, keeping respondent dropouts (or non-response) to a minimum and describing how your context differs from others by providing “data about the degree of similarity between various groups of people, places, and even times”(Trochim, 2006). Nevertheless, it is not always necessary that you would want to generalize your findings. You might just want to study one particular group in one place at a given time, you might want to conduct a programme evaluation or a rapid assessment and in such cases generalization is not a concern.

The group of your research interest, i.e. the group about which you would want to generalize your findings to is known as the **population** of your study. For instance, if you wish to study the effects of conflict-related displacement on gender-based violence and abuse, your study population would be all women displaced by conflict. It might not be feasible to access the entire population for your survey, therefore you will identify the accessible portion of the population, i.e. the **sampling frame**, from which you will draw your sample. Continuing with the previous example, your sampling frame could be women displaced due to conflict during a certain time period, say during the 2009 military action in Khyber Pakhtunkhwa (KPK) and the tribal areas, who have been relocated to 5 camps identified in KPK. For the sampling frame you might have some existing list on which you can rely to select your sample,

such as registered Afghans or registered internally displaced persons (IDPs), or you might have to do your own census to generate a sampling frame. The latter would require interacting with the community of interest, walking around, and asking questions.

4.1.1 Typology of Sampling Techniques

How the sample is selected from within the sampling frame depends on the kind of data you require; i.e. whether you want data about individuals or you want cultural data. “Individual data are about attributes of individuals in a population”, such as age, income, education, etc (Bernard, 2000: 144). Cultural data requires experts “who can offer expert explanations and who represent intra-cultural variations that we find in all societies” (Bernard, 2000: 144); for instance why do parents discriminate against girls in some cultures, or why do some women choose to go to unsafe health service providers for delivery while others do not?

When individual data is required, to determine specific characteristics of a population, a sample needs to be drawn scientifically, so that every individual/unit of analysis in the sampling frame has an equal chance of being selected. This is known as **random sampling** or **probability sampling** and is the preferred method of sample selection in social science research. However where cultural data is required or where random sampling is just not possible, **non-probability sampling** techniques are used. The most common probability and non-probability methods are (Bernard, 2000):

Probability Sampling

- **Simple Random Sampling:** All units in the sampling frame are assigned a unique number and selected one by one until the desired sample size is reached. For instance if you have 1,000 individuals in your sample frame and you wanted a sample of 100, you would either use a computer program, such as STATA or SPSS⁸, or random number tables to select these 100 individuals. Computer programs are more convenient, especially for larger samples. A list of random numbers is generated and applied to the sampling frame; numbers of individuals that correspond with the random numbers generated will be selected.

For instance, there are 1,000 registered household heads in a particular IDP camp. Their particulars are entered into a database and each one is assigned a number from 1 to 1,000. Your computer program generates a list of a 100 random numbers for you to apply to the sampling frame. Say the first 10 numbers in the list are: 8, 12, 5, 46, 22, 3, 59, 10, 36 and 73. Then the first 10 household heads to be included in your sample will be the ones who have been assigned these numbers in your database.

- **Systematic Random Sampling:** Where numbered lists of sampling units might not always be readily available, simple random sampling becomes unfeasible. One alternative is systematic random sampling, which requires a random start and a sampling interval. Most researchers usually use systematic random sampling instead of simple random sampling. The sampling interval is obtained by dividing the required sample size by the number of units in the sampling frame. If a sample of 100 is needed from a sampling frame of 10,000, every 100th unit will be selected starting from a random number within the sampling interval of 1 to 100.

Suppose, in the example used to illustrate simple random sampling above, there is no list of registered household heads. In this case, you enter the names of the members of the potential sampling frame, assign them numbers from 1 to 1,000, select a random number to start with, for example 55 and then keep adding every 10th member to the sample (i.e. 65th, 75th, 85th and so on) until you have a sample of one hundred respondents.

- **Stratified Random Sampling:** By dividing the sampling frame into sub-groups/sub-frames, this method ensures that important sub-populations are included in your sample. These sub-frames

⁸ SPSS (Statistical Package for Social Sciences) and STATA are both computer-based statistical data analysis programs.

are created on the basis of key independent variables such as age groups, gender, urban/rural residence, etc. and then random samples are taken from every group. Through stratification, i.e. sampling from smaller groups, we create more homogenous sub-groups thereby **minimizing within-group variance** and at the same time introduce more variance, in terms of key independent variables, between the sub-groups, i.e. we **maximize between-group variance**. This technique also reduces sampling error.

For instance, in our sample of household heads, we want to ensure that female headed households are also represented. For this purpose you divide your sampling frame into male and female household heads and then do a random selection of respondents from each strata (group). Suppose there are 300 households that are headed by females. When selecting your sample members you can either have a **proportionate** or a **disproportionate** representation of the female and male household heads. If you chose *proportionate stratified random sampling technique* then 30% of your sample will be female and 70% male. *Disproportionate random sampling* is used when you feel that an important sub-population will be underrepresented in a simple stratified random sample. Suppose in this example, female headed households only represented 5% of the population of household heads. Using proportionate random sampling you would only have 5 female household heads in your sample. To ensure you have sufficient data on this important sub-group you could increase the number of female household heads in your sample by a factor of 5. Therefore the proportion of females and males in the sample will be 25% and 75% respectively, i.e. there will be disproportionately greater female representation.

- **Cluster Sampling:** Also known as multistage cluster sampling, this method is used when there are no frames or lists available to choose the sample from. Naturally existing groups or clusters such as geographic demarcations (tehsil, union councils, mohalla/neighborhood, etc) or institutions (schools, mosques, an industry, etc.) are used to collect the sample in smaller, more manageable, homogenous groups. For instance, you might not be able to find a single list of registered IDPs in all relief camps set up in a province however you might find a list of the camps themselves. You would begin by sampling camps to narrow the list and then sample IDPs within the selected camps. Here, again, the idea is to narrow down from large heterogeneous chunks to smaller, more manageable, homogenous clusters.

Non-Probability Sampling

- **Quota Sampling:** As with stratified random sampling, in quota sampling you divide the population into sub-groups that are of importance to your research. Then you decide on the proportion of each sub-group in your entire sample. Unlike random stratified sampling, in this case the participants are not picked randomly beforehand rather they are selected on the spot. For instance, you go into the field knowing you have to interview 25 female household heads and 75 male household heads within a particular camp, not knowing beforehand whom you will interview. You select respondents, using some predetermined criterion (for instance go to every 5th house to the left of the first selected household) and continue till you have fulfilled the quota.
- **Purposive/Judgment Sampling:** This method is generally used when there are no reliable sampling frames or your target population is likely to be hidden (such as drug addicts, trafficking victims, etc.) or special (such as religious leaders, professionals such as lawyers). You identify the purpose that you want the respondents to fulfill and then use your judgment (or that of field experts) to identify respondents who you think will satisfy your research purposes. This method is also useful when conducting pilot studies or where a few critical/intensive case studies are required. For instance, to identify victims of domestic violence/abuse, you could contact a female councilor or community-based rights activist to assist you in developing the appropriate strategy to approach and identify the right respondents for your study.
- **Convenience/Haphazard Sampling:** This technique relies on interviewing anyone who would be willing to participate in your survey. This is best suited to preliminary/exploratory research and

pretesting of questionnaires. The danger in using this method is that your sample is unlikely to be representative and it lacks purpose.

- **Snowball Sampling:** In this sort of sampling a few key respondents are identified who are then asked to identify others like them who could participate in the research. This technique is particularly useful where social networks are to be studied or where the population of interest is otherwise difficult to locate (or hidden) or is too small. Again, taking the example of respondents for the domestic violence study, once you are able to identify a couple of relevant respondents, you could request them to identify other similar victims who would be willing to partake in the study. You keep on going till you have exhausted all the possible connections or till you reach the desired number of participants.

4.1.2 Sample Size Determination

Determinants of Sample Size

In order to be considered precise a sample not only needs to be unbiased (as discussed in the Module 3, Section 3.3 and in Section 4.1.1 above) but also of an appropriate size. Although there is no clear-cut answer to how big a sample should be, increasing sample size does increase the reliability of survey estimates (Fowler, 2009). Nevertheless researchers have identified some important factors that can be used to guide sample size determination:

- Degree of Variability or Heterogeneity of the Population.** The more heterogeneous the population of interest in terms of variables being measured (e.g. education, socio-economic status, ethnicity), the bigger the required sample will be, and the lesser the variability in the attributes being studied (i.e. the more homogenous the population), the smaller the sample will be (Israel, 1992). For instance if you are studying the role of caste or ethnicity in people’s attitudes towards migration then you need to ensure you have significant variation in your sample in terms of ethnicity of respondents; respondents of one caste or ethnic background will not fulfill the purpose. Therefore the more ethnicities you decided to include in your study the bigger your sample.
- Number of Subgroups Required in the Analysis.** (Bernard, 2000) If you are studying several different independent variables with various attributes, you will need a reasonable number of subjects in each subgroup for your estimates to be meaningful. Consider the following table of independent variables and their respective attributes:

Table 4.1 Variables and Attributes

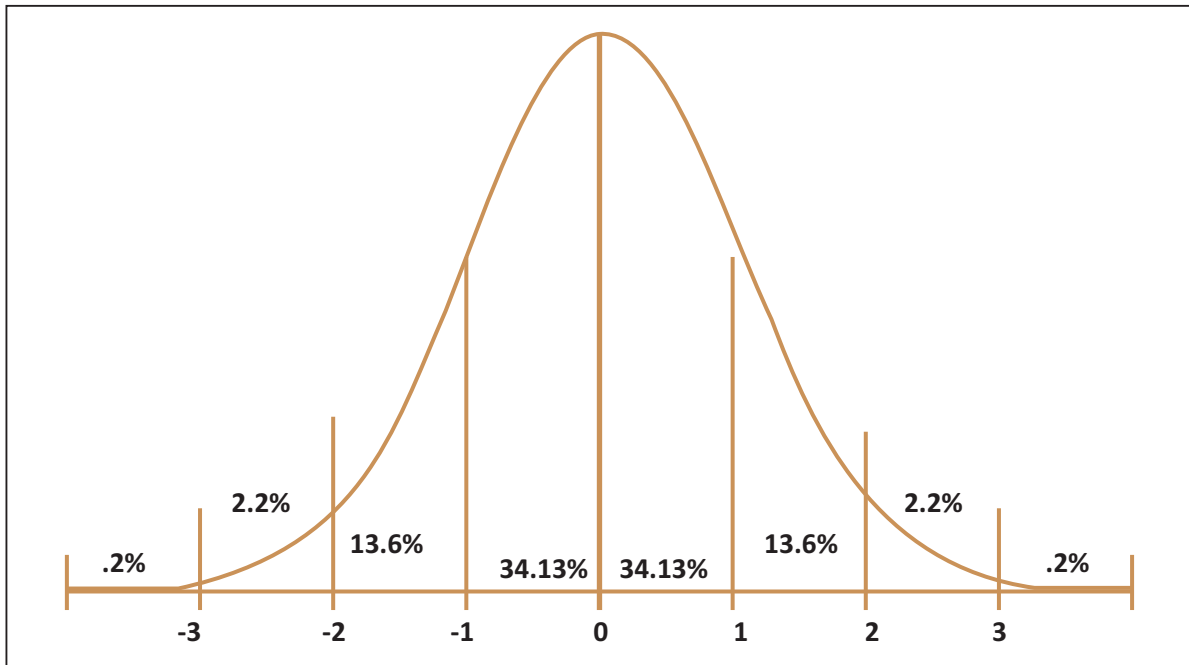
Independent Variables	Attributes
Household head	Male/Female
Age of household head	Can be any number reported by respondent
Education of household head	Can be anywhere between ‘No Education’ to ‘Graduate’ or beyond
Sources of livelihood	Can be any profession from farming to casual labourer to government service to private employment or business

For each of the four independent variables there at least two attributes therefore producing a multiple cell analysis design. If you gather a sample of say 50 respondents, there might not be enough subjects in each category of analysis (for instance female household heads between the ages of 30 and 35 years) to say anything meaningful.

- Standard Error or Precision.** To understand standard error we must first go over the concepts of **normal distribution, standard deviation, and the central limit theorem.**

The **normal distribution**, shown in the diagram below, has a mean of 0 and a standard deviation of 1. **Standard deviation** “is a measure of how much the scores in a distribution vary from the mean (average) score”. It tells you how homogenous or heterogeneous your population is. The greater the standard deviation the more heterogeneous the population (Bernard, 2000).

Figure 4.1: Percent of Normal Distribution Scores in Each Interval



In the diagram of the normal curve, the markings on the x-axis show the mean (0) and standard deviations from the mean, ranging from -3 to 3 and the percentages mentioned within the columns under the curve (34.13%, 13.6%, and so on) depict the area under the curve covered by the standard deviations. The area under the curve is determined using standard Z-score tables, such as the one provided in Appendix 3, Z-Score Table (or Table of Areas under the Curve). Using the table we see that the area under the curve for a standard deviation of 1 above the mean is 34.13%. Similarly standard deviation of 1 below the mean is also 34.13%. Therefore 68% (34.13%+34.13%) of all the scores in a normal distribution fall within ± 1 standard deviation of the mean. From Appendix 3 we also see that 95.44% of all scores fall within ± 2 standard deviations (47.72%+47.72%) and area covered by ± 3 standard deviations is 99.7%. Usually these figures are rounded off and used as 95% and 99% instead. According to Appendix 3, the areas under the curve corresponding with these percentages are 1.96 and 2.58 standard deviations respectively [Note: look for half of 95%, i.e. 0.4750 and half of 99%, i.e. 0.4951].

Now that we know what normal distribution is we can discuss the **central limit theorem**, which says that if many samples of sufficient sizes are taken from the same population: “(i) the mean and standard deviation of the sample means will usually approximate the true mean and standard deviation of the population; and (ii) the distribution of sample means will approximate a normal distribution” (Bernard, 2000:164). While overall sample means will converge to the true population mean, each sample drawn from the population will have a different sample mean. The standard deviation of the mean of a particular sample (i.e. of using the sample mean to approximate the population mean) from the overall population mean is the standard error of the sample mean. The general formula for calculating standard error is (Bernard, 2000: 169) [Note: while you might not always need to use these formulas, it is important to understand these concepts and know how standard error affects sample size]:

$$SE = \frac{SD}{\sqrt{n}} \text{ (Formula 1)}$$

Where SE= standard error, SD= (sample) standard deviation and n=sample size. Thus the equation for sample size would be:

$$\sqrt{n} = \frac{SD}{SE} \text{ (Formula 2)}$$

Going back to the discussion on normal distribution, we can now say that 95% of all samples, of a given size, selected from a given population, will produce an estimate of ± 1.96 standard errors and 99% of the samples will produce ± 2.58 standard errors. The 95% and 99% figures represent confidence levels. The respective confidence intervals are calculated using the standard deviation figures of ± 1.96 and ± 2.58 as multipliers with standard error, i.e. $1.96 \times (SE)$ or $2.58 \times (SE)$. Therefore the true population mean (μ) would be:

$$\mu = \bar{x} \pm 1.96 \times (SE) \text{ or } \bar{x} \pm 2.58 \times (SE), \text{ where } \bar{x} \text{ is the sample mean}$$

Using these equations we can say with some confidence that 95% or 99% of the confidence intervals will contain the true mean (μ) (Bernard, 2000:169). That is, we can use the estimate of mean, of any variable in our study, to determine the true mean of the variable in the population using the confidence intervals for standard deviation, ± 1.96 or ± 2.58 depending on the chosen confidence level, i.e. 95% or 99%.

From the standard error formula above we can see that if we want to reduce the standard error, we have to increase the sample size. It makes intuitive sense as well since the larger the sample size the closer the sample will be to the population hence chances of error will be lower. Furthermore if we want to increase our confidence level from 95% to 99% we can use the 2.58 multiplier instead of 1.96 with $\frac{SD}{SE}$ again increasing the sample size.

A common misconception is that a sample has to represent some considerable proportion of the target population (Fowler, 2009). The size of the population from which the sample is to be selected might not be important in determining sample size; a sample of 150 people can describe a population of 15,000 as well as a population of 150,000. This fact has been demonstrated by researchers as far back as the 1960s/70s. For instance, show that “as the population increases the sample size increases at a diminishing rate and remains relatively constant at slightly more than 380 cases” (Krejcie & Morgan, 1970). Therefore you should not determine the sample size by specifying some fraction of the population to be included in the sample since that will not reduce sampling errors.

How to Determine Sample Size

If you are working with a small population to begin with, say 100 to 200 people, you might just conduct a census (i.e. study the entire population) instead of selecting a sample, which might become a tedious process. If you are working with larger populations (say 10,000 or 100,000 people) it is more likely that you will have to take a sample. There are several ways in which you can determine your sample size. While the methods listed here might not all be truly scientific they can facilitate sample size determination- as you would have gathered by now, accurate and appropriate samples are difficult to design! Some of the strategies are:

- **Reviewing other similar studies.** During your literature review you will come across quantitative studies by others which might give you some ideas of how big or small your sample needs to be. While it is good to take guidance from such studies, it is not advisable to use the exact same sample sizes since the circumstances and constraints within which any two studies take place are different. When you take this approach it is extremely important to study and understand the survey design

processes of the reference studies to ensure you don't repeat any of the mistakes and that there are similarities between your study and those other studies.

- **Use of published sample size tables** (Israel, 2009). There are some tables available that have predetermined sample sizes for a given set of criterion, typically the size of the population, confidence interval/margin of error. An example of such a table and the formula used to determine sample sizes is given in Box 4.1. Using this table (Box 4.1), given a confidence level of 5,000 for required margin of error of 2.5% and confidence level of 95%, the appropriate sample size would be 1,176. For a larger margin of error like 5% the sample size will be smaller, i.e. 357. For a higher confidence level, i.e. 99% using the same margins of error the required sample size would go up.

These tables should however be used with caution since they assume that the sample will be randomly selected and that the attributes being measured are distributed normally (Israel, 2009). While you would have determined the overall sample size required for your study, if there are subgroups within your population you need to do some additional calculations to ensure that subgroups are represented adequately in your sample.

- **Use of online resources and software packages.** Some websites provide free online sample size calculators that can be used to estimate the appropriate sample size for a study. Usually you are required to feed in the population size, confidence level and desired precision (confidence interval/margin of error). A few of these websites are listed below:
 - MaCorr Research Solutions: [http://www.macorr.com/ss_calculator.htm]
 - Creative Research Systems: [<http://www.surveysystem.com/sscalc.htm>]
 - Raosoft: [<http://www.raosoft.com/samplesize.html>]

A more precise but more costly method of sample size determination is the use of software packages. One such package is the IBM SPSS SamplePower software⁹. If you plan to conduct quantitative studies on a regular basis, it might be worth investing in reliable sample size calculation software. These software packages come with user guides that take the user through the process of sample size calculation.

At this point you might ask, *what if the population in question is unascertainable?* There will not be a list of homeless people available, for example; neither will you find any records of trafficked persons; and what about 'illegal' migrants who lack any identifying documents whatsoever? What helps in such situations is a review of other studies on the subject or related subjects. Other studies might have tried to estimate the population size or they could at least tell you if you can expect the population to be small or large. As mentioned earlier, in case of large populations, the population size does not really impact the sample size; a sample of 400 would suffice for a population of 50,000 or 100,000. If no relevant statistics exist, you could consult field experts and get some sort of a consensus on the prevalence of the phenomenon in question.

Whichever method of sample size determination you choose, be mindful of your financial resources and time constraints. A larger sample will require more time and resources; while it might be ideal to include 1,000 people in your sample, you might only have resources to conduct 300 interviews. Whatever sample size you chose to work with, be clear on the sample selection methodology and be able to justify your sample size.

Sample Weights

Despite all the precautions you take in sampling there might still be a possibility that not all those selected in the sample will respond; some are likely to opt out of the study or might not be available at the time of the interview (non-response). To compensate, researchers often increase the sample size by 10% or more (Israel, 2009). While issues of non-response can be dealt with by increasing sample

⁹ For more information, see:

[http://www14.software.ibm.com/download/data/web/en_US/trialprograms/U741655I36057W80.html?S_CMP=rnav]

Box 4.1 Sample Size Table

Population Size	Required Sample Size*							
	Confidence = 95%				Confidence = 99%			
	Margin of Error				Margin of Error			
	5.0%	3.5%	2.5%	1.0%	5.0%	3.5%	2.5%	1.0%
10	10	10	10	10	10	10	10	10
20	19	20	20	20	19	20	20	20
30	28	29	29	30	29	29	30	30
50	44	47	48	50	47	48	49	50
75	63	69	72	74	67	71	73	75
100	80	89	94	99	87	93	96	99
150	108	126	137	148	122	135	142	149
200	132	160	177	196	154	174	186	198
250	152	190	215	244	182	211	229	246
300	169	217	251	291	207	246	270	295
400	196	265	318	384	250	309	348	391
500	217	306	377	475	285	365	421	485
600	234	340	432	565	315	416	490	579
700	248	370	481	653	341	462	554	672
800	260	396	526	739	363	503	615	763
1,000	278	440	606	906	399	575	727	943
1,200	291	474	674	1067	427	630	827	1119
1,500	306	515	759	1297	460	712	959	1376
2,000	322	563	869	1655	498	808	1141	1785
2,500	333	597	952	1984	524	879	1288	2173
3,500	346	641	1068	2565	558	977	1510	2800
5,000	357	678	1176	3288	586	1066	1734	3842
7,500	365	710	1275	4211	610	1147	1960	5165
10,000	370	727	1332	4899	622	1193	2098	6239
25,000	378	760	1448	6939	646	1285	2399	9972
50,000	381	772	1491	8056	655	1318	2520	12455
75,000	382	776	1506	8514	658	1330	2563	13583
100,000	383	778	1513	8762	659	1336	2585	14227
250,000	384	782	1527	9248	662	1347	2626	15555
500,000	384	783	1532	9423	663	1350	2640	16055
1,000,000	384	783	1534	9512	663	1352	2647	16317
2,500,000	384	784	1536	9567	663	1353	2651	16478
10,000,000	384	784	1536	9594	663	1354	2653	16560
100,000,000	384	784	1537	9603	663	1354	2654	16584
300,000,000	384	784	1537	9603	663	1354	2654	16586

* Copyright. The Research Advisors (2006). All rights reserved

$$n = \frac{\chi^2 * N * P * (1-P)}{(ME^2 * (N-1)) + (\chi^2 * P * (1-P))}$$

Where:

n = sample size

χ^2 = Chi-square for the specified confidence level at 1 degree of freedom

N = Population Size

P = population proportion (.50 in this table)

ME = desired Margin of Error (expressed as a proportion)

Source: Accessed online on July 27, 2011 at: [<http://research-advisors.com/tools/SampleSize.htm>]

size, imperfections that relate to non-coverage of groups within the population can be dealt with at the data analysis stage through the use of **sample weights**.

