OPHI
OXFORD POVERTY & HUMAN DEVELOPMENT INITIATIVE
www.ophi.org.uk

UNIVERSITY OF OXFORD

Summer School on Multidimensional Poverty Analysis

Oxford Poverty & Human Development Initiative, (OPHI), University of Oxford

3–15 August 2015, Georgetown University Washington DC

OPHI

OXFORD POVERTY & HUMAN DEVELOPMENT INITIATIVE

www.ophi.org.uk

UNIVERSITY OF OXFORD

Data Issues in Multidimensional Poverty Measurement

Adriana Conconi & Ana Vaz

OPHI

Outline

1. Sources of multidimensional data
2. Household surveys
3. Indicators' design
4. Applicable population
5. Combined measures
6. Handling missing values

1. Sources of Multidimensional Data

Census

- Advantages:
  - information with negligible sampling error;
  - highly disaggregated levels.

- Disadvantages:
  - have low frequency;
  - offer information on a small set of indicators;
  - micro data may not be available to researchers.

1. Sources of Multidimensional Data

Administrative Data

- Advantages:
  - cover virtually all population and in a continuous form;
  - no data collection costs; and
  - data for individuals who might not respond to surveys.
- Disadvantages:
  - information is limited and may not match