Weighting is an effort to bring the sample closer to the actual population. It requires multiplying survey responses by weighting factors or base weights (Yansaneh, 2003). These weights are calculated by dividing the proportion of a group in the population by the proportion of the group in the sample. To illustrate with an example, suppose you want to conduct a study on IDP youth in the 12 districts of KPK that are directly adjacent to the Federally Administered Tribal Areas (FATA) (i.e. Dera Ismail Khan, Tank, Banu, Lakki Marwat, Karak, Hangu, Kohat, Peshawar, Charsadda, Malakand, Dir). However due to security concerns you cannot cover some of the target districts, say three of them (Charsadda, Hangu and lower Dir). Not gathering data from these geographic areas might bias the outcomes of your study but you cannot risk the lives of your team members by sending them there. The way out is to assign weights to the cases (individuals or households) in your sample. To compensate for the non-coverage of the 3 districts you will begin by tabulating the population of all the target districts and determining the proportion of population that each district represents:

Table 4.2 Population and Proportion of Population Represented

District	Population (from official census records) A	Proportion of Population B
Dera Ismail Khan	A	a/X
Tank	B	b/X
Banu	C	c/X
Lakki Marwat	D	d/X
Karak	E	e/X
Hangu	F	f/X
Kohat	G	g/X
Peshawar	H	h/X
Charsadda	I	i/X
Malakand	J	j/X
Upper and Lower Dir	K	k/X
Total	X	

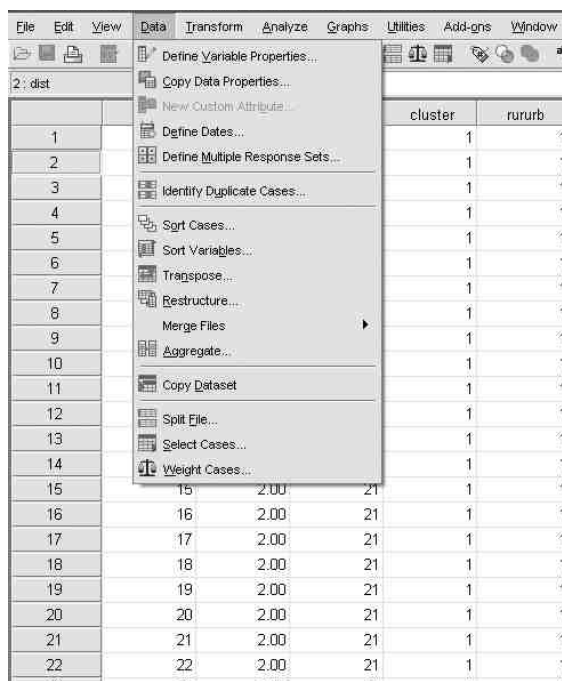
Next you will distribute the population of the non-covered districts among the districts that are represented in your sample. One way would be to add the populations of the non-covered districts to the populations of the geographically adjacent district, for instance the population of Charsadda could be added to Peshawar or Malakand, lower Dir could be added to Malakand and so on. Another possible way of redistribution could be through identifying districts that are ethnically similar to the non-covered districts and adding the population to the closest match. However you chose to redistribute you will now recalculate the new population proportions where the overall population will remain the same (i.e. X) but the proportions reported in Column B will change since the populations of 3 districts not represented in your sample have been redistributed. The new district population proportions will then be divided by the sample district proportions to calculate weights:

Table 4.3 Calculation of Weights

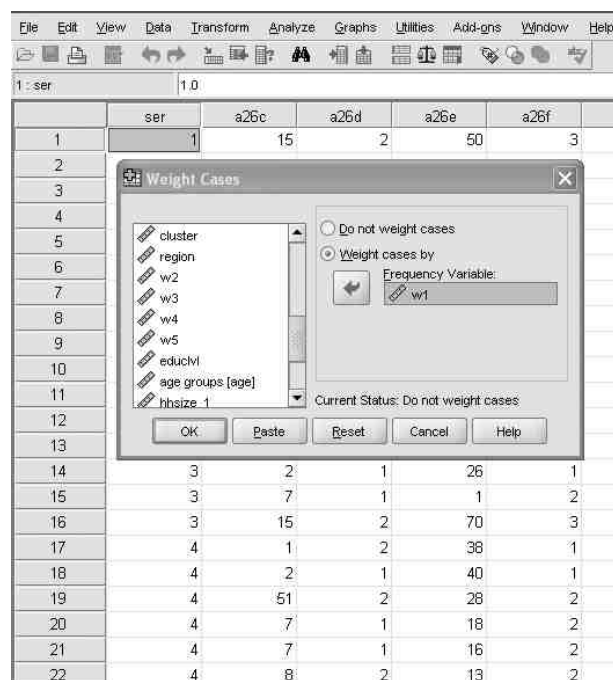
District	District Code for SPSS	Population A (from official census records)	Proportion of population B	Sample district population	Proportion of sample district population	Sample weight (w1)
Dera Ismail Khan	1	a1	$a1/X$	aS	aS/X	$(a1/X) \div (aS/X)$
Tank	2	b1	$b1/X$	bS	bS/X	$(b1/X) \div (bS/X)$
Banu	3	c1	$c1/X$	cS	cS/X	$(c1/X) \div (cS/X)$
Lakki Marwat	4	d1	$d1/X$	dS	dS/X	$(d1/X) \div (dS/X)$
Karak	5	e1	$e1/X$	eS	eS/X	$(e1/X) \div (eS/X)$
Kohat	6	g1	$g1/X$	gS	gS/X	$(g1/X) \div (gS/X)$
Peshawar	7	h1	$h1/X$	hS	hS/X	$(h1/X) \div (hS/X)$
Malakand	8	j1	$j1/X$	jS	jS/X	$(j1/X) \div (jS/X)$
Only Upper Dir	9	k1	$k1/X$	kS	kS/X	$(k1/X) \div (kS/X)$
Total		X		XS		

Once you have calculated the sample weights, they will have to be entered into SPSS like other variables so that they can be assigned to the corresponding groups, i.e. multiplied with the responses to get weighted results. In the SPSS 'Data' menu you are given the option of 'Weight Cases', which when selected asks you for the variable you want to use to weight your data, i.e. 'Weight cases by'. Here you would choose the newly created weight variable w1. Please follow screenshots 1 and 2 below:

Screenshot 1



Screenshot 2



Remember that applying weights is useful when “representative levels of statistics are desired, such as percentages, means, and medians; use of sample weights is inappropriate for estimating relationships, such as regression and correlation coefficients”¹⁰.

4.2 Non-Sampling Errors

Throughout the text of Module 3 and Module 4 we have referred to several errors that are associated with the survey process. For the sake of clarity, in this section, we will recap these errors and elaborate more on non-sampling errors. There are two types of errors associated with surveys, sampling errors and non-sampling errors, and the two taken together constitute total survey error. While sampling errors stem from the fact that data is collected from a sample rather than the whole population, non-sampling errors are a result of measurement, data recording and processing. Unlike sampling errors, chances of non-sampling errors increase with an increase in sample size (Banda, 2003). We have already gone over what sampling errors are and how they can be controlled. Among the possible sources of non-sampling errors are the following (Banda, 2003):

- Data descriptors being inconsistent with survey objectives
- Use of misleading definitions/concepts
- Inappropriate methods of interview/observations
- Omission of units of analysis (i.e. individual participants) due to non-coverage or non-response
- Duplication of units of analysis
- Untrained/under-trained enumerators/supervisors
- Errors in data entry into computer system, faulty coding, inadequate data checking/cleaning
- Errors in data analysis and presentation

Some of the ways in which these errors can be mitigated are:

- Non-coverage: use reliable sampling frames or make your own sampling frames/census; use of weights at the data analysis stage.
- Non-response: proper training of interviewers; clear respondent confidentiality statements; follow-up interviews for incomplete interviews; identification of alternative respondents (although enumerator discretion in selecting substitute respondents should be kept to a minimum), close supervision of fieldwork; keeping questionnaires as simple as possible.
- Measurement errors: clearly defining research objectives and the concepts central to your research; keeping questions unambiguous and concise; pre-testing survey to check how relevant and comprehensible the questions are; training enumerators to ensure they understand the purpose of the research as well as the objective of each question and are able to get the message across to the respondents.
- Processing errors: clear data management protocols; clear/unambiguous coding manuals or codebooks; supervision of data entry; performing data checks to clean entered data.

4.3 Descriptive Statistics

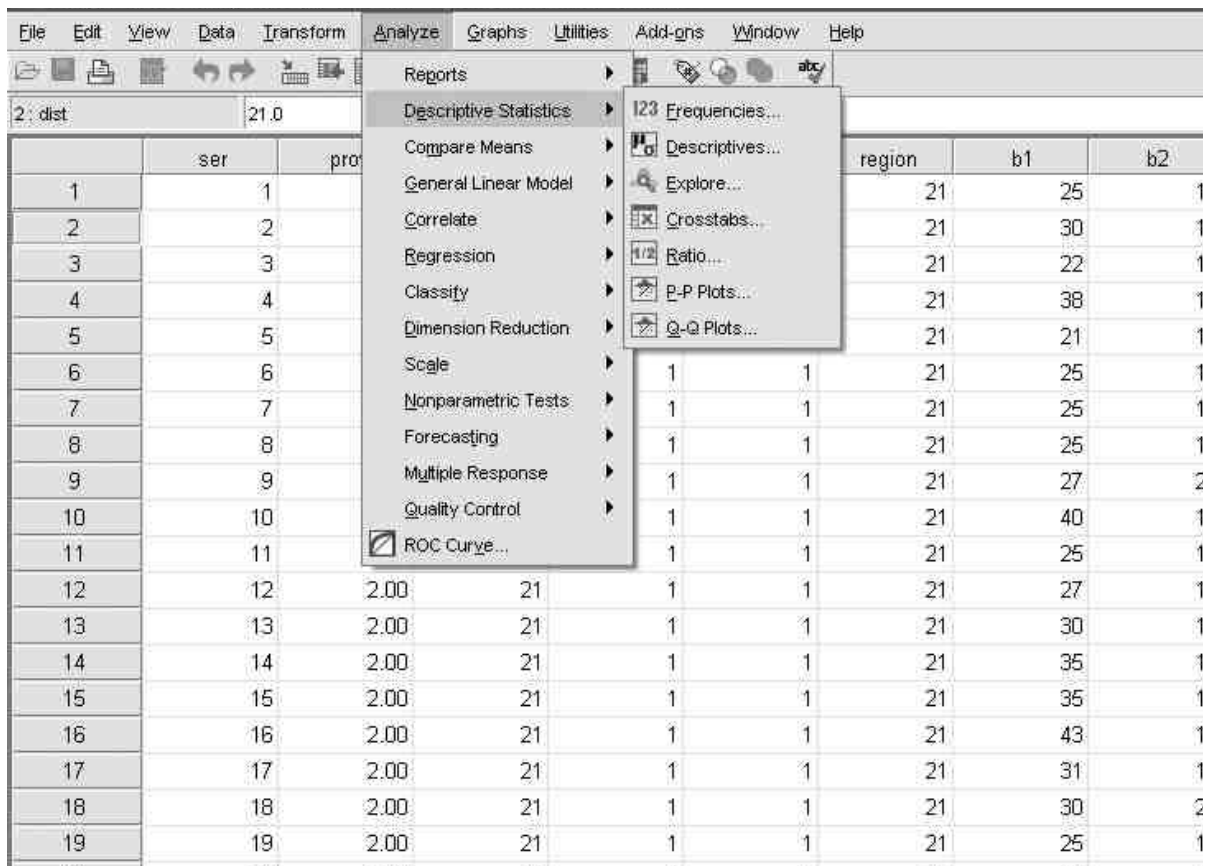
There are three types of analyses that can be performed on data: univariate, bivariate and multivariate. Univariate analysis involves examining variables precisely and really getting to know your data, whereas in bivariate analysis you would look at associations between pairs of variables and in multivariate analysis you would look at the “effects of more than one independent variable at a time on a dependent variable” (Bernard, 2000:502). Univariate analysis makes use of descriptive statistics such as frequency distributions, measures of central tendency and measures of dispersion.

¹⁰ DHS Statistics Live, Online Guide to DHS Statistics. [Accessed online on July 29, 2011 at : http://www.measuredhs.com/help/Datasets/index.htm#sampling_weights.htm].

4.3.1 Frequency Distributions and Measures of Central Tendency

Frequency distribution simply shows how a variable, such as age, education, income, is distributed amongst the sample. When looking at frequencies we are looking for variability in our data; variables that remain constant over the sample are of no interest in further analysis. For instance if a survey only interviewed female respondents then the 'gender' variable would be of no use in the analysis. This is however deliberate, i.e. the purpose of your survey might be to interview only females. What if your all-female sample is also similar in terms of other variables for instance, age, marital status? If the frequencies of these variables do not show any variability then it is best not to use them in further analysis.

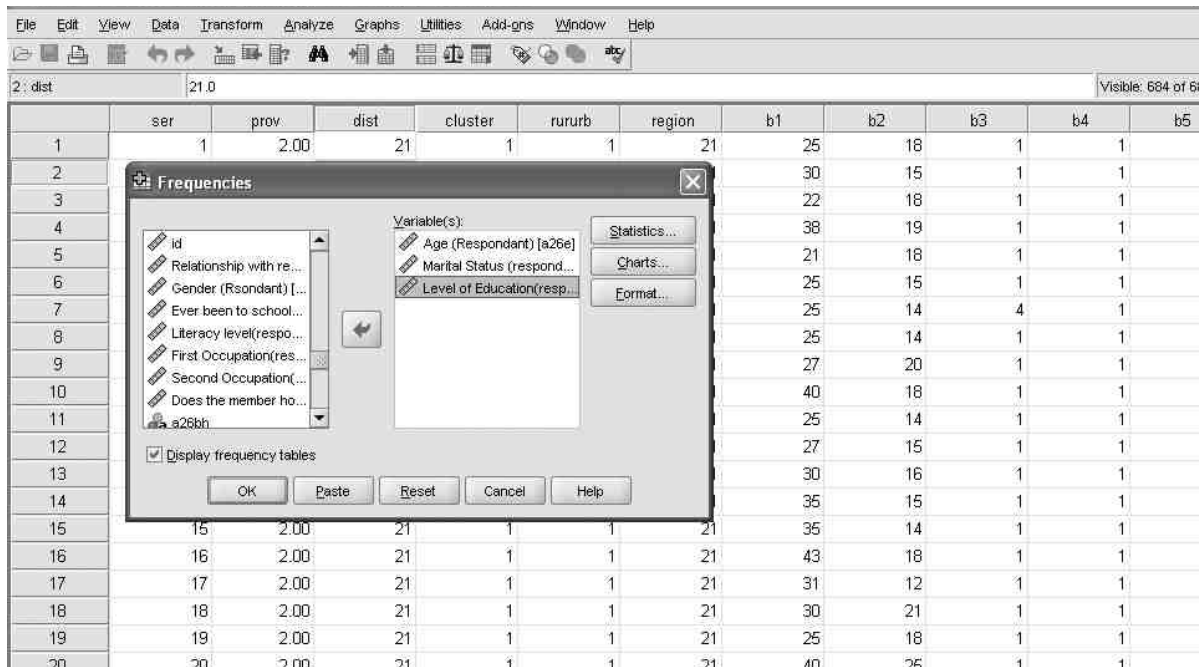
Screenshot 3



How are frequencies determined? Any statistical data analysis program will calculate frequencies for you. If you are using SPSS (version 13) then in the 'Analyze' data menu you will select 'Descriptive Statistics', under which you will select 'Frequencies'. Screenshot 3 demonstrates this process.

Once you have selected the 'Frequencies' option, the system will ask you to identify the variable for which you require frequencies. You can select more than one variable at a time. Please refer to screenshot 4:

Screenshot 4



The results will be displayed after you select 'OK' and will appear as¹¹:

Table 4.4 Marital Status

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Married	686	98.1	98.1	98.1
	Widowed	5	.7	.7	98.9
	Divorced	2	.3	.3	99.1
	Separated	6	.9	.9	100.0
	Total	699	100.0	100.0	

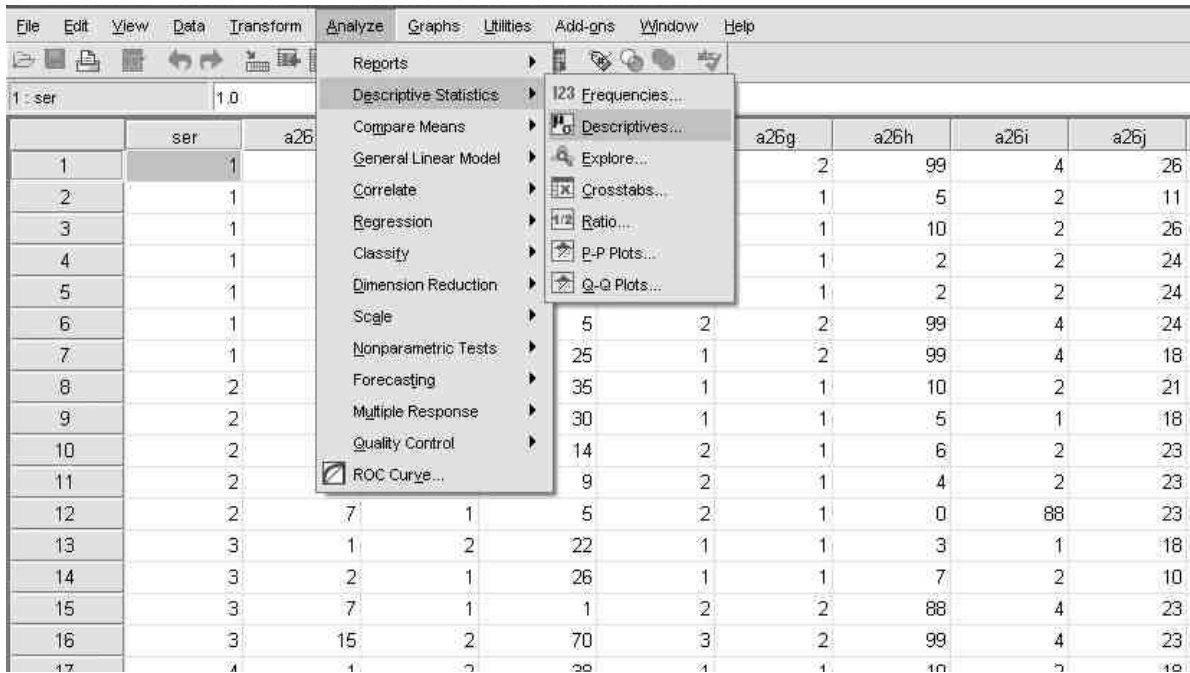
Looking at Table 4.4, we can see that 98% of the sample is currently married whereas close to 2% are divorced, widowed or separated.

Measures of central tendency measure the typical values of variables. The most common measures of central tendency are:

- i. **The Mean.** This is the average value for a variable, i.e. the sum of individual scores divided by the number of scores. You might want to know what the average age at marriage of your sample, the average income of your sample or average family size. Screenshot 5 shows how the mean is extracted using SPSS:

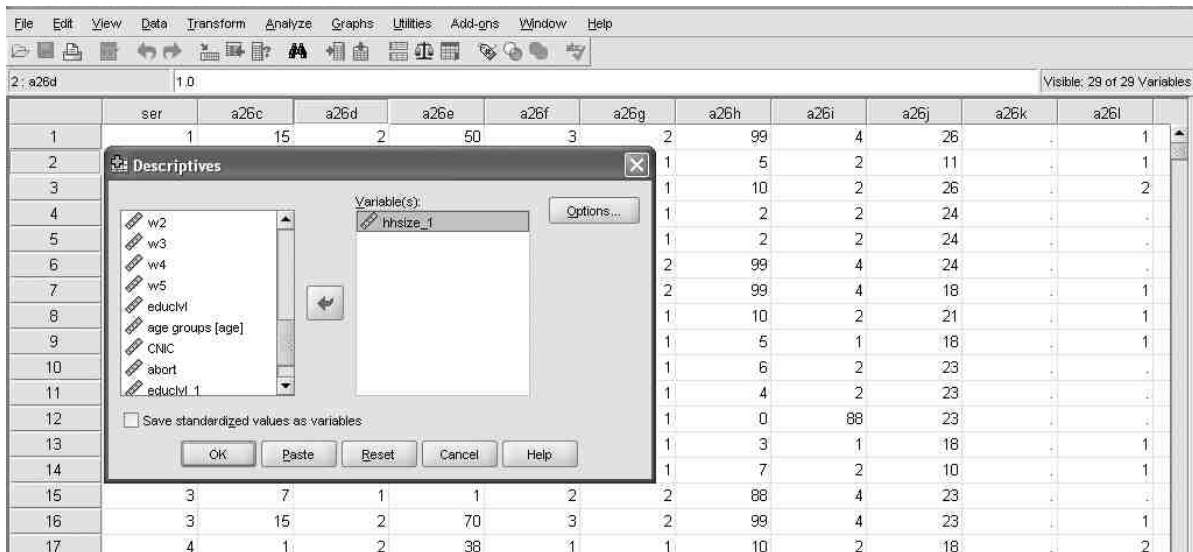
¹¹ Data used in these examples is the property of Collective for Social Science Research, Karachi.

Screenshot 5



After selecting 'Descriptives' you will select the variable, as shown in screenshot 6, for which you require the mean, in this example we want to know the average household size:

Screenshot 6



The resulting table produced by the system will look like this:

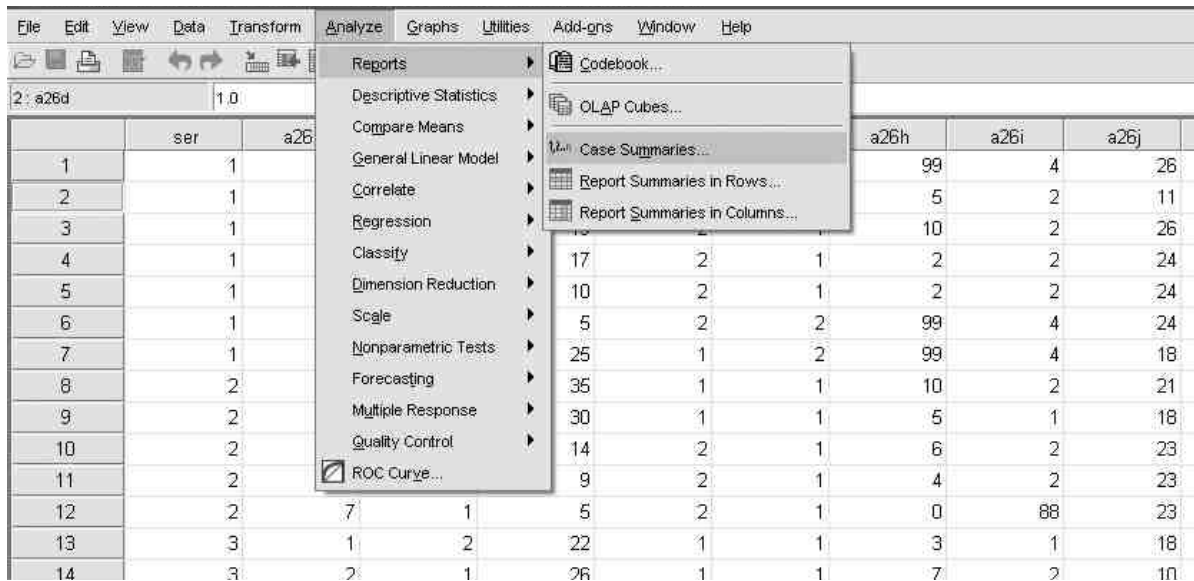
Table 4.5 Descriptive Statistics (1)

	N	Minimum	Maximum	Mean	Std. Deviation
Hhsize_1	5073	2.00	29.00	8.3885	3.60941
Valid N (listwise)	5073				

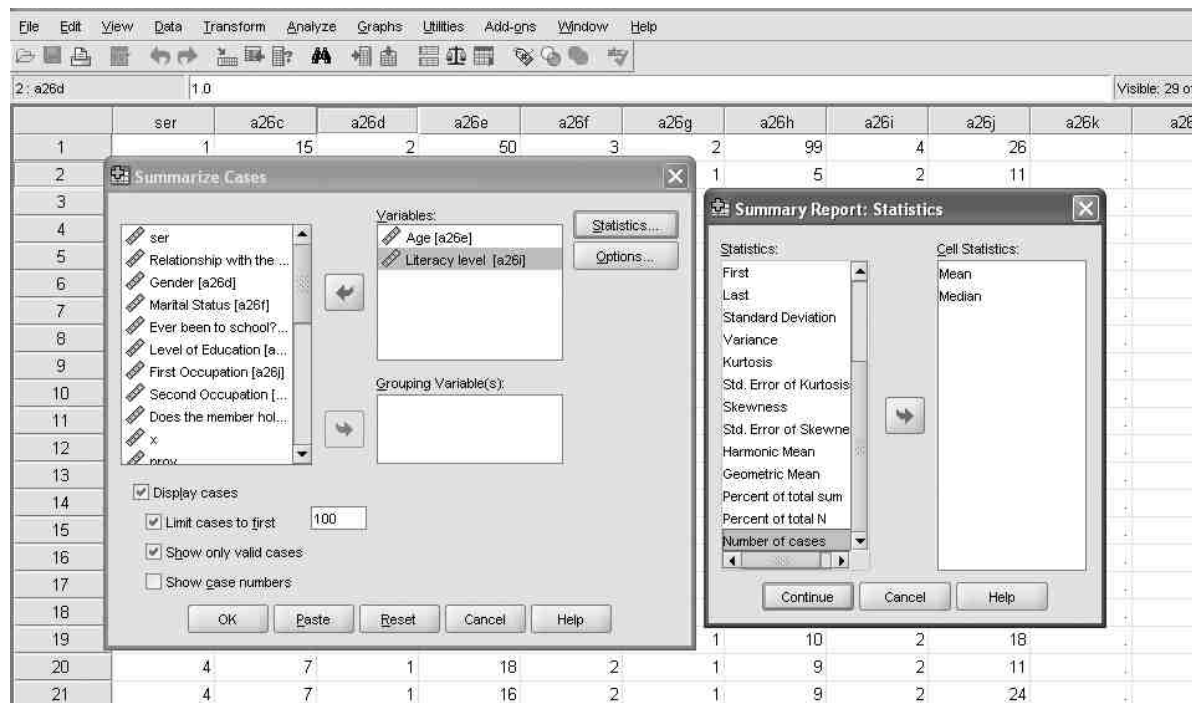
We can see from the table 4.5 that the average household size in the sample is 8.4 members with a minimum of 2 household members and a maximum of 29.

- ii. **The Median.** This is the mid-point of the distribution of any variable, i.e. there are equal numbers of scores (observations) above and below this point. To determine the median, under 'Analyze' you will select 'Reports' and then 'Case summaries', as in Screenshot 7, which will open up the window where you select the relevant variables as well as specify the statistics you require. As you can see several statistics from Screenshot 8, including the mean and median, can be extracted through 'Case summaries':

Screenshot 7



Screenshot 8



- iii. **The Mode.** This is the value or attribute that occurs most frequently for a variable. It is observed by looking at the frequency distribution of the variable. Going back to Table 4.4 to give an example, we see that majority of the sample is in the “married” category i.e. 98% of the respondents.

4.3.2 Measures of Dispersion

Measures of dispersion tell us how homogenous or heterogeneous our sample is, i.e. how dispersed our variable attributes are. You want to know how spread out your data is and therefore how rich it is. **Range** is one measure of dispersion. The range gives you the extreme values of variable attributes. For instance, from Table 4.5 we see that the minimum family size is 2 members and the maximum is 29 members, hence the range for household size is 2-29 people.

The most commonly used measure of dispersion is **standard deviation**, usually denoted by SD. In Section 1.2 we defined what standard deviation is and how it affects sample size. “SD is a measure of how much, on average, the scores in a distribution deviate from the mean score” and is calculated using variance, (generally denoted by S^2), which is the average squared deviation from the mean and is calculated as follows (Bernard, 2000:526):

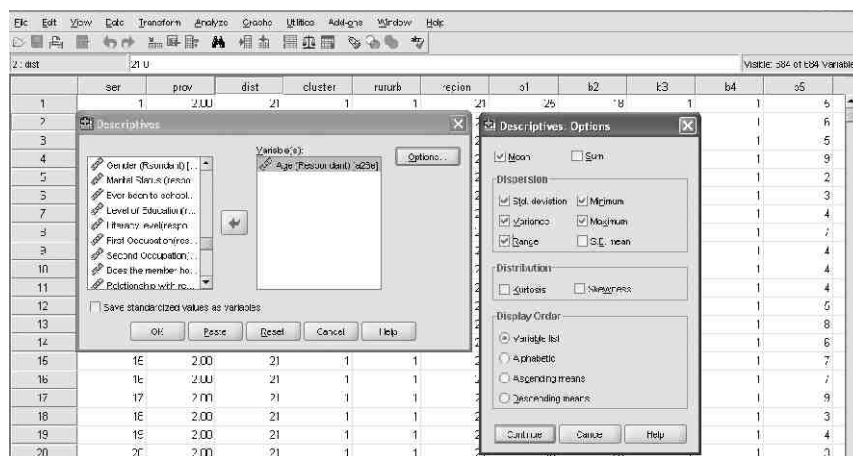
$$S^2 = \frac{\sum (x - \bar{x})^2}{n-1} \text{ (Formula 4)}$$

Where x is the score in a distribution of observations (for instance the age of one respondent, say 25 years), \bar{x} is the mean of the distribution of all scores (say the mean age of the sample is 30 years) and n is the total number of observations/sample size and \sum represents sum (i.e. the sum of differences between the ages of all respondents and the mean of 30 years). Therefore to measure variance we subtract the mean of the set of values for a variable from the value of each individual observation, square it to get rid of negative numbers, add up all the differences and divide by the sample size minus 1. Standard deviation is simply the square root of variance (S^2):

$$SD = \sqrt{\frac{\sum (x - \bar{x})^2}{N-1}} \text{ (Formula 5)}$$

Thankfully, you will not have to use these formulas to calculate variance or SD since your data analysis program will do it for you. Nonetheless it is important to understand how variance and SD are calculated since these basic formulas form the basis of further analysis. To calculate SD or variance in SPSS you will go to ‘Analyze’ then select ‘Descriptive statistics’ followed by ‘Descriptives’, as in Screenshot 9. Once you do so a window will appear asking you for the variables you want described. After selecting the variables (‘Age’ in this example) you will go into ‘Options’ and check the types of statistics you want the system to generate:

Screenshot 9



The resulting table (Table 4.6) gives us the total number of cases (N), the range for ages of respondents (minimum and maximum), the mean, variance and standard deviation:

Table 4.6 Descriptive Statistics (2)

	N	Range	Minimum	Maximum	Mean	Std. Deviation	Variance
Age (Respondent)	699	45	15	60	31.96	6.783	46.005
Valid N (list-wise)	699						

When presenting the analysis, the SD is reported along with the mean. In this case we would say that the average age of respondents was 32 years (SD or $s=6.8$).

4.4 Hypothesis Testing

In the section on sampling we have discussed the importance of our sample being representative of the population from which it is drawn. Recall from the discussion on central limit theorem that *if many samples of sufficient sizes are taken from the same population, the mean and standard deviation of the sample means will usually approximate the true mean and standard deviation of the population*. So how do we determine whether our sample mean represents the population mean? This is where **hypothesis testing** comes in. The steps involved in hypothesis testing are (Bernard, 2000:526):

1. Stating your **null hypothesis**, i.e., there is no difference between the sample mean and the population mean, except for chance, usually denoted by H_0 .
2. Stating your **alternative hypothesis**, i.e., there is a difference between the sample and the population means that is not due to chance, usually denoted by H_1 .
3. Determine whether you are only interested in the magnitude of the difference stated in H_1 or also the direction of the difference between the two means, i.e. whether the sample mean is greater than or smaller than the population mean. If you are not interested in the direction then it is okay to simply state your alternative hypothesis as done in step 2 above. However if you are interested in the direction of the difference then H_1 should be stated as: there is a difference between the sample and the population means. The sample mean is greater than/smaller than the population mean. This will determine whether you will use the **one-tailed** or the **two-tailed test** to check for difference. When the direction of the difference is important then the one-tailed test is used (i.e. you know it's either bigger than or smaller than the true mean and hence you only use one tail/direction).

To understand the concept, recall the bell curve or the normal distribution curve in Figure 1 of Section 1.2. The sample means are distributed across this curve, some larger than the true mean, some smaller and some exactly the same. What we are interested in are the tails of the curve i.e. the extremes, i.e. whether the sample mean is too small or too large compared to the population mean. How do we determine if our sample mean is in the thin area under one of the tails or under the bulge in the curve? For one-tailed tests the cut-off point is 0.05 (i.e. 5%) and for two-tailed tests it is 0.10 (i.e. 10%). These percentages represent significance levels.

4. Determine the **significance level** for the hypothesis test. If the probability of an outcome occurring is less than 5% in a hundred tries then, as a rule, the outcome is considered to be statistically significant. This is known as the p value (or the probability value) and is written as 0.05 level of significance. A p value of 0.01 (i.e. a 1% chance of occurring) is more significant than the 0.05 level.

A significance level of 0.05 in a hypothesis test would be interpreted as follows: if the mean from a sample is likely to occur more than 5% of the time then **we cannot reject the null hypothesis**, i.e. *we cannot reject* the statement that the sample mean is the same as the population mean. On the other hand if the mean from a sample is likely to occur less than 5% of the time (i.e. less than 5% likely to be found in the region under the bulge of the curve) then **we reject the null hypothesis**, i.e. *we reject the statement* that the sample mean and the population mean are the same. In both cases we either reject or fail to reject the null, but we never try to prove the alternative hypothesis.

There are two types of errors that one should be mindful of while running these tests; **Type I error** which is when we reject the null hypothesis when it is actually true and **Type II error** which is when we fail to reject the null hypothesis when it is actually false. To reduce the chances of Type I error we can increase the significance level, say from 0.05 to 0.01 so that there is a greater chance of the mean occurring outside the defined limits. Box 4.2 below illustrates this principle with an example.

Box 4.2 Calculation of Null Hypothesis¹²

Suppose you select a sample of 150 married women between the ages of 15-49 years, from 3 refugee camps in KPK and ask them to report the type of assistance they had at the time of their latest live delivery. Suppose on average 29% (SD= 8.5) of them say their deliveries were attended by doctors. You want to check how close your sample is to entire population of married women of reproductive age in KPK. From a recent nationally representative survey you know that 27.8% women reported being assisted by doctors at the time of their latest delivery. From the same survey, you estimate the size of the target population to be 5 million women.

Supposing your null hypothesis is: Based on our sample the mean (i.e. births attended by doctors) of the population from which the sample is drawn is 27%. Your alternative hypothesis is: Based on our sample the mean (i.e. births attended by doctors) of the population from which the sample is drawn is not 27%. From our sampling discussion we know that standard error is the error made in estimating a population parameter using sample statistics. The standard error of the mean (SEM) is SD divided by the square root of the sample size:

$$SEM = \frac{SD}{\sqrt{n}}$$

In this case the SEM would be:

$$SEM = \frac{8.5}{\sqrt{150}} = 0.694$$

Using what we learned about confidence intervals in Section 1 we calculate the confidence limits at 95% confidence level:

$$29 \pm 1.96 (0.694)$$

$$29 + 1.36 = 30.36$$

$$29 - 1.36 = 27.64$$

We can expect that 95% of all samples of 150 taken from 5 million women of reproductive age between the ages of 15 and 49 years will fall between 27.64% and 30.36%. Since both the sample mean (29%) and the population mean (27.8%) fall within the confidence limits we cannot reject our null hypothesis that our sample mean and population mean are the same, i.e. that the percentage of births attended by doctors in our sample was the same as the percentage in the population.

¹² Formulas taken from: Bernard, H. Russell. *Social Research Methods: Qualitative and Quantitative Approaches* (Thousand Oaks, CA, Sage Publications, 2000), Chapter 14: Univariate Analysis at 532.

4.5 Practice Exercises

1. Scenario 1: You want to study the impact of displacement on refugee women (not IDPs) particularly in terms of gender-based violence and human rights abuses. The aim is to assess whether displacement has made these women more vulnerable to violence and abuses.

Scenario 2: An international aid agency plans to implement a programme to improve the food security and nutritional status of IDPs housed in relief camps in different parts of KPK. Your organization/department has been selected to assist the agency in the design and execution of the programme in your district. As part of the ground work you are required to conduct a situational analysis survey of the existing food security situation in the camps in your district.

For the above scenarios:

- (i) Explain how you would select your sample, what would be your sampling frame, the selected sampling technique, and whether or not you will incorporate any comparison groups in your study (Note: you can use any of the techniques/methods discussed in this Module as long as you justify the use of the technique);
 - (ii) Explain how you will determine sample size; and
 - (iii) Identify some of the sampling and non-sampling errors that could be associated with the studies and explain how you would try to minimize those errors
2. You read the following statement: In a sample of *300 respondents chosen properly from a population of 150,000*, 32% reported that they had at least one family member who had migrated to the neighboring country whereas 68% said that none of their family members had migrated to the neighboring country.

Does this statement give you enough information to conclude whether or not the sample is representative of the population? What additional information will you require to determine whether the data accurately describes the population, especially with regards to the italicized information?

4. For a sample of 500 households, collected from four provinces of the country, the average household size comes to 8.4 members ($SD=3.61$). From the latest *Pakistan Demographic and Health Survey (PDHS 2006-07)* you know that the average national household size is 7.2 members and that the estimated national population is approximately 170 million. Based on this information:
 - (i) Develop your null and alternative hypotheses regarding whether or not the sample represents the population from which it is drawn based on the mean household size;
 - (ii) Determine confidence intervals for this sample for 95% and 99% confidence levels and test your null hypothesis at both levels (i.e. either reject or fail to reject the null).

Further Reading

On Sample Weighting:

Yansaneh, Ibrahim S. *Construction and Use of Sample Weights* (United Nations Secretariat, Statistics Division: 2003). UN Doc. ESA/STAT/AC.93/5. [Accessed online at: http://unstats.un.org/unsd/demographic/meetings/egm/Sampling_1203/docs/no_5.pdf]

On Non-Sampling Errors:

Banda, Jeremiah P. *Nonsampling Errors in Surveys* (United Nations Secretariat, Statistics Division: 2003). UN Doc. ESA/STAT/AC.93/7. [Accessed online at: http://unstats.un.org/unsd/demographic/meetings/egm/Sampling_1203/docs/no_7.pdf]

On Standard Errors/Sampling Errors:

National Institute of Population Studies & Macro International Inc. *Pakistan Demographic and Health Survey 2006-07 Report* (Islamabad/Calverton, Maryland: June 2008) at 197-208. See: Appendix C, Estimates of Sampling Errors. [Accessed online at: <http://www.measuredhs.com/pubs/pdf/FR200/FR200.pdf>]

References

Banda, Jeremiah P. *Nonsampling Errors in Surveys* (United Nations Secretariat, Statistics Division: 2003). UN Doc. ESA/STAT/AC.93/7. [Accessed online on July 29, 2011 at: http://unstats.un.org/unsd/demographic/meetings/egm/Sampling_1203/docs/no_7.pdf]

Bernard, H. Russell. *Social Research Methods: Qualitative and Quantitative Approaches* (Thousand Oaks, CA: Sage Publications, 2000).

DHS Statistics Live, Online Guide to DHS Statistics. [Accessed online on July 29, 2011 at: http://www.measuredhs.com/help/Datasets/index.htm#sampling_weights.htm]

Fowler, J. Floyd. *Survey Research Methods* (Thousand Oaks, CA: Sage Publications, 2009).

Israel, Glen D. *Determining Sample Size*. University of Florida, Institute of Food and Agricultural Sciences. Originally published November 1992, reviewed April 2009. [Accessed online on July 23, 2011 at: <http://edis.ifas.ufl.edu/pdffiles/PD/PD00600.pdf>]

Krejcie, Robert V. & Morgan, Daryle W. "Determining Sample Size for Research Activities," (1970) *Educational and Psychology Measurement* 30 at 607-610.

Trochim, William M. *The Research Methods Knowledge Base*, 2nd Edition. [Accessed online at: <http://www.socialresearchmethods.net/kb/external.php>] (version current as of October 20, 2006). See: Sampling/External Validity.

Yansaneh, Ibrahim S. *Construction and Use of Sample Weights* (United Nations Secretariat, Statistics Division: 2003). UN Doc. ESA/STAT/AC.93/5. [Accessed online on July 23, 2011 at: http://unstats.un.org/unsd/demographic/meetings/egm/Sampling_1203/docs/no_5.pdf]



Module

5

Qualitative Methods of Data Collection

Module 5 - Qualitative Methods of Data Collection

Purpose: To achieve an introductory knowledge of qualitative research methods and their usage.

Qualitative research broadly refers to any kind of research that does not rely on quantitative methods, such as statistical procedures or any other means of quantification. (Strauss and Corbin, 1990: 17) Qualitative research seeks illumination and understanding of situations, while quantitative research seeks to determine the cause of an event or occurrences and predict future ones. (Hoepf, 1997)

In contrast to the setting up of a scientific experiment, qualitative research gains its findings from observing people in their natural or real-world settings with as little interferences from the researcher as possible. It consists of a three part process of collecting data, writing notes, and engaging in ongoing analysis. This is in addition to the preparatory work of doing background reading and the final written analysis at the end of the research. (Ahmed, 2006)

Critics sometimes argue that its methods are “soft”, lacking in rigour, as compared to quantitative research. For this reason field notes and observations, along with any more documentation about what takes place in the field, must be carefully maintained and are essential to the success of the methods used. Researchers argue that it is not helpful to dwell on a dichotomy between qualitative and quantitative research, because any research process takes place on many levels and can combine methods that emerge from both scientific and social sciences. (Hammersley, 1992)

Qualitative research is commonly used in one of these ways:

- i) To inform the design of a quantitative survey;
- ii) To examine in further depth the outcomes of a quantitative survey;
- iii) To do a thematic and holistic analysis of a research topic.

The reason you may select to do a qualitative research study depends on factors such as these¹³:

- i) The nature of the research question: if you are not trying to establish causality or a relationship between variables, and instead are posing your question in terms ‘how’ or ‘what’.
- ii) The topic at hand needs to be explored. There are no obvious variables and no simple theories to explain the behavior of your participants.
- iii) You need to present a detailed view of the topic. One perspective will not suffice, you need to have multiple perspectives and sources of information.
- iv) You want to study individuals in their natural setting. This means going out to the field, accessing your participants and gathering your material.
- v) You wish to write in a literary style, acknowledging yourself as the writer, and possibly presenting your findings in a storytelling form.
- vi) You had sufficient time and resources to spend in the field, conduct extensive data collection, and detailed data analysis of your textual information.

¹³ <http://www.computing.dcu.ie/~hruskin/RM2.htm> this comes from Cresswell (1994)

- vii) Audiences are receptive to qualitative research, e.g. publishers, academics, readers who respond to qualitative approaches.
- viii) You wish to share your findings through presenting the participants' views, thereby acknowledging their subjectivity and yourself as an active learner, rather than as a disengaged expert wishing to pass judgment on them.

Qualitative research draws on methods, approaches and techniques gleaned from other disciplines, such as anthropology, sociology, history, and cultural studies. The process of going into a community, and learning about its members through the use of methods such as conducting interviews, establishing rapport, observations, compiling a journal, and so on, is the work of an ethnographer, i.e. one whose work is the scientific description of human cultures¹⁴. But beyond this, the job of the researcher is to ascribe *meaning* to human behavior as studied. He emerges with “a rich and extensive set of details concerning methodology and context provided in a research report”, which is known as a “*thick description*”¹⁵.

The notion of **reflexivity** is essential to qualitative methods in use today. It refers to an acknowledgement by the researchers of the central role they play in the research process. Stories, i.e. interviews, are told to particular people for particular reasons, and thus the researcher him/herself is part of the research context as well as its limitations. This need not undermine the quality of the research itself, since it is not based upon the standard of objectivity found in the natural sciences.

This is of particular concern when representatives of various organizations, including service-providers, enter communities of refugee/displaced people in order to conduct research. Those from whom information is being solicited are extremely vulnerable -- physically, psychologically, economically and also politically. Those conducting the interviews are in a unique position of power, since the information they gather will have some clear link with policies and programmes intended to benefit the interviewees. If you are an interviewer in such a situation, a high degree of reflexivity needs to be maintained in order to interpret your data in the light of this power differential or imbalance. You may be of a different class, language, religion, colour, ethnicity or cultural background than the person whom you are interviewing, and this will all come to bear on the content of the interview.

5.1 Research Methods

The methods of qualitative research discussed below draw on the work of anthropology. However, in traditional anthropological research the time spent understanding and living in a community can extend for months and even a year, and you will conduct your research over a period of days or weeks only. This does not allow for a strong rapport to be developed with your respondents. Nonetheless these methods have been adapted very successfully to the goals of the development sector, as you will see.

While methods of qualitative inquiry will vary, there are some rules that remain constant. The researcher needs to be respectful at all times, ask open-ended questions, avoid passing any judgment on what people say, allow them to complete their train of thought, avoid short and sharp questions that people feel obliged to answer, give people a chance to ask their own questions about the process, and use appropriate eye and body language to put people at ease.

5.1.1 Participatory Research Methods

Participatory research (PR) has two objectives. “The first is to produce knowledge and action directly useful to a group of people.” (Reason, 2001: 1) The second aim is to empower people, i.e. help them to use their own sources of knowledge in order to see how they have been deprived of knowledge and

¹⁴ <http://www.answers.com/topic/ethnography>

¹⁵ writing.colostate.edu/guides/research/glossary/

power under existing social and political circumstances, with a view to resisting their own oppression and marginalization. This type of activist social science research was developed in the 1970s and 1980s, mainly by researchers from third world countries who sought to empower the poor. (Cancian, 1993: 93)

The emphasis on this “bottoms up” approach to knowledge generation, i.e. with a focus on locally defined priorities and local perspectives has begun to be used by many NGOs and international agencies, in response to ineffective policy and programme recommendations that have occurred in developing countries in the past. (Cornwall and Jewkes, 1995) This makes PR a somewhat controversial research strategy, because all knowledge that is generated is consciously located within the subjectivity of the participants and with an awareness of the context of power relations in which they live.

The role of the researcher in PR is to empathize with the needs of the community, with its members arrive at an articulation of the problem that needs to be resolved, and together with them arrive at possible solutions. In effect, the researcher is a catalyst for people helping themselves. The divide between knowing and doing is bridged through this approach, and the people become active agents in the generation of knowledge and its applicability. (Park, 1993: 3)

Most participatory methods use visualization and group work, engage in a process of open-ended enquiry that focuses on a set of challenges, and arrive at potential solutions to these problems. All of them require an empathic researcher, who knows both how to engage in active listening to the community and also how to engage in a real dialogue with the community so as to arrive at knowledge-based solutions to the problem at hand.

The most engaged type of participatory research is **cooperative inquiry**, “in which all those involved in the research endeavor are both co-researchers, whose thinking and decision-making contributes to generating ideas, designing and managing the project, and drawing conclusions from the experience; and they are also co-subjects, participating in the activity which is being researched.” (Reason, 1998: 2) In simple terms, this means research that is done *with* people (working together in a group), rather than *on* people (involving an outside researcher studying a group). It is not a commonly used method of qualitative research in Pakistan, but it has enormous potential to generate new kinds of knowledge beyond what our traditional methods of inquiry have been able to do thus far in social science research. The kinds of research tools that would be employed in cooperative inquiry include extended episodes of group discussions to determine the research agenda and explore the issues among those conducting the inquiry. For example, a group of community-based health workers in a displaced persons camp in northern Pakistan discuss their common interests in a group format and then emerge with a clear sense of which research questions (leading to improved outcomes) they would like to explore. They would then proceed to test their questions in their own spheres of work and discover through experimentation what works.

PR methods include *visualizations*, which provide opportunities for local people to illustrate in their own terms how they conceptualize a given subject. For example, researchers may ask a group of women to draw on a chart paper a time line, and then mark out what proportion of a day they spend on specific types of activities. Seasonal calendars may be used to illustrate the connections between time of year, types of diseases that occur in particular seasons, seasonal migration patterns, and food availability at different times of the year. A group of men or women may be asked to draw a human figure and then label its parts, which allows them an opportunity to illustrate their conceptualization of their bodies which may significantly differ from what the researchers had in mind.

The notion of “community”, as a distinct, homogenous and integrated group of people, must be critiqued here. For example when outside researchers coming from Islamabad engage with local people in KP, they might have conceptualized a community as a group of displaced people from one locality in upper Dir who are now living in a camp outside Peshawar as a community. Upon closer

scrutiny, though, they are in fact a heterogeneous group of people of different class and caste backgrounds, wealth, age, education levels, and so on. This need not deter them from the research, as long as they recognize and acknowledge the existing power structures within the group and the competing needs and priorities of the people. Here the 'agency' of the researcher becomes important, who consciously chooses who to work with and how to employ participatory methods to ensure that a full range of voices, including the most marginalized, will be heard.

5.1.2 Oral History

Oral history is a specialized method of qualitative research that focuses on interviewing specific individuals about their lives for the purpose of creating a historical narrative. This is an extremely creative method, it is an exercise that connects life and times, a personal history with a public history. Events from a person's life serve to convey something larger or definitive for the narrator, these events as told serve to create an archive of memory for the narrator and a testament of history for the researcher. (Portelli, 1997)

Oral histories can serve as a complement to quantitative surveys. For example Miller (2007) writes about his use of them to augment data that emerged from a panel survey (a quantitative survey conducted with the same group of respondents over more than one time period). The histories do so by providing a holistic view of a family, and extending the time dimension of the panel study by asking respondents about events before the panel survey and about events beyond it into the future.

The most common narrative that emerges from oral history is that of war, during which the respondent has played a role or been directly affected. For example, an oral history research study has collected fifty oral histories of Afghan women refugees who fled to Pakistan during the course of violent conflict in their country. Their life stories unfold in the context of the larger political and social upheavals of their times, allowing for the interviewer to gain deeper understanding into the connection between the personal and historical context which might not be possible through any other method. The fifty testimonies also serve as a historical record of what took place, what was witnessed, and what were women's perspectives.

The creativity in this method arises out of the dynamic between the interviewer and narrator, which influences what will get told and what may get left out. The personal relationship that develops between the interviewer and respondent, given the fact that the oral history is often taken over a course of many interview sessions, also has a great deal of influence over what is selected for narration. Oral history refers to "what the source and the historian do together at the moment of their encounter in the interview." (Portelli, 1997:3) When a person is asked to provide a narrative of his life, and it emerges in the form of an oral history, it is likely to be the first time such a narrative has ever been formed. In fact, you may find that the respondent/narrator is deeply moved by having her story listened to so attentively and giving her a chance to create some coherence to the story of her life.

Portelli (1997: 12) has noted that the less the interviewer reveals about himself and his views, the more likely that the respondents will reveal superficial layers of their lives and material that is in the public domain. More personal knowledge and experience is more likely to emerge from a more critical or challenging interview. Hence the role of the interviewer is so valuable in shaping the kind of narrative that emerges.

Oral histories are best recorded, providing the respondent agrees to this. The questions do not need to follow a chronological order with respect to the person's life. Some level of free association can remain in place, allowing the respondent to meander through what he/she thinks are important life events. Interviews can take place over a course of days, weeks or even years, depending on the purpose of your study.

5.1.3 Visual Documentation

It is entirely valid for qualitative research to employ a variety of methods based on analysis of visual material, not necessarily text or verbal material. For example, a study could look at the portrayal of women's bodies in the popular media, and for that purpose it may use photographs, videos, and films in order to arrive at a coherent analysis. Further, textual and other information can be used to verify the findings using multiple methods.

5.2 Qualitative Research Tools

5.2.1 Community Profiling

A detailed social, economic and political profile of each locality selected for fieldwork, prepared on the basis of interviews, observations, and background information on the sites. A sample data collection sheet is in Box 5.1. The collection of community data allows for basic information about the researched communities to be compiled in one format. It must include, in addition to services available, roads, access to markets, also information about caste and class composition and political affiliations or activities in the community. This will help in interpreting interview material, as well as serving as a guide.

5.2.2 Mapping

An introductory exercise conducted by at least two field researchers with a small group of randomly selected community members. The purpose is usually to map out with marker and paper the salient details of the locality, its meeting places, areas of differing activity, and other information pertinent to the research goals. [This is known as "social mapping."] The exercise also helps the researchers to break the ice with community members and begin to select potential interview respondents. [See Participatory Action Research for more discussion.] Perhaps even more important, it is generated by community members themselves and provides a framework for discussion, raising awareness and debate over issues, and helping the researchers to identify leaders who may be participants in focus group discussions. (Jones, 1996)

Other types of social mapping include *kinship maps* that identify caste, tribe, and kinship ties that enable genealogies to be understood. For example, in a community of migrants, a researcher can gain an excellent understanding of the ties within the community by sitting down with elders and through discussion ask them to map out how they are related to one another. *Skills maps* identify those community members who have skills that are relevant to the research. If a community of new migrants needs to be helped in establishing income-generating projects, group interviews with people of various ages in which they are asked to illustrate their skills on a chart paper can be very helpful indeed. *Education maps* illustrate literacy levels and numbers of children in households attending schools; these are particularly useful if researchers are attempting to assess the needs of migrant children. *Health maps* can do the same for the physical health of household members, and identify a wide range of chronic and infectious diseases, or addictions.

Topical resource maps can examine specific aspects on the social maps, or explore issues such as access to land. This could include agriculture, water, livestock, forestry, land, and these maps can also identify boundaries of control and disputes over these resources. (Jones, 1996)

Futures maps can illustrate the changes people would like to see and participants can build upon this map to identify the steps that need to be taken in order to achieve these goals.

Mental maps show what is most important to people and provide insight to their perceptions. For example, they can be asked to draw areas where they feel safe and unsafe at different times of day, and the maps can be used to compare perceptions among different interest groups.

Mobility maps are enormously helpful in determining gender and class differences in mobility outside the home. Women are invited to indicate on a paper or other surface which places they are able to go to and the frequency of such outings. This can include places within and outside their residential areas.

Mapping is usually done with a group of 8-10 participants, commonly divided along lines of gender and age, but there is always room for adjustment depending on the circumstances. The discussion generated by mixed groups can also become engaging and lively, although the exclusion of less dominant voices in the discussion needs to be avoided. The use of paper and markers/pencils for drawing works well, especially if two copies of the map are made and one remains with the community members. Flipcharts, stickers of different shapes and colours, and other items are useful as well. (Jones, 1996)

Information from the maps needs to be “triangulated” i.e. cross-checked through holding a discussion around it, sharing it with other groups in the community to invite their perspectives, and comparing maps across groups.

We should be mindful that in our research on refugees and displaced people, there may be concerns about sharing of information regarding women, assets, personal/tribal/political disputes. Mapping should only take place when these risks are minimal.

5.2.3 Interviews

Interviews are a means of communication that seeks to generate information from individuals and groups. While there are many different styles and types of interviews, all of them need to conform to the norms mentioned in Section 1. Qualitative interviewing is useful to access an individual’s attitudes and beliefs, and explore issues with depth and nuance that are not possible through quantitative research methods. The interview is useful to the researcher not only for what it says, but also for what is left out.

i. Focus Group Discussions: A formal scheduled recorded discussion with a group of 8-10 community members, purposefully selected on the basis of certain homogenous characteristics necessary for the research goals, e.g. sex, age, ethnicity, length of time in Pakistan.¹⁶ An FGD may take place at any time during the research process. It is useful to obtain general background information on a research topic, as well as to generate a new hypothesis and ideas. Finally, it may also be used as a means to explore further the findings of quantitative research (Stewart and Shamdasani, 1990: 15).

Focus group members should be selected with careful attention to the research purpose. There should be enough homogeneity within the group that variation does not disrupt the group dynamics and level of the discussion. This is particularly difficult in Pakistan, where arranging a group with married women ages 20-35, for example, may end up turning into a group of married women along with some of their children and mother-in laws! Researchers will need to politely invite those who have intruded upon the group to an informal discussion separately, in order to allow the FGD to take place as planned.

The interview guide should reflect the research purpose, but it should also be flexible to respond to new material that may emerge from the discussion. Questions need to be open-ended and free of judgment. One researcher moderates the discussion and another may take notes. A recording of the discussion, if permitted by the participants, is extremely helpful to ensure accuracy of interpretation.

The moderator should begin by introducing himself/herself, and then invite the note-taker and each group member to do the same. Next the moderator should go over the FGD rules, i.e. that every group member should have a chance to speak, there should be no interruptions, and the moderator should facilitate the discussion based on certain topics to be covered. At this point permission to record the discussion can be requested.

¹⁶ There are other group methods which are not discussed in this Manual, e.g. see the reference to Stewart and Shamdasani (1990) at the end of this Module.

The moderator will then introduce the subject of the research, and explain how it is relevant to the participants and how the research findings may help to arrive at solutions to the problem at hand. There may be feedback from group members. The moderator can then begin the discussion and invite the views of all participants, however, this is not to be done on a turn-by-turn basis, but in a conversational way that encourages all to participate.

The ability of the moderator to create a comfortable atmosphere conducive to the sharing of views in a non-threatening atmosphere will determine the quality of the FGD. (Stewart and Shamdasani, 1990: 100)

The discussion can last 60-90 minutes, although FGDs of lesser duration may take place depending on the time available to the participants. Remember that the moderator is not conducting an interview of individual participants regarding their own personal experiences alone, but rather trying to gauge from the discussion what most of the participants feel about attitudes, norms and practices.

ii. Key Informant Interviews: This interview is with a member of the community or locality that has extensive knowledge of the research community and can help to identify new issues for researchers to follow up. It is a discussion over one or more sittings, based on interview guidelines developed for other interviews, but generalized to encourage information about the community as a whole instead of specific individuals. One researcher conducts the interview and takes notes. A recording of the interview, if permitted, is very useful.

iii. Informal Group Discussion: This tool is an adaptation of the above-mentioned FGD, but has been found to be a flexible and useful alternative in a complex field situation. It is also usually scheduled in advance, and brings together up to five respondents and one to three field researchers for approximately half an hour. The interview guidelines form the basis of the discussion and notes are taken.

iv. In-Depth Interview: This is a scheduled one on one interview of approximately one hour. A second field researcher may be present to assist in note-taking. The questions are open-ended. Audio recording of the interview is ideal. A complete in-depth interview can involve multiple meetings with the respondent, particularly if an oral history or case study is being compiled.

v. Short Interview: This scheduled interview follows the same format as the above detailed interview but is on average half an hour long. It may involve a couple of meetings, but does not serve the same analytic purpose as an in-depth interview. It can help in the triangulation of data, however, and also yield information that can be further explored in other interview formats.

vi. Interaction: This unscheduled short interview is less than twenty minutes in duration, conducted by one field researcher who takes notes. The formal documentation of chance encounters in the field forms a vital part of the research data. The term “interaction” arose specifically out of fieldwork conducted in Pakistan, in response to the large number of random conversations that are constantly taking place in a community setting between researchers and residents. If these interactions are formalized by being noted as part of the research data, they can be used better in the analysis to follow.

5.2.4 Participant Observation

Simple observation in the setting of your research project can elicit very valuable insights. For example, if you enter a particular environment, such as a newly established camp for displaced persons, there is an enormous amount of new information that you will have to absorb. Rushing through qualitative research is always a mistake, so it is best to take your time and observe what is going on first. For example, you may sit down at a small *chai khana* that has sprung up in the camp, and simply observe the comings and goings of people and listen to the snippets of conversation. If you do so without taking notes it is unobtrusive and people are unlikely to mind. Afterwards, record these observations in your field diary, they will prove very useful to interpretation of your interview material at the analysis stage of your research.

Another example of participant observation could take place when you, the researcher, simply accompany a respondent on an outing or activity. It could be walking alongside a woman in a camp as she goes out in search of water, or observing a teacher in a makeshift school in the camp during one day of school, and so on. The notes based on your observation work as a parallel to your interview material, and they may provide you with the information needed to modify your methodology, methods and analytic perspective during the research process.

5.2.5 Case Studies

Case studies can be used to illustrate findings or in comparison with one another to draw out finer points. They can also serve a useful purpose during the analysis stage of research, when illustrative uses will come in handy. A case study could be a profile of a migrant/displaced family, for example, whose story illustrates many of the themes that you have seen emerge through your other interview material. A case study could also be the story of a particular group's experience of displacement, used to illustrate the complexity of the obstacles in the way of building a livelihood. A detailed household case study can also serve as a good way to convey the complex livelihood strategy of a household that depends on a broader family network for its livelihood. An organizational case study may also be useful – e.g. one that highlights the work of a specific NGO in the refugee camp environment whose work addresses some of the issues raised in your research.

5.2.6 Field Notes

Writing notes is the core link between data collection and analysis. Field notes include observations, descriptions of 'interactions', personal notes to oneself, interview and oral history notes (with emphasis on contexts, observations, emphases), participant observation notes, maps from mapping exercises and timelines. It may be helpful to organize field notes thematically or chronologically. Oral notes can be recorded, as can group discussions about findings, if there is an arrangement for transcription. (Ahmed: 2006) Field notes can describe the context, circumstances, tones, silences, of the research material and your reflections on it to be used in the later analysis.

5.2.7 Visual and/or Document Collection

There is usually a set of written or visual material that is useful to informing the qualitative line of inquiry. For example, if you are studying the impact of conflict on women's lives in a certain geographical zone, such as Malakand, then it would follow that in order to enhance your contextual understanding of their lives you would turn to newspaper, magazine, NGO, or relevant government reports for further details and background. Nowadays in Pakistan the internet and television have become widespread and they are leaders in relaying information to the public, often before the government or newspapers get a chance to do so. It is useful and sometimes essential to refer to these sources as well.

For example, it became a huge issue among the public when a video was posted on YouTube and television a few years ago showing a young woman getting flogged in Swat by militants, for allegedly being sexually involved with a man who was not her husband. While the truth of this video could not be established, its effect corroborates an environment of fear that women in Swat reported during this time. A connection such as this you may wish to explore further in your interviews, and it may become an important part of your analysis later.

5.3 Getting Started

The qualitative research process needs to be adapted further to suit different settings, purposes and methodologies. In the context of conducting research in Pakistan, and that too among people who have been displaced from their homes due to conflict, disaster or persecution, here are some guidelines that may be useful.

i) Your research approach will depend upon your *questions and plans* for seeking answers. First you will need to become familiar with the research that already engages with these questions and identify where there are gaps that you may be able to fill through your research. Your selection of research sites will also depend upon what you learn through this background work.

ii) *Explain the purpose of the research* and the philosophy of learning “from experience and for action” to the people that are involved, which include primarily the research participants. When you are in the field, explain what the use of the knowledge will be, and any feedback mechanisms in place to further share and debate emerging findings.

iii) Begin the process of *participatory mapping* (see above) and work with a range of different interest groups. Getting started with the community profile and key informant interviews is also a good beginning.

iv) *Locate interview respondents*: The technique you use will depend on how qualitative methodology fits into your larger research goals. For example, if you are conducting a small qualitative research study to explore the findings of a quantitative survey, then you may wish to target those households from the survey that are of special interest to you. Alternately, you can go into the field without a plan for locating respondents, but assuming that your group mapping exercises will orient you to the community and allow you to identify interview respondents based on these early interactions. You could then proceed from one respondent to the next, having the latter identified by the former. This is known as snowball sampling technique.

Yet another option is to work with a community organizer or similar type of individual who has personal knowledge of the residents in the community, and can help you to identify households and respondents that are relevant for your research. Using a go-between helps to establish trust between yourself and the respondent when you do not have too much time to let this develop naturally. As a rule you must document the process and explain the reasons for having adopted a given approach when you write up your analysis. Using this approach, it is useful to plan on a target number of interviews you would like.

v) *Obtain informed consent* before conducting any formal interviews is essential. This may be done in a number of ways, depending on how literate the respondents are and the purpose of your research. [See Module 2 for details.] Respondents must know that they can opt out of participating in the research at any time, and that rules of confidentiality will be adhered to. When working with vulnerable communities in Pakistan, which include migrant and displaced populations, this may seem to be an unnecessary formality since respondents may not express any interest in issues of confidentiality at the time of being interviewed. Yet this is an important step to take in the process of increasing the professionalism of social science research in Pakistan, and to ensure our work meets international standards.

vi) *Confidentiality*: The researchers must assure respondents that their identities will remain confidential and so will the content of their interviews. Choosing the right location for the interview is an important part of maintaining the respondent’s confidence in the interview process. Ideally you would like to conduct the interview in a relaxed space where the interview can proceed undisturbed. Such a place is hard to find in a camp environment. Interviews conducted within the home, or on the premises of offices, i.e. on more neutral ground, bring their own set of possible disturbances and these need to be carefully considered. In the home setting other family members usually try to join in an interview. In a public setting, such as a tea shop or a school, other community members may wish to listen in to the interesting discussion. It is highly desirable to maintain privacy during an interview, but very difficult to ensure!

vii) *Problems* will arise in the collection of data for various reasons. Most common is the need for a translator to conduct the interviews, because field researchers may not be versed in the language that is most comfortable for the respondent to speak in. For example, Pakistani researchers in Afghan

refugee camps are unlikely to speak Dari. One possible solution is for the researcher to conduct the interview and take his/her own notes, but allow the conversation to be conveyed through an interpreter. Another option is to hire a Dari-speaker to conduct the interview and take notes along the guidelines the researcher has specified, which will subsequently get translated into the language being used for analysis by the researcher. In either case, the final “text” of the interview will have been mediated through more than one person, who is not the main researcher conducting the analysis. In most research studies, the data will also need to be translated, or summarized, in English – which may involve yet another person.

5.4 Collecting the Data

It is very important to rigorously document interview material, observations, field notes, and other impressions you have about the time spent in the field conducting your research. Qualitative research is often labeled “soft” as compared with quantitative research because it lacks the accuracy of numbers and scientific sampling associated with the latter. However, once you begin your fieldwork you will find that collecting and organizing your data, let alone analyzing it, is a more complex task than filling in survey questionnaires. Here are some examples of how to document some of your basic findings.

Box 5.1 is an example of a **community profile**, which is an essential component of your data because it provides an accurate context of the conditions in which your research is taking place. If you analyze your interviews across more than one community, the profile becomes key to comparing one community with the other. While you may collect more information than is written in Box 5.1, this should be the minimal amount that will provide broad contextual information about each community. The data is also useful for *triangulation* during your fieldwork and later analysis. Does the information that you have gathered for the community profile complement what your key informants are telling you? Are your in-depth respondents telling you about festering conflicts in the community that no one told you about otherwise? Try to make your community profile as accurate as possible based on your interview material and field observations.

Box 5.2 is an example of a **respondent profile**, or *basic information sheet* for all individuals with whom you conduct a formal interview, e.g. key informant, in-depth, or short interview. You may gather the relevant information in the form of field notes during the course of the interview and then fill in the profile sheet afterwards. These sheets allow for easy review in the field about the range and type of interviews conducted, and help to ensure you are focusing on the best sources for the kind of information you are trying to gather.

Box 5.3 is an example of **interview guide** that you will use in your discussions with all respondents, whether they are in-depth, or informal, or key informants. Even the focus group discussions, if you plan on having some, should cover the topics in these guidelines. In an ideal situation, you would not need to have these guidelines written in order to conduct your interviews because you will be so familiar with the purpose of the research that you will be able to carry out these interviews without any further support. But in Pakistan qualitative fieldwork is still relatively new, and there will be field researchers who are conducting this type of research for the first time. Written guidelines provide vital support to maintaining focus in what is often a complex and fluid context in the field. Nonetheless, the interviewer must remain open to digressions and additional questions arising out of the conversation, because the strength of qualitative research lies in its flexibility and openness to unexpected findings.

Observations and field notes should be maintained throughout the duration of the field research to complement all the interview material. In qualitative work, it may be the information that you are gathering alongside your interviews that provides depth to your analysis. For example, you may be conducting a focus group discussion with married women on female education in refugee camps. You observe that although they are participating in the discussion, they are reluctant to talk about their

experiences in any detail and you sense that there is more information they are reluctant to share. This observation is essential to analysis of the FGD later, and it also indicates you will need to find another way to facilitate an open conversation on the topic with these women.

Another example is the observation of mood and body language during an interview. You may conduct an in-depth interview with a woman who has lost her husband in the current fighting in Afghanistan, and is left trying to support her children on her own in Peshawar. The interview is long and interesting, but you note the woman appears severely depressed and possibly in need of some professional help. You can record this mood in your field notes, so that it is included in your data analysis at a later stage.

Field notes and observations are the best way to record your own subjective experiences of the research process and any role you may have played in affecting your findings. In qualitative research, especially of the participatory kind, reflexivity is an important component of the process and can help to achieve your research goals. The level of reflexivity that you aim for depends on how participatory your research is – as you have read earlier in this chapter, the more collaborative and collegiate the study, the greater your role in shaping its outcome. (Cornwall, 1995: 1669) Field notes also record the details of your informal interactions, short interviews and key informant interviews, which can be compiled and organized after your return from the field.

Box 5.1 Community Profile

Date: _____ Researcher: _____

Location of Community: _____

Population: _____

Ethnic/Linguistic Distribution: _____

Social Hierarchy: _____ [e.g. Class, Caste]

Migration History of Community: _____

Income-earning/generating Activities: _____

Social, Political, or Other Conflict Areas Among Population: _____

Types of Shelters/Houses: _____

Access to Services:

Educational Facilities: _____

Health Services: _____

Market: _____

Access to Utilities:

Water: _____ Gas: _____ Flush Latrines: _____ Electricity: _____

Additional Information: _____

Box 5.2 Respondent Profile

Date: _____

Interviewer Name: _____

Type of Interview: _____

Community Location: _____

Respondent Name: _____

Age: _____

Ethnicity: _____

Religion (Sect): _____

Language: _____

Place of Origin: _____

Education: _____

Job + Income: _____

Household member: _____

No. of years In Pakistan: _____

Route of entry into Pakistan: _____

Box 5.3 Interview Guidelines for Study on Afghan Refugee Livelihood Strategies (by CSSR/AREU)

We are mainly interested in looking at the transition from refugee to migrant status.

- 1) Refugee history
 - a. Flight, destination, support system, arrival in Peshawar.
 - b. Peshawar as a transit point: in relation to Gulf, Iran and the west
 - c. Refugee camps, decrease in national and international support, movement from camps into city settlements
 - d. Education and Health facilities in Camps
 - e. Lack of citizenship
 - f. Social, political networks and social influences on local culture
- 2) Work history and livelihood
 - a. Occupation/employment status: labour market segmentation, role of kinship groups in acquiring employment.
 - b. Relations with Pakistani co-workers, business sector within Peshawar and other big cities such as Karachi.
 - c. Housing arrangements
- 3) Links with Afghanistan
 - a. Family left behind
 - b. Family returned after having migrated to Pakistan
 - c. *Hawala/hundi*, financial arrangements
 - d. Issues within community and with locals
 - e. Perception of Afghanistan
 - f. Questions of identity with reference to children

4) Return Migration

- a. How viable (economically) is it to go back
- b. How do they view Pakistani migration to Afghanistan: Punjabis in Kandahar
- c. Plans for return
- d. Fears and threats from the view point of women
- e. Migration due to marriage
- f. What are the benefits of returning
- g. What are the expected employment opportunities
- h. Personal conflicts hindrance to return

5.5 Practice Exercises

- i. Write a research question and explain why the answers should be found through qualitative, rather than quantitative, methods.
- ii. Prepare a set of initial guidelines that you will use to conduct interviews in the field.
- iii. If you approach a community for the first time, with the above research question in mind, what qualitative methods would you use to begin your fieldwork? How would you gain access to respondents?
- iv. How would you win the confidence of a respondent despite the obvious power imbalance between yourself and the respondent?
- v. Think about locations for interviews and focus groups conducive to maintaining confidentiality of the interview and the privacy of the respondent.
- vi. Conduct a focus group discussion: once as moderator and once as note-taker.
- vii. Conduct an oral history interview.
- viii. Conduct an in-depth interview.
- ix. Conduct a key informant interview.

References and Further Readings

- Ahmed, Tania. *Proceedings of Workshop on Qualitative Research Methodology* (Karachi: Collective for Social Science Research, 2006). (unpublished)
- Angen, MJ. "Evaluating Interpretive Inquiry: Reviewing the Validity Debate and Opening the Dialogue," (2000) *Qualitative Health Research* 10(3) at 378-395.
- Cancian, Francesca M. "Conflicts Between Activist Research and Academic Success: Participatory Research and Alternative Strategies," (Spring, 1993) *The American Sociologist* 24(1) at 92-106.
- Chambers, R. *Rural Appraisal: Rapid, Relaxed and Participatory*. Discussion Paper No. 3 (Brighton: Institute for Development Studies, 1992).
- Cohen D, Crabtree B. "Qualitative Research Guidelines Project" (July, 2006). [Accessed online on September 5, 2011 at: <http://www.qualres.org>]
- Cornwall, A and R Jewkes. "What is Participatory Research?" (1995) *Social Science and Medicine* 41(12) at 1667-1676
- Creswell, J.W. *Qualitative Inquiry and Research Design Choosing Among Five Traditions*. (Thousand Oaks, CA: Sage Publications, 1998).
- Denzin, N.K. & Lincoln, Y.S. "Introduction: Entering the Field of Qualitative Research," in Denzin, N.K. & Lincoln, Y.S. (eds.) *Strategies of Qualitative Inquiry*. (Thousand Oaks, London and New Delhi: Sage Publications, 2003).
- Denzin, N.K. *Sociological Methods* (New York: McGraw-Hill, 1978).
- Freire, Paulo. *Pedagogy of the Oppressed* (New York: Continuum, 2000).
- Goetz, Anne Marie. "Feminism and the Claim to Know: Contradictions in Feminist Approaches to Women in Development" in Gran, Rebecca & Newland, Kathleen (eds.), *Gender and International Relations* (Celtic Court, Buckingham: Open University Press, 2000).
- Hammersley, M. What's Wrong with Ethnography? *Methodological Explorations* (London and New York: Routledge, 1992).
- Hoepf, Marie C. "Choosing Qualitative Research: A Primer for Technology Education Researchers," (1997) *Journal of Technology Education* 9 (1) at 47-63.
- Holliday, A.R. *Doing and Writing Qualitative Research* (London: Sage Publications, 2002).
- Jick, Todd D. "Mixing Qualitative and Quantitative Methods: Triangulation in Action," (1979) *Administrative Science Quarterly* 24 at 602-11.
- Jones, Carolyn. "Mapping: A Participatory Appraisal "Methods" Paper" (Brighton: Institute for Development Studies, University of Sussex, 1996). [unpublished]
- Lee, R.M. *Doing Research on Sensitive Topics* (London: Sage Publications, 1993).
- Lewins, Ann and Christina Silver. *Choosing a CAQDAS Package*. 6th Edition. CAQDAS QUIC Working Papers. (Surrey, UK: Qualitative Innovations in CAQDAS, April, 2009). [Accessed online at: <http://www.surrey.ac.uk/sociology/research/researchcentres/caqdas/files/2009ChoosingaCAQDASPackage.pdf>]
- Lincoln, Y.S. & Guba, E.G. *Naturalistic Inquiry* (Newbury Park, CA: Sage Publications, 1985).
- Marcus, George. *Ethnography Through Thick & Thin* (Princeton: Princeton University, 1998).
- Mays, N. & Pope, C. "Qualitative Research in Health Care: Assessing Quality in Qualitative Research,"

(2000) *British Medical Journal* 320(7226) at 50-52.

Miller, Robert. *Using Family Histories to Understand the Intergenerational Transmission of Chronic Poverty*. CPRS Research Centre Working Papers (Manchester, UK: Chronic Poverty Research Centre, 2007).

Morgan, David (ed.). *Successful Focus Groups: Advancing the State of the Art* (London: Sage Publications, 1993).

Morgan, David L. *Successful Focus Groups* (London: Sage Publications, 1993).

Park, Peter. "What is Participatory Research? A Theoretical and Methodological Perspective," in Park, P., Brydon-Miller, M., Hall, Budd, & Jackson, Ted (eds.) *Voices of Change: Participatory Research in the United States and Canada* (Toronto, OISE Press, 1993).

Patton, M.Q. "Enhancing the Quality and Credibility of Qualitative Analysis," (1999) *HSR: Health Services Research* 34(5) Part II at 1189-1208.

Patton, M.Q. *Qualitative Evaluation and Research Methods* (2nd Edition) (Thousand Oaks, CA: Sage Publications, 2001).

Pederson, Duncan. "Qualitative and Quantitative: Two Styles of Viewing the World of Two Categories of Reality?" in Scrimshaw, N. & Gleason, G. (eds.). *RAP: Rapid Assessment Procedure* (Boston: International Nutrition Foundation for Developing Countries, 1992) at 39-49.

Plummer, Ken. *Documents of Life: An Introduction to the Problems and Literature of a Humanistic Method*. (London: George Allen & Unwin, 1993).

Portelli, Alessandro. *The Battle of Valle Giulia: Oral History and the Art of Dialogue* (University of Wisconsin Press, 1997).

Ponterotto, Joseph G. "Brief Note on the Origins, Evolution, and Meaning of the Qualitative Research Concept "Thick Description"," (2006) *The Qualitative Report*. 11(3) at 538-549.

Presser, L. "Negotiating Power and Narrative in Research: Implications for Feminist Methodology," (1994) *Signs* 30(4) at 2067-2090.

Ragin, Charles C. *The Comparative Method: Moving Beyond Qualitative and Quantitative Strategies* (Berkeley: University of California Press, 1987).

Reason, Peter. *Integrating Action and Reflection through Cooperative Inquiry* (Bath, UK: University of Bath, Centre for Action Research in Professional Practice, 1998).

Reason, Peter. "Learning and Change through Action Research," in Henry, J. (ed.) *Creative Management* (London: Sage Publications, 2001).

Reissman, C. "When Gender is Not Enough: Women Interviewing Women," (1981) *Gender and Society* 2 at 172-207.

Schatzman, Leonard & Anselm, L. Strauss. *Field Research: Strategies for a Natural Sociology* (Englewood Cliffs: Prentice-Hall Inc., 1973).

Silverman, D. *Doing Qualitative Research: A Practical Handbook* (London: Sage Publications, 2000).

Slim, H. & Thompson, P. 'Ways of Listening: The Art of Collecting Testimony,' in Slim, H., Thompson, P., Bennett, O. & Cross, N. (eds.), *Listening for a Change: Oral Testimony and Development* (London: PANOS, 1993).

Spradley, James P. *The Ethnographic Interview* (New York: Harcourt Brace Jovanovich College Publishers, 1979).

Stewart, David W. & Shamdasani, Prem N. *Focus Groups: Theory and Practice*. Applied Social Research Methods Series, Vol. 20 (London: Sage Publications, 1990).

Stone, L. & J. Gabriel Campbell. 'The Use and Misuse of Surveys in International Development: An Experiment from Nepal,' (1984) *Human Organization*, 43(1) at 27-37.

Strauss, A. & Corbin, J. *Basics of Qualitative Research: Grounded Theory Procedures and Techniques*. (Newbury Park: Sage Publications Inc, 1990).

Tedlock, B. "From Participant Observation to the Observation of Participation: The Emergence of Narrative Ethnography," (1991) *Journal of Anthropology Research* 47(1) at 69-94.

Web Sources on Qualitative Interviewing:

Forum Qualitative Research

[<http://www.qualitative-research.net/fqs/fqs-eng.htm>]

Martin Ryder's Qualitative Research Site

[http://carbon.cudenver.edu/~mryder/itc_data/pract_res.html]

Oral History Society (UK)

[<http://www.oralhistory.org.uk/>]

Oral History Association (USA)

[<http://www.dickinson.edu/organizations/oha/>]



Module

6

Analysis and Presentation of Quantitative Data

Module 6 - Analysis and Presentation of Quantitative Data

Purpose: To understand the technique of basic quantitative analysis

6.1 Further Quantitative Analysis

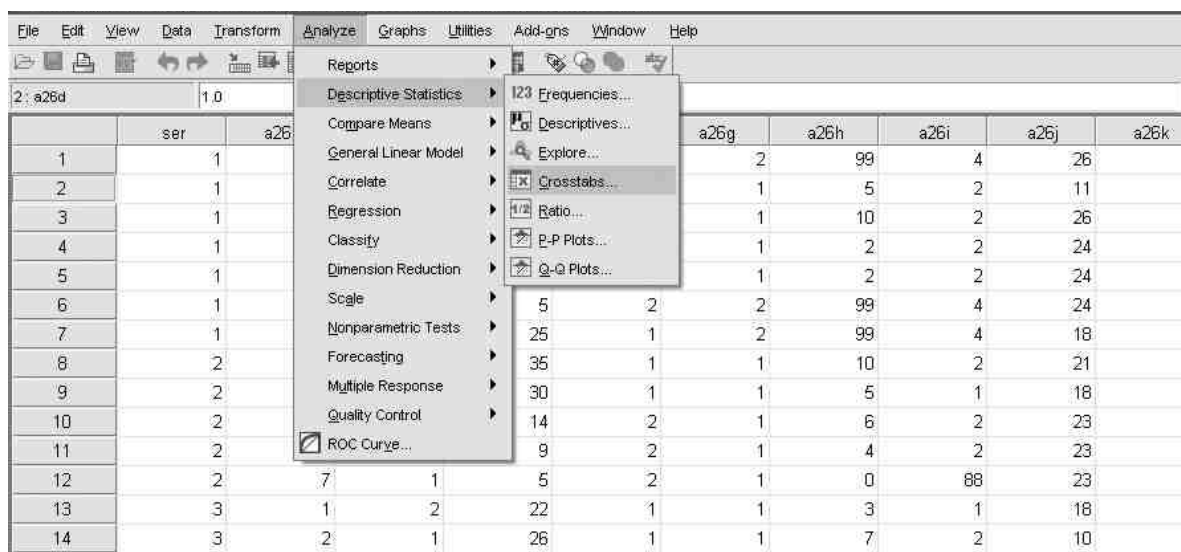
In Module 3 (Section 3.2.8 Data Coding, Entry and Cleaning) we discussed how data collected from the field is prepared for statistical analysis, i.e. how it is coded, entered and cleaned before it can be explored using statistical analysis programs such as SPSS. To recall, **coding** is the process of transforming variables and responses in your questionnaire into numbers and **data entry** is organizing the coded variables and responses in a format that can be processed by a computer program such as SPSS. While data is being entered and even after all data has been entered, several rounds of checks are performed to remove inconsistencies and ensure data accuracy, the process known as **data cleaning**.

Module 4 (Section 4.3 Descriptive Statistics) introduced to the reader the types of statistical analyses that are used to explore the (coded and cleaned) data; i.e. **univariate**, **bivariate** and **multivariate** analyses. Whereas in Module 4 the emphasis was on univariate analysis (i.e. measures of central tendency and measures of dispersion) the discussion in this Module will be around the most commonly used bivariate and multivariate analysis techniques, i.e. where the analysis goes beyond mere exploration of individual variables towards defining relationships and associations between variables.

6.1.1 Cross-tabulation

One of the simplest ways to look at the relationship between two nominal or ordinal variables (refer to Glossary) is through cross-tabulations. Cross-tabulation (crosstabs for short in SPSS) can be a 2x2 or even 3x2 bivariate table (i.e. a table that presents information on two variables) which represents percentages and column and row totals. The 'crosstab' command in SPSS can be found under 'Descriptive Statistics' (Screenshot 1):

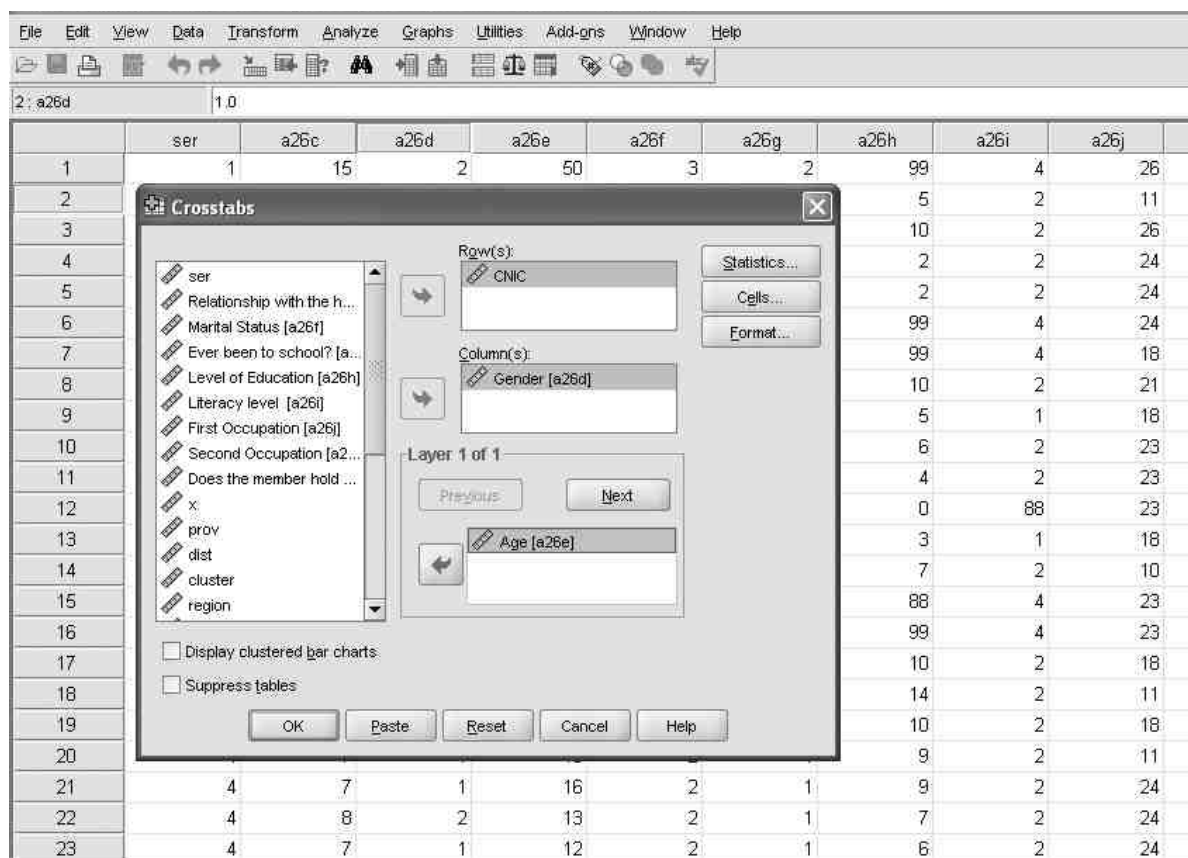
Screenshot 1



	ser	a26	a26g	a26h	a26i	a26j	a26k			
1	1	1								
2	1	1	2	99	4	26				
3	1	1	1	5	2	11				
4	1	1	1	10	2	26				
5	1	1	1	2	2	24				
6	1	1	1	2	2	24				
7	1	1	5	2	99	4	24			
8	1	1	25	1	2	99	4	18		
9	2	1	35	1	1	10	2	21		
10	2	1	30	1	1	5	1	18		
11	2	1	14	2	1	6	2	23		
12	2	1	9	2	1	4	2	23		
13	3	1	7	1	5	2	1	0	88	23
14	3	2	22	1	22	1	3	1	18	
			26	1	1	7	2	10		

Following this prompt, you will be required to identify the two variables you wish to analyze. For instance you want to determine the percentage of respondents who possess a computerized national identification card (CNIC) and you want this information disaggregated by gender.

Screenshot 2



Screenshot 2 shows how the relevant variables will be selected. If you want to determine how many of the female and how many of the male respondents hold CNICs then ‘Gender’ will be selected under Column and ‘CNIC’ under Row, i.e. in this case ‘Gender’ will be the independent variable. [If, however, you want to determine the proportion of males and females within those who hold and those who don’t hold CNICs then ‘CNIC’ will be the independent variable and thus will be selected as the Column variable whereas ‘Gender’ will be the Row variable]. The resulting Table 6.1 is as follows:

Table 6.1 CNIC * Gender Cross-Tabulation

			Gender		
			Male	Female	Total
CNIC	Yes	Count	921	788	1709
		% within Gender	83.5%	69.7%	76.5%
	No	Count	182	342	524
		% within Gender	16.5%	30.3%	23.5%
	Total	Count	1103	1130	2233
		% within Gender	100.0%	100.0%	100.0%

From Table 6.1 we can see that out of the 2,233 respondents 76.5% hold CNICs and that 83.5% of the male respondents and 69.7% of the female respondents hold CNICs. This shows that being male increases the likelihood of person being a CNIC holder. You can add further ‘layers’ to the analysis by including variables such as ‘Age’ to the crosstabs as shown in Screenshot 2. In this case the resulting table will show you the proportion of male and female CNIC holders by age (Table 6.2 below):

Table 6.2 CNIC * Gender * Age Groups Cross-Tabulation

age groups				Gender		Total
				Male	Female	
15-19 years	CNIC	1.00	Count % with Gender	33 34.7%	18 22.0%	51 28.8%
		2.00	Count % with Gender	62 65.3%	64 78.0%	126 71.2%
	Total	Count % with Gender	95 100.0%	82 100.0%	177 100.0%	
20-24 years	CNIC	1.00	Count % with Gender	90 63.8%	86 45.7%	176 53.5%
		2.00	Count % with Gender	51 36.2%	102 54.3%	153 46.5%
	Total	Count % with Gender	141 100.0%	188 100.0%	329 100.0%	
25-29 years	CNIC	1.00	Count % with Gender	102 87.2%	129 64.5%	231 72.9%
		2.00	Count % with Gender	15 12.8%	71 35.5%	86 27.1%
	Total	Count % with Gender	117 100.0%	200 100.0%	317 100.0%	
30-34 years	CNIC	1.00	Count % with Gender	128 89.5%	192 79.3%	320 83.1%
		2.00	Count % with Gender	15 10.5%	50 20.7%	65 16.9%
	Total	Count % with Gender	143 100.0%	242 100.0%	385 100.0%	

Table 6.2 shows the proportion of male and female CNIC holders by age-groups, where 1 means ‘yes’ and 2 means ‘no’. For instance we see that in the 20-24 years age group 64% of the males have CNICs whereas 46% of the females are CNIC holders; and overall, 54% of the sample in the 20-24 years age group has CNICs.

Through cross-tabs we see that there is a relationship between variables but not necessarily a cause-and-effect relation, i.e. the dependent variable is not the result (or effect) of the independent variable (or the cause). For instance holding a CNIC is not necessarily the result of being male, although being male increases the chances of a person holding a CNIC.

To summarize:¹⁷

- cross-tabs show the relationship (not causal) between two or more nominal or ordinal variables;
- in cross-tab tables the rows represent dependent variables and columns display independent variables; and
- percentage totals are for columns and are read across rows

¹⁷ Bernard (2000: 559).

Box 6.1. Interpreting Cross-Tab Tables

The table below, taken from PDHS (2006-07) [See footnote 2], shows an elaborate cross-tab table, showing the percentage of urban and rural households and population of the households in the sample by the type of toilet facilities they have access to. Notice here also that the percentage totals are down the columns and should not be interpreted across rows. Below percentage totals are given the actual numbers in each group, for example, the number of total urban households in the sample is 3,159 and the number of rural households is 6,096. The total population of sample households is 66,145 of which 43,757 lives in rural areas. Overall, 35.8% of the rural households have access to improved, non-shared toilet facilities, whereas the proportion of urban households that enjoy such facilities is 77.8%. Only 3.6% of urban households do not have access to any toilet facility whereas 43.3% of rural households reported having no toilet facility.

Table 2.10 Household sanitation facilities

Percent distribution of households and de jure population by type of toilet/latrine facilities, according to residence, Pakistan 2006-07

Type of toilet/latrine facility	Households					Population				
	Total urban	Major city	Other urban	Rural	Total	Total urban	Major city	Other urban	Rural	Total
Improved, not shared facility										
Flush/pour flush to piped sewer system	59.6	77.0	36.3	11.7	28.0	59.3	77.4	36.5	12.2	28.2
Flush/pour flush to septic tank	11.2	5.1	19.4	16.3	14.6	11.5	5.3	19.3	17.3	15.4
Flush/pour flush to pit latrine	5.7	3.6	8.4	6.0	5.9	6.1	3.9	8.8	6.6	6.4
Ventilated improved pit (VIP) latrine	0.3	0.0	0.7	0.8	0.7	0.4	0.0	0.9	0.8	0.7
Pit latrine with slab	1.0	1.1	0.8	1.0	1.0	1.1	1.3	0.9	1.1	1.1
Non-improved facility										
Any facility shared with other households	9.3	10.4	8.0	4.9	6.4	8.6	9.6	7.4	4.4	5.9
Flush/pour flush not to sewer/septic tank/pit latrine	2.2	0.8	4.1	2.6	2.5	2.5	0.7	4.8	2.8	2.7
Pit latrine without slab/open pit	0.8	0.1	1.7	3.7	2.7	0.8	0.1	1.7	4.0	2.9
Bucket	0.5	0.2	0.9	1.8	1.3	0.5	0.1	0.9	1.9	1.4
Hanging toilet/hanging latrine	5.3	0.3	11.9	7.4	6.7	5.2	0.3	11.3	7.1	6.5
Not facility/bush/field	3.6	0.8	7.2	43.3	29.8	3.4	0.8	6.7	41.2	28.4
Other	0.0	0.0	0.0	0.2	0.1	0.0	0.0	0.0	0.2	0.1
Missing	0.7	0.7	0.6	0.3	0.4	0.7	0.5	0.8	0.2	0.4
Total	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
Number	3,159	1,806	1,350	6,096	9,255	22,389	12,485	9,904	13,757	66,145

In Module 4 (Section 4.4) we discussed the significance and mechanics of hypothesis testing. Testing the difference between the means of two sample groups is a similar exercise. Whereas in hypothesis testing we try to determine whether our sample mean represents the population mean, in testing the difference between means (of two groups) we want to estimate the significance of the difference between group means. In this case, the null hypothesis is that the means of the two groups “come from random samples of the same population—that there is no difference, except for sampling error, between the two means”. The alternative hypothesis is that the two means “come from random samples of truly different populations”. (Bernard, 2000: 548)

Why do we need to test the difference between means of two sample groups? Such tests are most useful when we want to compare the performance of two groups, particularly in two-group experimental or quasi-experimental research studies, where performance on specific variables is measured before and after an intervention. For instance, a pilot study is conducted wherein a selected group of migrants becomes the beneficiary of a cash transfer programme (CTP). This group is the treatment group. Another group of migrants, with similar characteristics as the treatment group, is selected for comparison and is not the beneficiary of the CTP. The latter group is the control group. Before the introduction of the CTP certain key variables are measured for both groups, such as, household income and expenditure levels, health indicators, food security and possession of fixed assets. Six months after the CTP is introduced, the same variables are measured again for both groups. The means of variables for both groups, i.e. household income and expenditure levels of group1 and group2, health indicators for group1 and group2, and so on, are compared with each other to determine whether or not there are any differences, i.e. are the beneficiaries of CTP better off than the non-beneficiaries? If there is a difference between means then is it statistically significant?

Even if the study is non-experimental in nature, testing the difference between means is still a useful exercise. For example, if you were to assist an international aid/humanitarian agency or the government in designing repatriation strategies you would have to study the target population for effective strategy design. You will probably encounter two groups: (a) those who want to go back to the country of origin; and (b) those that want to stay in the host country or do not want to go back. In this case you would want to know how these groups differ. The former group might have stronger family ties in the home country or might have business concerns/employment they want to return to or there might have been a decline in their standard of living post migration. You would want to explore how significant these differences are and then devise appropriate strategies.

The process described in Section 4.4 (Hypothesis Testing) of Module 4, i.e. to evaluate the probability that the mean of a sample reflects the mean of the population from which it is drawn, is known as ‘the one-sample t-test’. (Bernard, 2000) To test the significance of the difference between the means of two groups the **two-sample t-test** is used. Although your statistical analysis package, such as SPSS, will calculate the t-statistic for you (and we will get to that shortly), it is useful to know how this value is calculated. The t-statistic is basically a ratio, calculated as follows¹⁸:

$$\frac{\text{Difference between group means}}{\text{(Variability of groups)}} \quad (\text{Formula 1})$$

The numerator is the difference between the means of the two groups and the denominator is a measure of the variability of scores or the standard error of the difference¹⁹. Recall from Section 4.3.2 of Module 4 (Measures of Dispersion) the calculation of variance (i.e. the average squared deviation from the mean). To calculate standard error of difference the variance of group 1 and group 2 are added and then squared. Formula 1 can alternatively be written as:

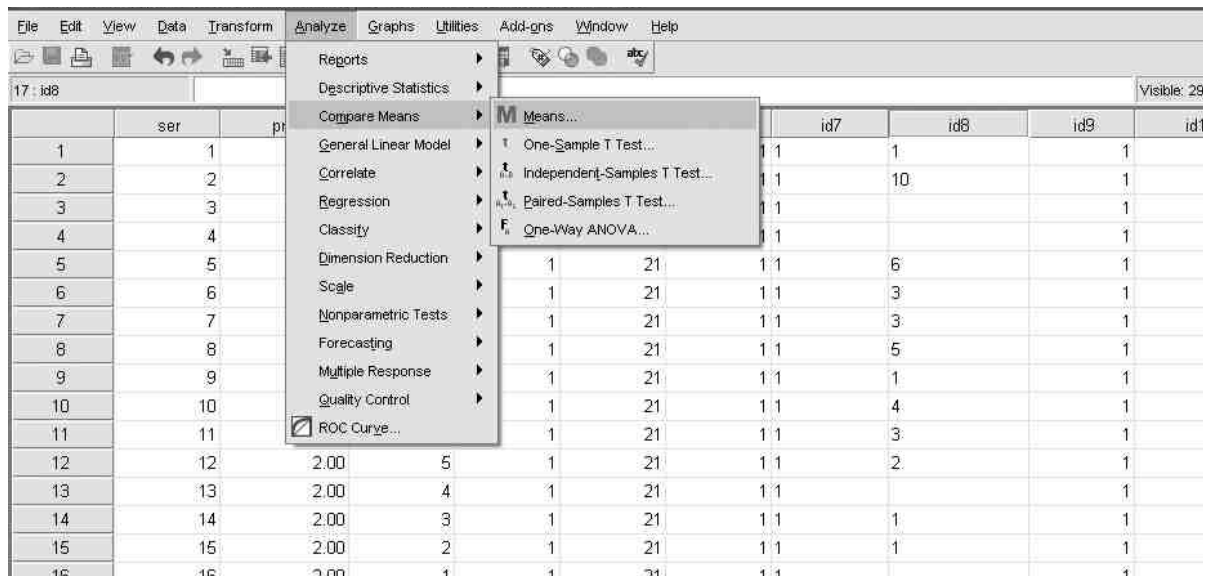
$$t = \frac{\text{mean of group1} - \text{mean of group2}}{\sqrt{\text{variance of group1} + \text{variance of group 2}}} \quad (\text{Formula 2})$$

Once we arrive at the t-value as such, we have to look it up in a table of significance such as that shown in Appendix 4 (Student’s T Distribution). As mentioned before, the t-statistic is calculated by the statistical package and you do not need to use these formulas. To conduct this test in SPSS, select ‘Analyze’ then ‘Compare Means’. This will give you several options as shown in Screenshot 3:

¹⁸ Trochim, William M. The Research Methods Knowledge Base, 2nd Edition, Analysis; The T-Test. [Accessed online at: http://www.socialresearchmethods.net/kb/stat_t.php] (version current as of October 20, 2006).

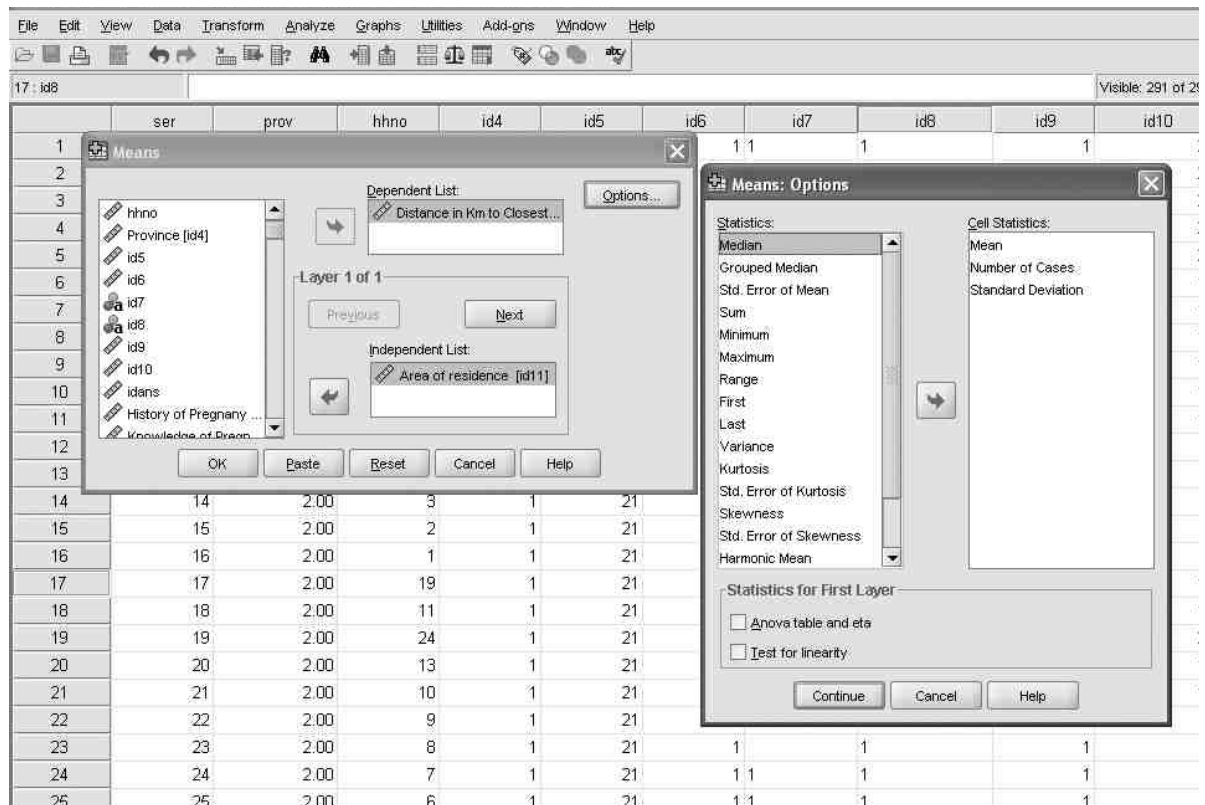
¹⁹ Ibid.

Screenshot 3



The type of test you select will depend on the nature of your study and analysis. 'Means' is the simplest form of t-test where you select the variable that divides the sample into two groups (the Independent list) and the variable whose mean you want to compare between the two groups (the Dependent list). See Screenshot 4 below:

Screenshot 4



As seen in Screenshot 3, the second option under 'Compare means' is the 'One Sample T Test', which tests whether the mean of a single variable differs from a specified constant. For instance if you are measuring the nutritional status of children in a particular community, you could use values of health and nutrition indicators, such as weight-for-age, height-for-age, set by international health organizations (such as WHO), as constant and measure your findings regarding these indicators against these constant values.

The third option is 'Independent-Samples T Test'. This procedure compares means for two groups of cases in which subjects are randomly assigned to those groups, so that any difference in response is due to the treatment (or lack of treatment) and not to other factors. However, this is not the case if you compare, for instance, the average income for males and females; a person is not randomly assigned to be a male or female.

The fourth option, 'Paired-Samples T Test', is used when the means of two variables for a single group are to be compared. It computes the differences between values of the two variables for each case and tests whether the average differs from 0. For instance in a study on nutritional status, all children in a community are measured at the beginning of the study, given supplements, and measured again. Thus, each subject has two measures, i.e. the before and after measures. Often such studies will also involve a control group, i.e. the group that does not receive treatment, whose subjects are also measured before and after the introduction of an intervention, to be compared with the subjects of the treatment group.

The final option, the 'One-Way ANOVA (or Analysis of Variance)', produces a one-way analysis of variance for a quantitative dependent variable by a single factor (independent) variable. Analysis of variance is used to test the hypothesis that several means are equal. This test can also be used to determine the difference between means. There are two types of tests for comparing means: a priori contrasts (tests set up before running the experiment) and post hoc tests (tests that are run after the experiment has been conducted).

Box 6.2. Example of t-test in SPSS

The tables in this box were computed using SPSS to compare means. The sample population has been divided into two groups, 'Rural' and 'Urban' and compared on the variable 'Distance in kilometers to the closest hospital a patient can be taken to in an emergency'. We want to see whether: (a) there is a difference between the average distances; and (b) the difference is significant.

From the first table (Group Statistics) we know that there are 301 urban and 398 rural households (from column 'N') and that the distances vary more for rural households than they do for urban households (from the 'Standard Deviation' column). We also see that the average distance from an urban household to the nearest health facility a person can be taken to in an emergency (2.56 km, SD 4.19) is less than the average distance of a similar health facility in rural settings (3.92 km, SD 4.46). Is this difference significant?

From the second table (Independent Samples Test), column 'Sig (2-tailed)' we know that this difference is significant. The significance value is 0.00 which is less than 0.05 (given 95% confidence level) hence the difference in mean distances for the two groups is significant (refer to discussion on significance levels in Section 4 of Module 4).

Group Statistics

	Area of residence	N	Mean	Std. Deviation	Std. Error Mean
Distance in Km to Closet hospital anyone can be taken to in an emergency from household	Urban	301	2.5656	4.18540	.24124
	Rural	398	3.9209	4.45514	.22332

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
Area of residence		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
Distance in Km to Closet hospital anyone can be taken to in an emergency from household	Equal variances assumed	5.551	.019	-4.087	697	.000	-1.35524	.33160	-2.00629	-.70419
	Equal variances not assumed			-4.123	665.280	.000	-1.35524	.32874	-2.00073	-.70975

Correlation describes the magnitude and direction of relationship between two or more variables. For instance the relationship between wealth and education, where education has a **positive correlation** with income, i.e. the higher the level of education of a person the higher their income. Suppose a study shows us that the coefficient of correlation between education and income is +2.5. This will be interpreted as saying that each additional year of education increases average income of a person 2.5 times. Taking another example, suppose a study tells you the coefficient of correlation between mother’s years of schooling and infant mortality is -0.5, i.e. there is a **negative correlation** between mother’s level of education and infant mortality. This means that with each additional year of mother’s schooling chances of an infant dying are reduced 0.5 times.

In addition to being positive or negative, correlations are either **linear** or **non-linear**. For instance the correlation between mother’s education and infant mortality is **linear and negatively correlated**. If plotted on a graph, it will roughly look like a line running diagonally from the *y-axis* to the *x-axis*. A **non-linear** correlation (positive or negative) is when an increase (or decrease) in one variable leads to a decrease (or increase) in the other variable, peaks at a certain point and then the relationship is reversed.

However, one needs to be careful in interpreting correlations. While there might be a relationship between two variables, it might not necessarily be a **causal relationship**, i.e. one variable might not be causing the other. Taking the relationship between mother’s education and infant mortality, on the surface it seems like increasing mother’s education reduces infant mortality, however there might be other related but hidden factors contributing to the reduction. One such variable could be mother’s age, another could be indigenous knowledge (knowledge she’s acquired from her mother or others around her), yet another could be access to health facilities or availability of qualified medical practitioners. Leaving out such causal variables from the analysis is known as the **omitted variable bias**. To control for this bias, all the important causal variables should be included in the analysis.

One also needs to be mindful of **reverse causality** while analyzing the relationship between two variables. A study might assert that conflict causes poverty, however it can very well be that poverty led to the conflict initially, which is now exacerbating poverty levels.

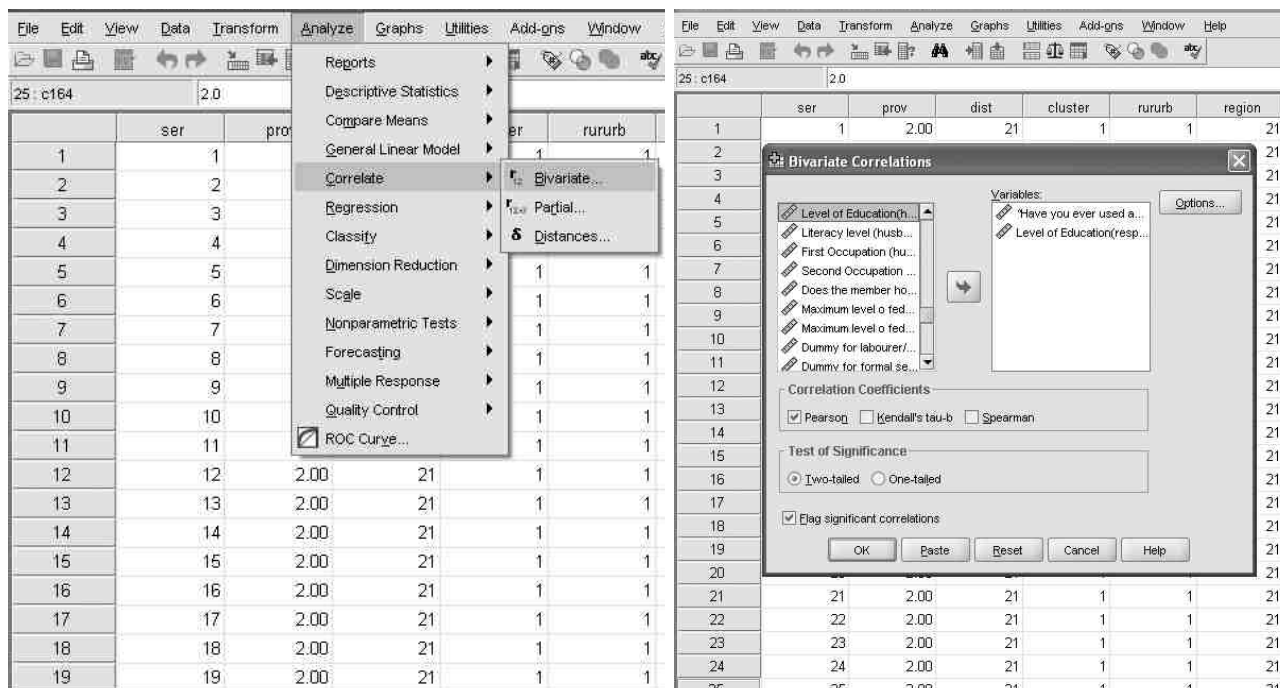
To explore correlations in SPSS, select ‘Analyze’ then ‘Correlate’ and then select the relevant variables. Please see Screenshot 5. In this example we will analyze the correlation between a married woman’s education level and ever use of contraception. Table 6.3 shows the correlation between the two variables and the significance of this correlation.

Table 6.3: Correlations (1)

		"Have you ever used any family planning method?"	Level of Education (respondant)
"Have you ever used any family planning method?"	Pearson Correlation	1	.152**
	Sig. (2-tailed)		.000
	N	699	699
Level of Education (respondant)	Pearson Correlation	.152**	1
	Sig. (2-tailed)	.000	
	N	699	699

** Correlation is significant at the 0.01 level (2-tailed).

Screenshot 5



A woman’s level of education and use of family planning are positively correlated and the coefficient of correlation is 0.152, i.e. if a woman’s education is increased by one grade the probability of a woman using contraception (family planning method) increases by 0.152. This correlation is also statistically significant at the 0.01 confidence level (denoted by the two asterisks with the coefficient and also the note at the bottom of the table).

You can also use the ‘Partial correlation’ option under ‘Analyze’. This option is used to describe the linear relationship between two variables while controlling for the effects of one or more additional variables. Taking the same example of the relationship between education and use of contraception, education might not be the only factor affecting the use of contraception, there might be other influences on the use of contraception like education level of husband or age of the woman. Controlling for these factors we get the following results:

Table 6.3: Correlations (1)

Control Variables	Level of Education (respondant)	"Have you ever used any family planning method?"
Age (Respondant) & Level of Education (husband)	Level of Education (respondant)	Correlation
	Correlation	1.000
	Correlation	.160**
	Correlation	.160**
	Correlation	1.000

** Correlation is significant at the 0.01 level

The coefficient of correlation is still positive and statistically significant. This shows that there is a direct relationship between education level of woman and use of contraception. However, if the value of the coefficient or the level of significance had decreased after controlling for other variables it would have meant that the factors that have been controlled for and not only the education level of the woman affect the use of contraception.

6.1.3 Regression

"Correlations are best understood in the context of regression" (Bernard, 2000: 606). Regression analysis is one of the most commonly used statistical analysis techniques in social science research. The objective of this analysis is to explore the relationship between a dependent variable and one or more independent variables. More specifically the analysis helps one predict if and how the value of the dependent variable changes when any one of the independent variables is varied, while other independent variables, if any, are held fixed. For instance suppose you want to identify the variables that have the greatest impact on internal migration. In this case the dependent variable will be the decision to migrate and the independent variables could be financial considerations, security concerns, family ties, better education, and so on. Your analysis will show you which of these independent variables actually have an impact on migration decision and what type of relationship they have with the dependent variable, i.e. whether it's positive or negative, weak or strong. Regression analysis, at times, can also be used to infer causal relationship between the independent and dependent variables.

There are several forms of regression analysis, the most common one being the linear regression or Ordinary Least Squares (OLS) analysis. As Bernard (2000: 582) explains, the mathematical equation for linear regression analysis is as follows:

$$y = a + bx$$

Where y is the dependent variable, a is constant, b is the regression coefficient and x is value of the independent variable. a is calculated as follows:

$$a = \bar{y} - b\bar{x}$$

Where \bar{y} is the mean of the dependent variable and \bar{x} is the mean of the independent variable and b is calculated as follows:

$$b = \frac{N(\sum xy) - (\sum x)(\sum y)}{N(\sum x^2) - (\sum x)^2}$$

Where N is the sample size and \sum denotes sum. Again, you will not need to apply these formulas yourself since the computer program will calculate these statistics for you; the purpose of showing these formulas here is to facilitate comprehension of how the system generates these statistics.

The values produced by the regression equation contain an **error or variance** that results from the difference between the predicted number for the dependent variable and its actual measurement. "This is also called the residual- that is what is left over after making your predictions using the regression equation". (Bernard, 2000: 584) This quantity is referred to as r^2 (r squared), which tells us how much of the variance in the dependent variable is accounted for by the scores of the independent variable(s).

Before we move on to a practical example of regression analysis, another important aspect to cover is the concept of dummy variables. "A **dummy variable** is a numerical variable used in regression analysis to represent subgroups of the sample in your study...Dummy variables are useful because they enable

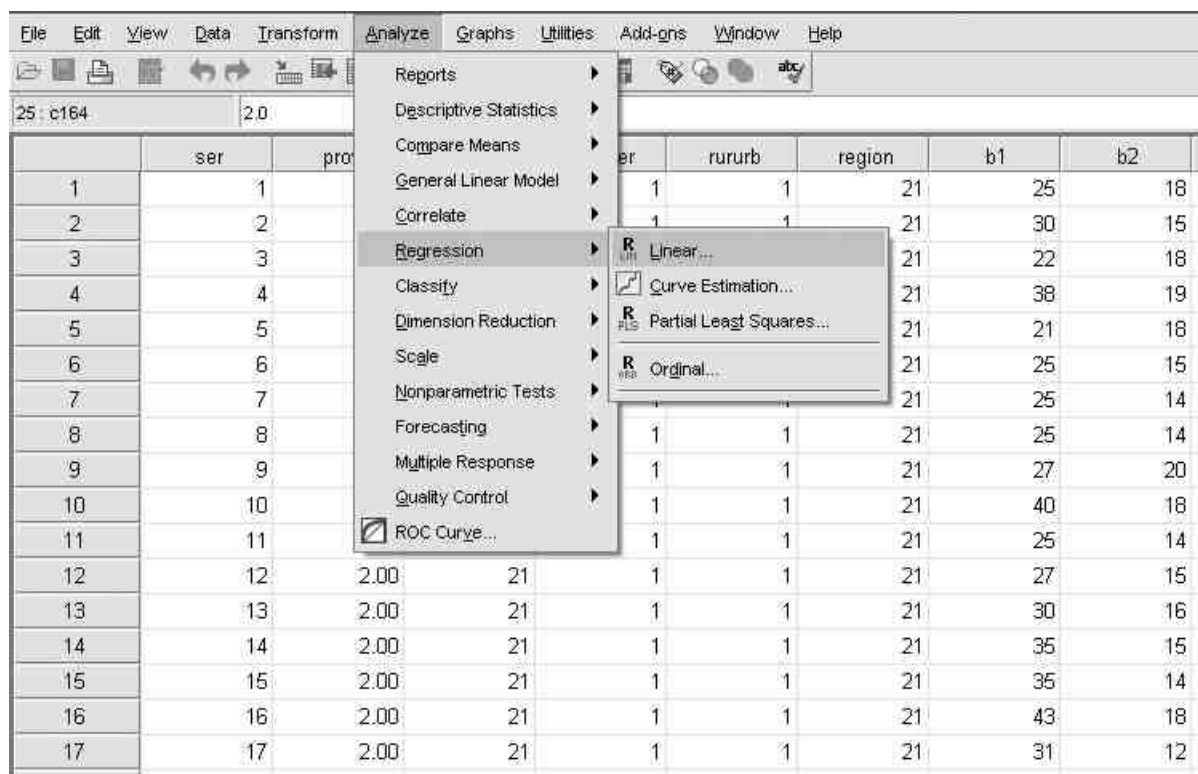
²⁰ Trochim, William M. *The Research Methods Knowledge Base*, 2nd Edition, Analysis; Dummy Variables. [Accessed online at: <http://www.socialresearchmethods.net/kb/dummyvar.php>] (version current as of October 20, 2006).

us to use a single regression equation to represent multiple groups"20. These variables are nominal i.e. 1, 0, which means that a condition is either present or absent. If you have, for instance males and females in your sample you could create a dummy variable 'female' which assigns a value of '1' to all females in the sample and '0' to all the males. Another example, suppose you have a list of health providers your respondents visit when they need treatment which includes doctors, nurses, lady health workers, hakim, pir/maulvi and compounder. You could create a dummy variable to differentiate between respondents who go to safe providers (i.e. doctors, nurses, LHWs) and those who go to unsafe providers (i.e. hakim, compounder, pir/maulvi). Suppose you call this variable 'safe', which assigns a value of '1' to all respondents who use services of safe providers and '0' to all those who don't. This way you will include all groups of respondents in your analysis without having to conduct the analysis separately for each type of respondent.

To demonstrate with an example from SPSS, suppose we want to identify the factors that have the greatest impact on a woman's decision to seek treatment for her illness. In this example our dependent variable is 'seeking treatment for illness' and our independent variable is 'distance from health facility'. The sequence of commands to give SPSS is:

Analyze → Regression → Linear (See Screenshot 6)

Screenshot 6



In the next command window, you identify the dependent and independent variables, along with the types of statistics that you want. SPSS will generate the following 3 tables once you hit press OK:

Table 6.5: Model Summary (1)

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.105 ^a	.011	.005	.00752

a. Predictors: (Constant), remotnes quintile

Table 6.6: ANOVA^b (1)

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	.286	1	.286	1.902	.170 ^a
	Residual	25.829	172	.150		
	Total	26.115	173			

a. Predictors: (Constant), remotnes quintile

b. Dependent Variable: Dummy Sought treatment for complication

Table 6.7: Coefficients^a (1)

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	.903	.070	-.105	12.942	.000
	remotness quintile	-.029	.021		-1.379	.170

a. Dependent Variable: Dummy Sought treatment for complication

The first table gives us the values of R, R squared and adjusted R squared. R is the correlation between the observed and predicted values of the dependent variable. The values of R range from -1 to 1. The sign of R indicates the direction of the relationship (positive or negative). The value of R indicates the strength, with larger values indicating stronger relationships. Here the value of R is 0.105, which is not a very strong correlation. R squared is the proportion of variation in the dependent variable explained by the regression model. The values of R squared range from 0 to 1. Small values indicate that the model does not fit the data well, which seems to be the case here (R square is 0.11). Adjusted R squared attempts to correct R squared to more closely reflect the goodness of fit of the model in the population. R Squared is used to determine which model is best.

We can also use the second table, ANOVA (analysis of variance) to determine how useful our model is (i.e. is it a good fit?). If the significance value of the F-statistic is less than 0.05 then we can assume that the independent variable does explain, to a large extent, the variation in the dependent variable. In this case the significance value of the F-statistic is 0.170, reiterating that this model does not explain well the variation in the dependent variable.

Finally, the last table shows us the correlation between the dependent and independent variables. The correlation coefficient is -0.029, which means that as remoteness from health facility increases by a factor of 1 (for instance by 5 kilometers) the probability of seeking treatment goes down by 2.9%. However, this correlation is not statistically significant at the 95% confidence level (significance value > 0.05).

To reduce the error/residual, usually two or more variables are introduced into the regression equation, thus making it a **multiple regression equation**. Continuing with the same example, we

introduce four more variables into the equation; the severity of illness, respondent's freedom of movement, involvement in own marriage decision and respondent's education level. The following tables are produced by SPSS:

Table 6.8: Model Summary (2)

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.385 ^a	.148	123	.36391

a. Predictors: (Constant), Agency in marriage decision, severity, Mobility score, remotnes quintile, Education of respondent

Table 6.9: ANOVA^b (2)

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	3.867	5	.773	5.840	.000 ^a
	Residual	22.248	168	.132		
	Total	26.115	173			

a. Predictors: (Constant), Agency in marriage decision, severity, Mobility score, remotnes quintile, Education of respondent

b. Dependent Variable: Dummy Sought treatment for complication

Table 6.10: Coefficients^a (2)

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	.165	.163		1.016	.311
	remotness quintile	-.011	.021	-.039	-.538	.591
	Mobility score	.023	.010	.169	2.357	.020
	severity	.175	.056	.224	3.125	.002
	Education of respondent	.010	.006	.123	1.624	.106
	Agency in marrage decision	.051	.022	.175	2.289	.023

a. Dependent Variable: Dummy Sought treatment for complication

From the first table, Model Summary, we see that the values of R, R squared and adjusted R square have improved with the introduction of the additional variables. While the earlier model explained 1.1% (value of R squared) of the variance in the dependent variable, this model explains 14.8% of the variance. From the second table we see that the significance value of the F-statistic has also improved, i.e. it is significant at the 95% confidence level. From the third table, coefficients, we can identify three variables that have a statistically significant impact on the decision to seek treatment for illness; the respondent's mobility, the severity of illness and the respondent's involvement in her own marriage decision (since all three have significance values that are less than 0.05), the most significant one being severity of illness. What this means is that the more mobile a woman is, the more say she has in important decisions of her life (such as marriage) and the more severe the illness is, the more likely she is to seek treatment for her illness. Other factors i.e. distance from health facility and education level of the respondent, do not seem to have any significant impact on the decision to seek treatment.

There are a few underlying assumptions for regression analysis²¹:

- On average, the errors in estimation balance out;
- The independent variables are non-random and have fixed and finite values;
- The independent variables are linearly independent, i.e. no independent variable in the equation can be expressed as a combination of the remaining independent variables, a condition which is known as multi-collinearity; and
- The variance of errors is constant across observations.

6.2 Conclusion

Quantitative data analysis is a complex exercise and one that requires practice. However it enables us to quantify concepts that are often abstract and support our arguments with numbers and quantifiable targets. Regarding statistical significance, bear in mind that the significance levels commonly used (i.e. 0.05 and 0.01) are arbitrary; a statistic that is not significant at these levels is not necessarily redundant, it might still be an important finding, one that might warrant further exploration. What these tests of significance do is give us some level of confidence in our research findings. In trying to find correlations between variables and trying to find the best fitting models we tend to eliminate outliers (i.e. those variables that are weakly correlated to the outcome of interest and lower the goodness of fit of a model). However these outliers might make the story more interesting and thus should not be completely excluded from the analysis and you might consider reporting them with your results.

²¹ Economics Laboratory Software Archive, University of California Berkeley; Regression Analysis:
[<http://elsa.berkeley.edu/sst/regression.html#OLS>]

Further Reading

On Multivariate and Bivariate analysis:

Bernard, H. Russell. *Social Research Methods: Qualitative and Quantitative Approaches*, (Thousand Oaks, CA: Sage Publications, 2000), at Chapters 14-16.

Tabachnick, Barbara G and Fidell, Linda S. *Using Multivariate Statistics*, Fourth Edition. (Boston: Allyn and Bacon, 2001) [Accessed online at: <http://www.er.uqam.ca/nobel/r16424/PSY7102/Document3.pdf>]

On Regression Analysis:

StatSoft , *Electronic Statistics Textbook*. [Accessed online at: <http://www.statsoft.com/textbook/multiple-regression/>]

Economics Laboratory Software Archive, University of California Berkeley. [Accessed online at: <http://elsa.berkeley.edu/sst/regression.html>]

Chatterjee, Samprit and Hadi, Ali S. *Regression Analysis by Example* (Hoboken, NJ: A John Wiley & Sons, 2006).

References

Bernard, H. Russell. *Social Research Methods: Qualitative and Quantitative Approaches* (Thousand Oaks, CA: Sage Publications, 2000).

Economics Laboratory Software Archive, University of California Berkeley. [Accessed online at: <http://elsa.berkeley.edu/sst/regression.html>]

National Institute of Population Studies & Macro International Inc. *Pakistan Demographic and Health Survey 2006-7* (Islamabad, 2008).

Trochim, William M. *The Research Methods Knowledge Base*, 2nd Edition, Sampling/External Validity. [Accessed online at: <http://www.socialresearchmethods.net/kb/external.php>] (version current as of October 20, 2006).

Analysis of Qualitative Data

Module 7 - Analysis of Qualitative Data

Purpose: To learn the basics for analyzing qualitative data

7.1 Managing Your Data

In qualitative research managing and organizing your data for analysis is probably the most difficult part of the research. Since all of your observations are converted into text, e.g. interview transcript or summary, community and respondent profiles, field journal observations, commentary on interviews, etc. there will be a lot of written material to make sense of before you can start analyzing it. What follows are a series of suggested steps based on experiences of working in Pakistan.

- i) Transcribe all recorded material or write a summary in the language of analysis (assume it is English) of all recorded interviews.
- ii) Summarize all focus group discussions and interviews, in English. However, highlight some original quotes that you think may be useful to illustrate an aspect of your analysis.
- iii) Write up your case studies – whether it is a particular case study of a family, based on your interview material and observations, or a case study of an organization, write it up and keep it for use later.
- iv) Put together all notes pertaining to one community: i.e. the community profile, respondents profile, interview summaries, focus group discussion summaries, field observations, and journal notes/commentaries. These can be further organized into the material relevant from individual interviews.
- v) Keep all of your original data, in its original form, safely, but in such a way that the identity of the respondents can remain confidential within your research team.

Another factor related to the above is building into your research budget and programme enough funds to have the data prepared for analysis (e.g. translated, typed) and sufficient time allocated for this purpose.

7.2 Data Triangulation

Qualitative research allows for different kinds of triangulation, a process that helps you to establish the validity of your data. It has been defined as “the combination of methodologies in the study of the same phenomenon.” (Denzin, 1978: 291) In effect triangulation is the use of multiple methods to secure an in-depth understanding of the phenomenon in question.” (Denzin and Lincoln, 2003) It can allow for new insights that would not have been discovered through one method alone. Triangulation of your data can be used to verify your findings and find converging answers from different sources to the same questions. It does not follow the rule of scientific checking for validity and reliability in research, but it is built in to the methodology of qualitative investigation in another way. This includes the collection, analysis and interpretation of data from different perspectives. (Pederson, 1992: 44)

Denzin (1978) and Patton (1999) identify four types of triangulation:

7.2.1 Methods triangulation - checking out the consistency of findings generated by different

data collection methods. One such example is the use of qualitative and quantitative data in a study. Methods triangulation can bring out complementary aspects of the same phenomenon, or also inconsistencies that are of interest to the qualitative researcher to explore.

7.2.2 Triangulation of Sources – examining the consistency of different data sources from within the same method. For example, sources approached at different points in time, in public vs. private settings, or comparing people with different viewpoints.

7.2.3 Analyst Triangulation - using multiple analysts to review findings or using multiple observers and analysts. This can provide a check on selective perception and illuminate blind spots in an interpretive analysis. Here the goal is not to seek consensus, but to understand multiple ways of seeing the data.

7.2.4 Theory/Perspective Triangulation - using multiple theoretical perspectives to examine and interpret the data.

In the kind of qualitative research that you are most likely to do, your process of triangulation is likely to take these forms:

- Comparison with and cross-checking of findings from quantitative research
- Using different field tools to establish the most valid possible account of a phenomenon: e.g. community profile/key informant interview/in-depth interview.
- Conducting repeat field visits to check validity of data if necessary.

7.3 Analyzing Your Data

Qualitative data is best analyzed *thematically*. Themes can be identified through two main approaches. First, your interview guidelines will reflect your research questions and represent certain themes. Second, your data will contain material that relates to these themes, but it will also contain material that was unanticipated. If it does not, then you have probably not allowed your interviewing to remain open-ended. It may also be that certain themes you included in your interview guidelines do not resonate at all with your respondents. For example, you may have included a question on community member’s assumed fear of vaccination campaigns, but you find out in the field that no one expresses any such fear. Hence, there is no need to include in your analysis a detailed discussion of fear of vaccination campaigns, suffice it to simply note that community members did not exhibit it.

The unanticipated material from the field is often the most interesting. It may not have occurred to you to ask respondents about domestic violence, yet you may find that women giving interviews and focus group discussion are quite assertive in bringing this up as a problem they face in the camp setting. You will have to address this in your analysis and see how it fits in with your larger research questions, and possibly suggest further research on the subject.

7.3.1 Observations and Field Notes

Your written observations, in the field and in your journal, are integral to the analysis of interviews. These allow you to include all the material that the interview itself was not able to capture necessarily. What was the mood of the encounter? Were others present, was the interview interrupted, were events happening on the sidelines that had an impact? How do you reflect upon your own role in the interview? Did the respondent sense any bias from you? Did you have to rush it? Did the interview digress into areas that you didn’t feel equipped to handle?

All of this is valuable data, and can be included in what is termed a “commentary” that you can develop on the interview material. This is a compilation of your notes that also includes your reflections on the event and your initial thoughts as to the kind of analysis that may take shape. The commentary can be the source of the “codes” that you will use for analysis at the next stage.

7.3.2 Coding

Codes are names that you give to certain portions of the text of your interview, that represent a theme or analytical category you find useful. For example, you could highlight a sentence and place next to it the code “violence against women” when a woman talks about her experience of such violence. Here are some more detailed coding techniques you can employ for analyzing your data, drawn from Miller’s (1999: 145-158) suggestions on researching life stories and family histories.

A. Based on Interview Guidelines

Read through the interview transcript and mark codes in the margins that are based on the thematic analysis you had in mind when conducting the interview. Text can be included in more than one code. Note the links between topics, or codes, that are being made by the respondent during the interview. Identify any new ideas or thoughts regarding codes/categories that you may have while reading the interview.

B. Based on New Categories Emerging from the Data

Read through the same interview again, this time looking for text that corresponds to the new categories you had in mind. After this, refine your category system.

C. Apply Revised Codes to a New Interview

Try coding the next interview based on your revisions to the earlier code list, and see if it works or if you need to make further revisions to your concepts in order to capture what is happening in an interview.

The process of coding, categorizing, and testing concepts is ongoing and there is no clear endpoint for the qualitative researcher. You will get to a point at which patterns emerge more clearly and consistently across interview transcripts, you will find some repetitions in associations between codes, and then you will feel more confident to assert your findings.

D. Use of Computer-Aided Qualitative Data Analysis

There are a number of software programmes that allow you to analyze the text of an interview according to codes of your choosing. These codes can be thematic, descriptive, or anything else, and when applied across interviews allow for you to compare interview material and identify patterns in responses. Coding is extremely time-consuming, and for this reason in our context it is rarely used. When it is used, it is applied often to interview summaries rather than transcripts. However, once interviews are coded more possibilities for comparison and analysis open up that would not be possible unless you have such computer aided tools.

Some organizations have experience with coding qualitative research. It may be possible to get training through these organizations. One well-known software programme is called The Ethnograph and another one called NuDIST. They are expensive to purchase, but allow for updates regularly.

Here are what qualitative data analysis programmes can do:

- They can locate individual words or phrases within a transcript or across interviews, i.e. when a certain topic is mentioned.
- The process of coding, or categorizing your interview material is sped up if this is done not by hand but by computer, and allows for portions of text to have multiple codes.
- Keep a record of your notes to yourself, patterns you have identified in codes, and different stages of your research. (Miller, 150-53)

E. Narrative Approach

You can look at the same interview transcript that you have worked with for the above analytical approaches, but this time do not analyze the material by coding it.

- a. Instead, you will concentrate upon the interaction between the interview and interviewee, and how the interaction may have affected the conduct of the interview. What was the influence on the interview of factors such as these:

- i. The location – was it conducive to the conversation?
 - ii. The age and/or class difference between you and the respondent – how did it affect your probing?
 - iii. Were there topics that were avoided, by you or the respondent?
 - iv. How did your style affect what the respondent said; did you inadvertently steer the course of the interview for the respondent?
- b. You can look at the transcripts again and this time categorize the text by where the respondent is:
- i. Reporting facts or events
 - ii. Narrating a story
 - iii. Making an assertion or argumentation

Summarize what are the main points of arguments or assertions made and see whether they coincide or conflict with the conceptual structure you developed in the previous exercise.

7.4 Presentation of Findings

When you are ready with your analytic framework and know what it is you really would like to report as research findings, the next step is to decide on what is the best way to present them. You will not want to make tables with so-called quantitative results out of your sample of respondents, although this is a very common mistake. For example, if you have interviewed ten people, there is no point in presenting a table to show how many people out of ten felt a given way about a given item. Trying to generalize from an inadequate sample (in quantitative terms) is not recommended. Instead, there are a number of different forms in which you can present your analyzed data that draw on the strengths of qualitative research methodology. Some are as follows:

- *A thematic analysis*, using quotes from interviews, in a research report. This is best used in a report that elicits views and perceptions from respondents, or explores themes as part of an argument.
- Using life histories for *comparative analysis*.
- Visual findings, e.g. maps from the field, audio-visual material, photographs, used in reports and presentations.

No matter in which form you would like to present your analysis, it is always important to contextualize your research. If you are doing your research in a camp, take some time in your report to explain the origins of the camp, its social and political context, and details about the community. You will also need to clearly state what was the purpose of your research, your questions/hypothesis, and why you chose qualitative research methods to look for answers. You may quote directly from interviews, summarize and compare them, and more, but all in the support of an argument that you are making to buttress your conclusion.

7.5 Practice Exercises

- a. Summarize a focus group discussion based on your practice exercise in Module 5.
 - i. Do a preliminary coding of the summary.
 - ii. Prepare a preliminary code-list based on the summary.
- b. Read the transcript of an in-depth interview you conducted in Module 5.
 - i. Identify any gaps that you could fill by repeating the interview.
 - ii. Code the transcript based on the themes reflected in your interview guidelines.
 - iii. Code it again for new themes, and sub-categories under which you can organize your codes.

References and Further Reading

Bee, Robert and Benjamin F. Crabtree. "Using ETHNOGRAPH in Field Note Management," in Boone & Wood (eds.). *Computer Applications for Anthropologists* (Belmont, CA: Wadsworth, 1992).

Green, William H. *Econometric Analysis*, 5th Edition (New Jersey: Prentice Hall, 2002).

Khan, Ayesha. "Song of War," in (2002) *Meridians: feminism, race, transnationalism* 2(2) 259-265 .

---. "Gendering War Talk: 'We Are Scattered Like Seeds and the World is Full of Us,'" in (2003) *International Feminist Journal of Politics* 5(3).

---. "Afghan Refugee Women's Experience of Conflict and Disintegration," in (2002) *Meridians: feminism, race, transnationalism* 3(1) at 89-121.

Shirkat Gah. *Stories Derived from True Life: Violence, Health and Rights, the Three Important Aspects of a Woman's Life* (Lahore: Shirkat Gah, 2008)

Silverman, David. 2001. *Interpreting Qualitative Data: Methods for Analyzing Talk, Text and Interaction* (London: Sage Publications, 2001).

Denzin, N.K. & Y.S. Lincoln. "Introduction: Entering the Field of Qualitative Research," in Denzin, N.K. & Lincoln, Y.S. (eds.). *Strategies of Qualitative Inquiry* (Thousand Oaks, CA: Sage Publications, 2003).

Denzin, N.K. *Sociological Methods* (New York: McGraw-Hill, 1978).

Pederson, Duncan. "Qualitative and Quantitative: Two Styles of Viewing the World of Two Categories of Reality?" in Scrimshaw, N. & Gleason, G. (eds.). *RAP: Rapid Assessment Procedure* (Boston: International Nutrition Foundation for Developing Countries, 1992) at 39-49.

Cohen, D. & Crabtree, B. "Qualitative Research Guidelines Project" (July, 2006). [Accessed online on September 5, 2011 at: <http://www.qualres.org>].

Patton, M.Q. "Enhancing the Quality and Credibility of Qualitative Analysis," (1999) *HSR: Health Services Research* 34(5) at 1189-1208.

Miller, Robert L. *Researching Life Stories and Family Histories* (London: SAGE Publications, 1999).



Appendices

Migration Related Academic Journals

- Asian and Pacific Migration Journal* [Scalabrini Migration Center]
Disasters [Wiley-Blackwell]
Forced Migration Review [University of Oxford]
International Journal of Migration, Health, and Social Care [Pier Professional]
International Journal of Refugee Law [Oxford Journals]
International Migration [Wiley-Blackwell]
International Migration Papers [International Labour Organization]
International Migration Review [Wiley-Blackwell]
Journal of Internal Displacement [EV Research Inc.]
Journal of Refugee Studies [Oxford University Press]
Peace and Conflict: Journal of Peace Psychology [Routledge]
Refugee Transitions [STARTTS]
Refugee Survey Quarterly [Oxford University Press]
Studies in Transnational Legal Policy [American Society of International Law]

List of International Migration Research Institutes

Institutes

Afghan Research & Evaluation Unit
Kabul, Afghanistan
<http://www.areu.org.af>

Alan Shawn Feinstein International Center
Tufts University
Medford, MA, United States
<http://fic.tufts.edu/>

Asia Pacific Forum on Women, Law and Development (APWLD)
Labour and Migration Programme
Chiang Mai, Thailand
www.apwld.org

Asia Research Institute
National University of Singapore
http://www.ari.nus.edu.sg/article_view.asp?id=1

Asian Migrant Centre
Hong Kong
www.asian-migrants.org

The Brookings Institution and London School of Economics
Project on Internal Displacement
<http://www.brookings.edu/projects/idp.aspx>

Centre for Research in International Migration and Ethnic Relations (CEIFO)
Stockholm University, Sweden
<http://www.ceifo.su.se/pub/jsp/polopoly.jsp?d=5990>

Centre on Migration, Policy and Society
University of Oxford, United Kingdom
www.compas.ox.ac.uk

Development Research Centre on Migration, Globalization and Poverty
<http://www.migrationdrc.org/index.html>
Graduate Institute of International and Development Studies
Programme for the Study of Global Migration (Geneva)
<http://graduateinstitute.ch/globalmigration>

Institute for the Study of International Migration
Georgetown University, Washington, DC, USA
www.isim.georgetown.edu

Internal Displacement Monitoring Centre
Geneva, Switzerland
<http://www.internal-displacement.org>

Internally Displaced Persons Vulnerability Assessment and Profiling (IVAP)
<http://www.ivap.org.pk/Introduction.aspx>
International Center for Migration, Ethnicity and Citizenship (ICMEC)
New School University, United States
<http://www.newschool.edu/icmec/>

International Migration Institute
Oxford, UK
www.imi.ox.ac.uk

International Migration Research Centre
Wilfred Laurier University, Waterloo, Canada
http://www.wlu.ca/homepage.php?grp_id=2599

International Organization for Migration, Policy & Research
Geneva, Switzerland
<http://www.iom.int/jahia/Jahia/policy-research/lang/en>

Migration Policy Institute
Georgetown University, Washington, DC, USA
<http://www.migrationpolicy.org/>

Refugee and Migratory Movements Research Unit (RMMRU)
Dhaka, Bangladesh
www.rmmru.org

Refugees Studies Centre
Oxford University, UK
www.rsc.ox.ac.uk

Social Science Research Council (SSRC)
Migration Program
New York, USA
www.ssrc.org/programs/migration-program/

Scalabrini Migration Center
Quezon, Philippines
www.smc.org.ph

Sussex Centre for Migration Research
University of Sussex, United Kingdom
<http://www.sussex.ac.uk/migration/>

United Nations High Commissioner for Refugees
www.unhcr.org
Policy Development and Evaluation Service
<http://www.unhcr.org/pages/4a1d28526.html>
Protection Policy and Legal Advice Section <http://www.unhcr.org/pages/4a16b17a6.html>

Z-Score Table (or Table of Areas under the Curve)

z	0	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0	0	0.004	0.008	0.012	0.016	0.0199	0.0239	0.0279	0.0319	0.0359
0.1	0.0398	0.0438	0.0478	0.0517	0.0557	0.0596	0.0636	0.0675	0.0714	0.0753
0.2	0.0793	0.0832	0.0871	0.091	0.0948	0.0987	0.1026	0.1064	0.1103	0.1141
0.3	0.1179	0.1217	0.1255	0.1293	0.1331	0.1368	0.1406	0.1443	0.148	0.1517
0.4	0.1554	0.1591	0.1628	0.1664	0.17	0.1736	0.1772	0.1808	0.1844	0.1879
0.5	0.1915	0.195	0.1985	0.2019	0.2054	0.2088	0.2123	0.2157	0.219	0.2224
0.6	0.2257	0.2291	0.2324	0.2357	0.2389	0.2422	0.2454	0.2486	0.2517	0.2549
0.7	0.258	0.2611	0.2642	0.2673	0.2704	0.2734	0.2764	0.2794	0.2823	0.2852
0.8	0.2881	0.291	0.2939	0.2967	0.2995	0.3023	0.3051	0.3078	0.3106	0.3133
0.9	0.3159	0.3186	0.3212	0.3238	0.3264	0.3289	0.3315	0.334	0.3365	0.3389
1	0.3413	0.3438	0.3461	0.3485	0.3508	0.3531	0.3554	0.3577	0.3599	0.3621
1.1	0.3643	0.3665	0.3686	0.3708	0.3729	0.3749	0.377	0.379	0.381	0.383
1.2	0.3849	0.3869	0.3888	0.3907	0.3925	0.3944	0.3962	0.398	0.3997	0.4015
1.3	0.4032	0.4049	0.4066	0.4082	0.4099	0.4115	0.4131	0.4147	0.4162	0.4177
1.4	0.4192	0.4207	0.4222	0.4236	0.4251	0.4265	0.4279	0.4292	0.4306	0.4319
1.5	0.4332	0.4345	0.4357	0.437	0.4382	0.4394	0.4406	0.4418	0.4429	0.4441
1.6	0.4452	0.4463	0.4474	0.4484	0.4495	0.4505	0.4515	0.4525	0.4535	0.4545
1.7	0.4554	0.4564	0.4573	0.4582	0.4591	0.4599	0.4608	0.4616	0.4625	0.4633
1.8	0.4641	0.4649	0.4656	0.4664	0.4671	0.4678	0.4686	0.4693	0.4699	0.4706
1.9	0.4713	0.4719	0.4726	0.4732	0.4738	0.4744	0.475	0.4756	0.4761	0.4767
2	0.4772	0.4778	0.4783	0.4788	0.4793	0.4798	0.4803	0.4808	0.4812	0.4817
2.1	0.4821	0.4826	0.483	0.4834	0.4838	0.4842	0.4846	0.485	0.4854	0.4857
2.2	0.4861	0.4864	0.4868	0.4871	0.4875	0.4878	0.4881	0.4884	0.4887	0.489
2.3	0.4893	0.4896	0.4898	0.4901	0.4904	0.4906	0.4909	0.4911	0.4913	0.4916
2.4	0.4918	0.492	0.4922	0.4925	0.4927	0.4929	0.4931	0.4932	0.4934	0.4936
2.5	0.4938	0.494	0.4941	0.4943	0.4945	0.4946	0.4948	0.4949	0.4951	0.4952
2.6	0.4953	0.4955	0.4956	0.4957	0.4959	0.496	0.4961	0.4962	0.4963	0.4964
2.7	0.4965	0.4966	0.4967	0.4968	0.4969	0.497	0.4971	0.4972	0.4973	0.4974
2.8	0.4974	0.4975	0.4976	0.4977	0.4977	0.4978	0.4979	0.4979	0.498	0.4981
2.9	0.4981	0.4982	0.4982	0.4983	0.4984	0.4984	0.4985	0.4985	0.4986	0.4986
3	0.4987	0.4987	0.4987	0.4988	0.4988	0.4989	0.4989	0.4989	0.499	0.499

Read the table by using the rows to find the first digit, and the columns to find the second digit of a Z-score. To find 1.96, first look down the rows to find 1.9 and then across the columns to 0.06 and 0.4750 will be the result.

source: <http://www.statsoft.com/textbook/distribution-tables/>

Student's T Distribution

Tail Probabilities

One Tail	0.10	0.05	0.025	0.01	0.005	0.001	0.0005
Two Tails	0.20	0.10	0.05	0.02	0.01	0.002	0.001

D	1	3.078	6.314	12.71	31.82	63.66	318.3	637	1
E	2	1.886	2.920	4.303	6.965	9.925	22.330	31.6	2
G	3	1.638	2.353	3.182	4.541	5.841	10.210	12.92	3
R	4	1.533	2.132	2.776	3.747	4.604	7.173	8.610	4
E	5	1.476	2.015	2.571	3.365	4.032	5.893	6.869	5
E	6	1.440	1.943	2.447	3.143	3.707	5.208	5.959	6
S	7	1.415	1.895	2.365	2.998	3.499	4.785	5.408	7
	8	1.397	1.860	2.306	2.896	3.355	4.501	5.041	8
O	9	1.383	1.833	2.262	2.821	3.250	4.297	4.781	9
F	10	1.372	1.812	2.228	2.764	3.169	4.144	4.587	10
	11	1.363	1.796	2.201	2.718	3.106	4.025	4.437	11
F	12	1.356	1.782	2.179	2.681	3.055	3.930	4.318	12
R	13	1.350	1.771	2.160	2.650	3.012	3.852	4.221	13
E	14	1.345	1.761	2.145	2.624	2.977	3.787	4.140	14
E	15	1.341	1.753	2.131	2.602	2.947	3.733	4.073	15
D	16	1.337	1.746	2.120	2.583	2.921	3.686	4.015	16
O	17	1.333	1.740	2.110	2.567	2.898	3.646	3.965	17
M	18	1.330	1.734	2.101	2.552	2.878	3.610	3.922	18
	19	1.328	1.729	2.093	2.539	2.861	3.579	3.883	19
	20	1.325	1.725	2.086	2.528	2.845	3.552	3.850	20
	21	1.323	1.721	2.080	2.518	2.831	3.527	3.819	21
	22	1.321	1.717	2.074	2.508	2.819	3.505	3.792	22
	23	1.319	1.714	2.069	2.500	2.807	3.485	3.768	23
	24	1.318	1.711	2.064	2.492	2.797	3.467	3.745	24
	25	1.316	1.708	2.060	2.485	2.787	3.450	3.725	25
	26	1.315	1.706	2.056	2.479	2.779	3.435	3.707	26
	27	1.314	1.703	2.052	2.473	2.771	3.421	3.690	27
	28	1.313	1.701	2.048	2.467	2.763	3.408	3.674	28
	29	1.311	1.699	2.045	2.462	2.756	3.396	3.659	29
	30	1.310	1.697	2.042	2.457	2.750	3.385	3.646	30
	32	1.309	1.694	2.037	2.449	2.738	3.365	3.622	32
	34	1.307	1.691	2.032	2.441	2.728	3.348	3.601	34
	36	1.306	1.688	2.028	2.434	2.719	3.333	3.582	36
	38	1.304	1.686	2.024	2.429	2.712	3.319	3.566	38
	40	1.303	1.684	2.021	2.423	2.704	3.307	3.551	40
	42	1.302	1.682	2.018	2.418	2.698	3.296	3.538	42
	44	1.301	1.680	2.015	2.414	2.692	3.286	3.526	44
	46	1.300	1.679	2.013	2.410	2.687	3.277	3.515	46
	48	1.299	1.677	2.011	2.407	2.682	3.269	3.505	48
	50	1.299	1.676	2.009	2.403	2.678	3.261	3.496	50
	55	1.297	1.673	2.004	2.396	2.668	3.245	3.476	55
	60	1.296	1.671	2.000	2.390	2.660	3.232	3.460	60
	65	1.295	1.669	1.997	2.385	2.654	3.220	3.447	65
	70	1.294	1.667	1.994	2.381	2.648	3.211	3.435	70
	80	1.292	1.664	1.990	2.374	2.639	3.195	3.416	80
	100	1.290	1.660	1.984	2.364	2.626	3.174	3.390	100
	150	1.287	1.655	1.976	2.351	2.609	3.145	3.357	150
	200	1.286	1.653	1.972	2.345	2.601	3.131	3.340	200

Two Tails	0.20	0.10	0.05	0.02	0.01	0.002	0.001
One Tail	0.10	0.05	0.025	0.01	0.005	0.001	0.0005

Source: <http://www.math.unb.ca/~knight/utility/t-table.htm>, University of New Brunswick

Bibliography on Forced Migration in Afghanistan/Pakistan

INTERNALLY DISPLACED PERSONS - GENERAL

Office of the Commissioner for Humanitarian Assistance. *Guiding Principles on Internal Displacement* (Geneva: UN, 2004). [Accessed online at: <http://www.unhcr.org/43ce1cff2.html>]

UNHCR. *Global Report 2010 – Working with the Internally Displaced* (Geneva: 2010). [Accessed online at: <http://www.unhcr.org/4dfdbf380.html>]

UNHCR. *Handbook for the Protection of Internally Displaced Persons* (Geneva: 2010). [Accessed online at: <http://www.unhcr.org/4c2355229.html>]

IDPs IN AFGHANISTAN

Farr, Grant. “Afghanistan: Displaced in a Devastated Country” in Vincet, Mark & Sorenson, Birgitte Refslund (eds.). *Caught Between Borders: Response Strategies of the Internally Displaced* (London: Pluto Press, 2001) at 117-137.

Qadeem, Mossarat. “Afghanistan: The Long Way Home” in Banerjee, Paula et. al. (eds.). *Internal Displacement in South Asia: The Relevance of the UN’s Guiding Principles*. (New Delhi: Sage Publications, 2005) at 30-61.

Schmeidl, Susanne. “Between and Rock and a Hard Place: The Return of Internally Displaced Persons to Northern Afghanistan,” (Brookings-Bern Project on Internal Displacement, July 1, 2009). [Accessed online at http://www.brookings.edu/opinions/2009/0601_afghanistan_mundt.aspx].

Schmeidl, Susanne. “The Failure to Protect: Battle-Affected IDPs in Southern Afghanistan” (Brookings-Bern Project on Internal Displacement, June 22, 2009). [Accessed online at http://www.brookings.edu/opinions/2009/0622_afghanistan_mundt.aspx]

DISASTER-AFFECTED IDPs - PAKISTAN

Andrew Wilder, *Perceptions of the Pakistan Earthquake Response: Humanitarian Agenda 2015 Pakistan Country Study* (Medford, MA: Feinstein International Center, February 2008). [Accessed online at: <https://wikis.uit.tufts.edu/confluence/download/attachments/14553670/HA2015+Pakistan+Earthquake+Response.pdf?version=1&modificationDate=1218657045000>]

Brookings-Bern Project on Internal Displacement. *Protecting Persons Affected by Natural Disasters: IASC Operational Guidelines on Human Rights and Natural Disasters* (Washington, DC: 2006).

Earthquake Rehabilitation and Reconstruction Authority (ERRA), Government of Pakistan. *Social Impact Assessment: 8 October 2005 Pakistan Earthquake* (Islamabad: May 2007).

Internal Displacement Monitoring Centre. *Briefing Paper on Flood-Displaced Women in Sindh Province, Pakistan* (Oslo, 2011). [Accessed online at: <http://www.internal-displacement.org/briefing/pakistan>]

Walter Kälin, “Natural Disasters and IDPs’ Rights,” (July, 2005) *Forced Migration Review* at 10-11. [Accessed online at: <http://www.fmreview.org/FMRpdfs/Tsunami/full.pdf>]

CONFLICT-AFFECTED IDPs - PAKISTAN

Amnesty International. *As If Hell Fell On Me: The Human Rights Crisis in Northwest Pakistan* (London: April 2010). [Accessed online at: <http://www.amnesty.org/en/library/asset/ASA33/004/2010/en/1ea0b9e0-c79d-4f0f-a43d-98f7739ea92e/asa330042010en.pdf>]

Bruijn, Bart de. *Millenium Development Goals Analysis of Peoples to the Concern of UNHCR: The Living Conditions of Refugees, Asylum Seekers and IDP's in Armenia, Ecuador, Sri Lanka and Pakistan* (The Hague: Netherlands Interdisciplinary Demographic Institute, October 2006). [Accessed online at: <http://www.unhcr.org/cgi-bin/texis/vtx/home/opendocPDFViewer.html?docid=45adedde2&query=pakistan%20IDP>]

Human Rights Commission of Pakistan. *Internal Displacement in Pakistan: Contemporary Challenges* (Lahore: October, 2010). [Accessed online at: <http://www.hrcp-web.org/pdf/Internal%20Displacement%20in%20Pakistan.pdf>]

International Crisis Group. *Pakistan: The Worsening IDP Crisis* (Islamabad/Brussels: 16 September 2010). [Accessed online at: [http://www.crisisgroup.org/~media/Files/asia/south-asia/pakistan/B111%20Pakistan%20-%20The%20Worsening%20IDP%20Crisis.pdf](http://www.crisisgroup.org/~/media/Files/asia/south-asia/pakistan/B111%20Pakistan%20-%20The%20Worsening%20IDP%20Crisis.pdf)]

Najam-ud-Din. *Internal Displacement in Pakistan: Contemporary Challenges* (Lahore: Human Rights Commission of Pakistan, October 2010). [Accessed online at: <http://www.hrcp-web.org/pdf/Internal%20Displacement%20in%20Pakistan.pdf>]

United Nations & Government of North West Frontier Province. *Early Recovery Priority Interventions in NWFP/FATA: A Report of Evidence-Based Needs and Priority Interventions* (Peshawar: October 2009). [Accessed online at: http://www.pdma.gov.pk/documents/Final_CERINA_23_Feb2010.pdf]

Winthrop, Rebecca. *Pakistan's Displaced Girls and Women: An Opportunity for Education* (Washington, D.C.: Brookings-Bern Project on Internal Displacement, June 2009). [Accessed online at: http://www.brookings.edu/opinions/2009/0611_pakistan_education_winthrop.aspx]

World Bank & Asian Development Bank. *North West Frontier Province and Federally Administered Tribal Areas: Preliminary Damage and Needs Assessment - Immediate Restoration and Medium Term Reconstruction in Crisis Affected Areas* (Manila: November 2009). [Accessed online at: <http://www.adb.org/Documents/Reports/PAK-FATA-Damage-Needs-Assesment-Nov-2009.pdf>]

AFGHAN MIGRATION - GENERAL

Afghanistan Research & Evaluation Unit. *Report of Conference on Afghan Population Movements* (Kabul: April 27, 2005).

Afghanistan Research & Evaluation Unit. *Report of Conference on Afghan Population Movements* (Islamabad: February 14, 2006).

Amnesty International. *Afghanistan: Out of Sight, Out of Mind: The Fate of the Afghan Returnees* (London: 2003).

Barakat, Sultan. & Wardell, Gareth. "Exploited By Whom? An Alternative Perspective on Humanitarian Assistance To Afghan Women," (Oct., 2002) *Third World Quarterly* 23(5), Reconstructing War-Torn Societies: Afghanistan at 909-930.

Berg, Michelle. "Peace Will Come with Pens, Not Swords: Afghan Refugees and Returnees Seek Security, Shelter, Education and Opportunities," (November 2003) *Refugee Reports* 24(8) at 10-13.

Boesen, I. W. "What Happens to Honour in Exile? Continuity and Change Among Afghans Refugees," in Huld, B. & Jansson, E. (eds.). *The Tragedy of Afghanistan: The Social, Cultural, and Political*

Impact of the Soviet Invasion (London: Croom Helm, 1998) at 219-39.

Bose, Ashish. "Afghan Refugees in India," (Oct. 23-29, 2004) *Economic And Political Weekly* 39(43) at 4698-4701.

Couldry, Marion & Morris, Tim (eds.). "September 11th – Has Anything Changed?" (June 2002) *Forced Migration Review* Vol. 13.

Dupree, Louis. "Settlement and Migration Patterns in Afghanistan: A Tentative Statement," (1975) *Modern Asian Studies* 9(3) at 397-413.

Edwards, David Busby. "Marginality and Migration: Cultural Dimension of the Afghan Refugee Problem," (Summer, 1986) *International Migration Review* 20(2), Refugees: Issues and Directions at 313-325.

Emery, Margaret & Ruiz, Hiram. "Descent Into Disaster? Afghan Refugees," (Winter, 2001) *Middle East Report* 221 at 16-17.

Emmott, Sue. "'Dislocation', Shelter, and Crisis: Afghanistan's Refugees and Notions of Home," (Feb., 1996) *Gender and Development* 4(1), Urban Development at 31-38.

Ghobadi, Negar et. al. *Moving Out of Poverty: Migration Insights from Rural Afghanistan* (Kabul: Afghan Research and Evaluation Unit, August 2005).

Gee, Richard. *Refugee Enfranchisement in the 2004 Afghan Elections: A Strategy for Refugee Participation in the Future of Afghanistan*. Masters Project (Durham, N.C.: Duke University, Spring 2004).

Helton, Arthur C. "Rescuing the Refugees," (Mar.-Apr., 2002) *Foreign Affairs* 81(2) at 71-82.

Human Rights Watch. *Closed Door Policy: Afghan Refugees in Pakistan & Iran* (New York: 2002).

Hyman, Anthony. "The Afghan Politics of Exile," (Jan., 1987) *Third World Quarterly* 9(1) at 67-84.

International Crisis Group. *Afghanistan: What Now for Refugees* (Brussels: August 31, 2009).

IOM. *Trafficking in Persons: An Analysis of Afghanistan* (Kabul: 2003).

Jazayery, Leila. "The Migration-Development Nexus: Afghanistan Case Study," (2002) *International Migration* 40 (5), Special Issue 2 at 231-254.

Khoser, Khalid & Schmeidl, Susanne. *Displacement, Human Development & Security in Afghanistan*. (unpublished)

Lischer, Sarah Keynon. "Afghan Refugees: Three Decades of Political Violence," in Lischer, Sarah Keynon (ed.). *Dangerous Sanctuaries: Refugee Camps, Civil War, and the Dilemmas of Humanitarian Aid*. (Ithaca: Cornell University Press, 2005) at 44-72.

Long, Kate. *Voting With Their Feet: A Review of Refugee Participation and the Role of UNHCR in Country of Origin Elections and Other Political Processes*. (Geneva: UNHCR, September, 2010).

Macleod, Ewan. "Afghan Refugees in Iran and Pakistan," in Loesher, Gil et al. (eds.). *Protracted Refugee Situations: Political, Human Rights & Security Implications* (New York: United Nations University Press, 2008) at 333-350.

Majrooh, Sayd Bahaouddin. "Afghan Intellectuals in Exile: Philosophical and Psychological Dimensions," in Anderson, Ewan & Dupree, Nancy Hatch (eds.). *The Cultural Basis of Afghan Nationalism* (London: Pinter Publishers, 1990) at 71-83.

Malik, Salma. "Impact of International Jurisdiction on Afghan Refugee Rights," (August 2004) *Refugee Watch* Vol. 22 [Accessed online at <http://www.mcrg.ac.in/cata.htm>].

- Margesson, Rhoda. *Afghan Refugees: Current Status and Future Prospects* (Washington: Congressional Research Service Report for Congress, January 2007).
- Marsden, Peter. "Afghanistan's Complex Forced Migration" in Van Hear, Nicholas & McDowell, Christopher (eds.). *Catching Fire: Containing Forced Migration in a Volatile World* (Oxford: Lexington Books, 2006) at 115-139.
- McSpadden, Lucia Ann. "Reviewed Work(s): Aging and Family in an Afghan Refugee Community: Transitions and Transformations by Patricia A. Omidian," (Winter, 1997) *International Migration Review* 31(4) at 1139-1140.
- Monsutti, Alessandro. *War and Migration: Social Networks and Economic Strategies of the Hazaras of Afghanistan* (New York: Routledge, 2005).
- Monsutti, Alessandro. *Afghan Transnational Networks: Looking Beyond Repatriation* (Kabul: Afghan Research and Evaluation Unit, August 2006).
- Monsutti, Alessandro. "Afghan Migratory Strategies and The Three Solutions to the Refugee Problem," (2008) *Refugee Survey Quarterly* 27 at 58-73.
- Mougne, Christine. *Trees Only Move in the Wind: A Study of Unaccompanied Afghan Children in Europe* (Geneva: UNHCR, June 2010).
- Niland, Norah. "Taliban-Run Afghanistan: The Politics of Closed Borders and Protection" in Bayefsky, Anne F. (ed.). *Human Rights and Refugees, Internally Displaced Persons and Migrant Workers: Essays in Memory of Joan Fitzpatrick and Arthur Helton* (Leiden: Martinus Nijhoff Publishers, 2006) at 179-208.
- Opel, Aftab. *Bound for the City: A Study of Rural to Urban Labour Migration in Afghanistan* (Kabul: Afghan Research and Evaluation Unit, May 2005).
- Poppelwell, Teresa. *FMO Research Guide: Afghanistan*. (Forced Migration Online, July 2007).
- Rizwi, Gowher. "The Afghan Refugees: Hostages in the Struggle for Power," (1990) *Journal of Refugee Studies*, Vol. 3(3) at 245 - 261.
- Roy, Arpita Basu. "Afghanistan: Seven Phases of War and the Exodus," (June 1999) *Refugee Watch* 5 & 6 [Accessed online at <http://www.mcrg.ac.in/cata.htm>].
- Roy, Arpita Basu. "Taliban: Shelter Seekers or Refugee Warriors?" (June 1999) *Refugee Watch* 5 & 6 [Accessed online at <http://www.mcrg.ac.in/cata.htm>].
- Ruiz, Hiram. "Afghanistan: Conflict and Displacement 1978 to 2001," (2004) *Forced Migration Review* 13 at 8-10.
- Saito, Mamiko. *Second-Generation Afghans in Neighbouring Countries, From Mohajer to Hamwatan: Afghans Return Home, Case Study Series* (Kabul: Afghan Research and Evaluation Unit, December 2007).
- Saito, Mamiko. *Searching for my Homeland: Dilemmas Between Borders — Experiences of Young Afghans Returning "Home" from Pakistan and Iran, Synthesis Paper Series* (Kabul: Afghan Research and Evaluation Unit, July 2009).
- Schmeidl, Susanne. "Ethnic Dynamics in Forced Migration: A Comparison of Afghan and Palestinian Refugees," (1994) *International Journal of Group Tensions* 24(4) at 333-360.
- Schmeidl, Susanne. "(Human) Security Dilemmas: Long Term Implications of the Afghan Refugee Crisis," (Feb., 2002) *Third World Quarterly* 23(1) at 7-29.
- Schmeidl, Susanne & Maley, William. "The Case of the Afghan Refugee Population: Finding Durable

Solutions in Contested Transitions”, in Adelman, Howard (ed.) *Protracted Displacement in Asia: No Place to Call Home* (Hampshire: Ashgate Publishing Ltd., 2008) at 131-179.

Schöch, R. “UNHCR and the Afghan Refugees in the Early 1980s: Between Humanitarian Action and Cold War Politics,” (2008) *Refugee Survey Quarterly* 27 at 45-57.

Sharani, M. Nazif. “Afghanistan’s Muhajirin (Muslim “Refugee Warriors”): Politics of Mistrust and Distrust of Politics,” in Daniel, E. Valentine & Knudsen, John Chr. (eds.). *Mistrusting Refugees* (Berkeley: University of California Press, 1995) at 187-206.

Sherjai, Ajmal. “Displacement in Afghanistan” in Mishra, Omprakash (ed.). *Forced Migration in the South Asian Region: Displacement, Human Rights & Conflict Resolution* (Delhi: Manak Publications Pvt. Ltd, 2004) at 236-245.

Stigter, Elca & Monsutti, Alessandro. *Transnational Networks: Recognising a Regional Reality*, Briefing Paper (Kabul: Afghan Research and Evaluation Unit, April 2005).

Tapper, Richard. “Who Are the Kuchi? Nomad Self-Identities in Afghanistan,” (March, 2008) *The Journal of the Royal Anthropological Institute* 14(1) at 97-116.

Van Selm, Joanne. “Foreign Policy Considerations in Dealing with Afghanistan’s Refugees: When Security and Protection Collide,” (December, 1998) *Forced Migration Review* 13 at 16-18.

Weinbaum, Marvin G. “The Politics of Afghan Resettlement and Rehabilitation,” (Mar., 1989) *Asian Survey* 29(3) at 287-307.

AFGHANS IN PAKISTAN

Reproductive Health for Refugees Consortium, *If Not Now, When? Addressing Gender-Based Violence in Refugee, Internally-Displaced, and Post-Conflict Settings – A Global Overview* (New York, 2002) [Accessed online at: <http://helid.digicollection.org/en/d/Js2666e/6.1.html>].

Abrar, C. R. “The South Asia Declaration on Refugees,” in *Udbastu: The Uprooted, A Newsletter on Refugee and Migratory Movements*, Issue No. 27 (Dhaka: Refugee and Migratory Movements Research Unit, January to March 2004).

Allan, Nigel J.R. “Impact of Afghan Refugees on the Vegetation Resources of Pakistan’s Hindukush-Himalaya,” (Aug., 1987) *Mountain Research and Development* 7(3) at 200-204.

Author Unknown, *Afghan Refugees – Citizenship and Naturalization in Pakistan: A Legal Analysis* (Islamabad: undated).

Author Unknown. “Afghan Refugees,” *Fact File No. 21* (Islamabad: Islamabad Policy Research Institute, April 15, 2001). [Accessed online at <http://ipripak.org/factfiles/ff21.shtml>]

Afghan Research and Evaluation Unit. “Feature: The Registration of Afghans in Pakistan,” (July 2007) *Afghanistan Research Newsletter* 14 at 2-3.

Altai Consulting. *Study on Cross Border Population Movements with Pakistan*. (Kabul: 2009).

Benard, Cheryl. “Politics and the Refugee Experience,” (1986) *Political Science Quarterly* 101(4) at 617-636.

Centlivres, Pierre & Centlivres-Demont, Micheline. “Sociopolitical Adjustment Among Afghan Refugees in Pakistan,” (1987) *Migration World* 15(4).

Centlivres, Pierre & Centlivres-Demont, Micheline. “The Afghan Refugee in Pakistan: An Ambiguous Identity,” (1998) *Journal of Refugee Studies* 1(2) at 141-152.

Centlivres, Pierre & Centlivres-Demont, Micheline. “The Afghan Refugees in Pakistan: A Nation in Exile,” (1998) *The Sociology of Involuntary Migration* 36(2).

- Cheema, Iqbal Pervaiz. "Pakistan and The Afghan Refugees," *The News* (Islamabad: Oct 1, 2007). [Accessed online at <http://ipripak.org/articles/newspapers/pakistanafghanistan.shtml>]
- Chief Commissioner for Afghan Refugees, States & Frontier Regions Division, Government of Pakistan. *Official Handbook on Refugee Management in Pakistan* (Islamabad: 1981).
- Chief Commissioner for Afghan Refugees, States & Frontier Regions Division, Government of Pakistan. *Handbook on Management of Afghan Refugees in Pakistan*, Revised Edition (Islamabad: 1984).
- Collective for Social Science Research. *Afghans in Karachi: Migration, Settlement & Social Networks* (Karachi: March 2005).
- Collective for Social Science Research. *Afghans in Peshawar: Migration, Settlements and Social Networks* (Karachi: January 2006).
- Collective for Social Science Research. *Afghans in Pakistan: Broadening the Focus* (Karachi: January 2006)
- Collective for Social Science Research. *Afghans in Quetta: Settlements, Livelihoods, Support Networks and Cross-Border Linkages* (Karachi: January 2006)
- Connor, Kerry M. "Factors in the Residential Choices of Self-Settled of Afghan Refugees in Peshawar, Pakistan," (Winter, 1989) *International Migration Review* 23(4) at 904-932.
- Dupree, Hatch Nancy. "Demographic Report on Afghan Refugees in Pakistan," (1988) *Modern Asian Studies* 22(4) at 845-865.
- Enterprise for Business & Development Management. *Baseline Study On Illegal Migration, Human Smuggling And Trafficking In Pakistan* (Peshawar: June 2009).
- Farr, Grant. "Afghan Refugees in Pakistan: Definitions, Repatriation and Ethnicity," in Anderson, Ewan & Dupree, Nancy Hatch (eds.). *The Cultural Basis of Afghan Nationalism* (London: Pinter Publishers, 1990) at 134-143.
- Fielden, Matthew B. "The Geopolitics of Aid: The Provision and Termination of Aid to Afghan Refugees in North West Frontier Province, Pakistan," (1998) *Political Geography* 17(4) at 459-487.
- Ghufran, Nasreen. *Afghan Refugee Children in Pakistan*, Paper Presented at the Conference of Scholars and Other Professionals Working on Refugees and Displaced Persons in South Asia, held in Rajendrapur, Bangladesh on February 9-11, 1998 (Colombo: Regional Centre for Strategic Studies, 1998).
- Ghufran, Nasreen. *Afghan Refugee Women in Pakistan* (Peshawar: Hanns Seidel Foundation & Department of International Relations, Peshawar University, 1999).
- Ghufran, Nasreen. "Afghan Refugees in Pakistan: Current Situation and Future Scenario," (July-December, 2006) *Policy Perspectives* 3(2) (Islamabad: Institute of Policy Studies).
- Ghufran, Nasreen. "Afghans in Pakistan: A Protracted Refugee Situation," (April 2008) *Policy Perspectives* 5(2) (Islamabad: Institute of Policy Studies).
- Ghufran, Nasreen. *Politics of Afghan Refugee Repatriation from Pakistan: Complications and Challenges* (Saarbrücken: VDM Verlag Dr. Müller, 2010).
- Grare, Frederic. "The Geopolitics of Afghan Refugees in Pakistan," Stedman, Stephen John & Tanner, Fred (eds.). *Refugee Manipulation: War, Politics and the Abuse of Human Suffering* (Washington: The Brookings Institution, 2003) at 57-94.
- Habibi, Gulbadan & Hunte, Pamela. *Afghan Returnees from NWFP, Pakistan to Nangarhar Province*

(Kabul: Afghanistan Research and Evaluation Unit, March 2006).

Hatch-Dupree, Nancy. "A Socio-Cultural Dimension: Afghan Women Refugees in Pakistan," in Anderson, Ewan & Dupree, Nancy Hatch (eds.). *The Cultural Basis of Afghan Nationalism* (London: Pinter Publishers, 1990) at 121-133.

Human Rights Commission of Pakistan. *Afghan Refugees in Pakistan: Push Comes to Shove* (Lahore: April 2009).

Human Rights Commission of Pakistan. *Human Trafficking Through Quetta* (Lahore: 2009).

Hussain, Ijaz. "Current Issues In The Afghan Refugee Debate," (November, 2002) *Islamabad Policy Research Institute Journal* 2(1). [Accessed online at <http://ipripak.org/journal/winter2002/currentissuesinafgha.shtml>]

Insan Foundation Trust. *Footsteps: A Study on Afghan Refugee Children in Pakistan* (Rawalpindi: 2006).

Institute of Policy Studies. *Afghans in Pakistan: Plight, Predicament and the Way Forward*. (Islamabad: 2009).

Institute of Policy Studies & UNHCR. *International Refugee Law, Islamic Shariah and Afghan Refugees in Pakistan* (Islamabad: 2010).

International Medical Corps. "Gender-Based Violence Among Afghan Refugees: Summary of Post-Intervention Survey Findings in Three Camps in North-West Frontier Province, Pakistan," in *Knowledge, Attitudes and Exposure to GBV Among Afghan Refugees in Pakistan, Final Report* (Santa Monica, USA: December, 2009).

Islam, Muhammad. "A Socio-Political Study of Afghan Refugees in Islamabad City" in Mishra, Omprakash (ed.). *Forced Migration in the South Asian Region: Displacement, Human Rights & Conflict Resolution* (Delhi: Manak Publications Pvt. Ltd, 2004) at 246-263.

Kassam, Azaad & Nanji, Anar. "Mental Health of Afghan Refugees in Pakistan: A Qualitative Rapid Reconnaissance Field Study" (2006) *Intervention* 0(0) at 1-9.

Khan, Ayesha. "Afghan Refugee Women's Experience of Conflicts and Disintegration," (2002) *Meridians* 3(1) at 89-121.

Khattak, Saba Gul. "Violence and Home: Afghan Women's Experience of Displacement" in Behera, Navnita Chadha (ed.). *Gender, Conflict and Migration* (New Delhi: Sage Publications, 2006) at 116-136.

Khattak, Saba Gul. "Living on the Edges: Afghan Women and Refugee Camp Management in Pakistan," (Spring, 2007) *Signs* 32(3) at 575-579.

Khattak, Saba Gul. "Insecurity/Security Afghan Refugees and Politics in Pakistan" in *State, Society and Displaced People in South Asia* (Dhaka: The University Press Ltd, 2004) at 327-343.

Khattak, Saba Gul et. al. "Afghan Refugees: Policy Dilemmas and Solutions," in *SDPI Research and News Bulletin* 13(4 & 5) (Islamabad: Sustainable Development Policy Institute, July – October, 2006).

Kronenfeld, D. A. "Afghan Refugees in Pakistan: Not All Refugees, Not Always in Pakistan, Not Necessarily Afghan?" (2008) *Journal of Refugee Studies* 21 at 1–21.

Kundi, Dr. Akbar Mansoor. "Pakistan An Immigrant Country: Afghan And Its Impact On Baluchistan," (Summer 2005) *The Institute Of Strategic Studies Journal* 25 (2). [Accessed online at http://www.issi.org.pk/ss_Detail.php?dataId=344]

Langenkamp, Daniel. "The Victory of Expediency - Afghan Refugees and Pakistan in the 1990s,"

(2003) *Fletcher Forum of World Affairs* 229 27(2).

Latendrese, Simon. "The Border as Strategy: The Case of Afghans in Peshawar," in *Policy Perspectives* 4(1) (Islamabad: Institute of Policy Studies, January, 2007).

Lawrence, C. *Status of Afghan Women and Children Refugees*, Report to Women's Commission for Refugee Women and Children (Peshawar: International Rescue Committee, 1989).

Malik, Salma. "Refugee Rights Under International Jurisdiction: A Case Study Of Afghan Refugees," (Winter, 2005) *Islamabad Policy Research Institute Journal* 5(1). [Accessed online at <http://ipripak.org/journal/winter2005/refugee.shtml>]

Malik, Abdul Hamid. "Impelled Afghan Migration To Pakistan 1978-84," (Summer 2001) *The Institute Of Strategic Studies* 21(2). [Accessed online at http://www.issi.org.pk/ss_Detail.php?dataId=172]

Marsden, Peter. "Afghans in Pakistan: Why Rations Decline," (1992) *Journal of Refugee Studies* 5(3/4) at 289-299.

Ministry of States & Frontier Regions, Government of Pakistan, National Database & Registration Authority (NADRA) & UNHCR. *Registration Of Afghans In Pakistan 2007* (Islamabad: 2007).

Ministry of States and Frontier Regions, Government of Pakistan. *Management and Repatriation Strategy for Afghan Refugees in Pakistan for the Years 2010-2012*. (Islamabad, 2009).

Oberoi, Pia. "Afghan Refugees in Pakistan," in Oberoi, Pia. *Exile and Belonging: Refugees and State Policy in South Asia*. (New Delhi: Oxford University Press, 2006) at 136-169.

Parker, Elizabeth. "The Refugee Problem: Looking Towards Afghanistan's Long-Term Stability," (December 10, 2008) *South Asia Monitor* 125.

Phillips, Melissa. *The Role and Impact of Humanitarian Assets in Refugee-Hosting Countries*, New Issues in Refugee Research, Working Paper 84 (Geneva: UNHCR, March 2003).

Pstrusinska, Jadwiga. "Review Article: 'Refugees and Development' by Ernest E. Boesch & 'Tradition and Dynamism Among Afghan Refugees: A Report on Income-Generating Activities for Afghan Refugees in Pakistan' by Armin M.F. Goldschmidt," *Third World Quarterly* (Oct., 1986) 8(4) at 1451-1452.

Rogers, Tom. "Two Dimensions of a National Crisis: Population Growth and Refugees in Pakistan," (Oct., 1992) *Modern Asian Studies* 26(4) at 735-761.

Ruiz, Hiram A. "Afghan Refugees in Pakistan at Risk," (December 2001) *Refugee Watch* 16 [Accessed online at <http://www.mcrg.ac.in/cata.htm>].

Saito, Mamiko & Hunte, Pamela. *To Return or to Remain: The Dilemma of Second-Generation Afghans in Pakistan* (Kabul: Afghanistan Research and Evaluation Unit, June 2007).

Schoch, Rudiger. *Afghan Refugees in Pakistan During the 1980s: Cold War Politics and Registration Practice*. New Issues in Refugee Research, Research Paper No. 157 (Geneva: UNHCR, June 2008).

Shah, Babar. "Pakistan's Afghanistan Policy: An Evaluation," (Spring & Summer 2000) *The Institute Of Strategic Studies Islamabad Journal* 20(2)&(3). [Accessed online at http://www.issi.org.pk/ss_Detail.php?dataId=56]

Sheikh, Atta-ur-Rehman. "Towards a Legal Regime for Refugee Protection in Pakistan," (April 2003) *Refugee Watch* 18 [Accessed online at <http://www.mcrg.ac.in/cata.htm>].

Sheikh, Atta-ur-Rehman. "Developing a Legal Regime for Refugees in Pakistan," (December 2005) *Refugee Watch* 24, 25 & 26 [Accessed online at <http://www.mcrg.ac.in/cata.htm>].

Sustainable Development Policy Institute (Pakistan). *Report on Stakeholder Consultations on the*

Future of Afghan Refugees in Pakistan (Islamabad: September 2006).

Theodoropoulos, Christos. "The Development of Refugee Law in Pakistan," (January, 2003) *International Human Rights Perspective* 2(1) at 43.

UNHCR, Global Appeal 2010-2011 (Pakistan).

UNHCR. *Searching For Solutions: 25 Years of UNHCR – Pakistan Cooperation on Afghan Refugees* (Islamabad: June, 2005).

UNHCR. *Voluntary Repatriation from Pakistan* (Islamabad: 2005).

Zieck, Marjoleine. "The Legal Status of Afghan Refugees in Pakistan, a Story of Eight Agreements and Two Suppressed Premises," (2008) *International Journal of Refugee Law* 20(2) at 253-272.

Zieck, Marjoleine. "In the Aftermath of Voluntary Repatriation: The Fate of Afghan Refugees in Pakistan," (November 2007) *Pakistan Law Review* 3 at 1-51.

Zieck, Marjoleine. "Accession of Pakistan to the 1951 Convention and 1967 Protocol Relating to the Status of Refugees: 'Signing On Could Make All the Difference'," (March 2005) *Pakistan Law Review* 5 at 1-54.

The PDF version of
'Researching Forced Migration in Pakistan'
can be downloaded from
www.camp.org.pk

Community Appraisal & Motivation Programme (CAMP)
is a registered Pakistani NGO dedicated to working on peace and security,
sustainable development, human rights and emergency response at the
national and regional level. CAMP's activities range from research and
analysis, policy advocacy to implementation of projects.

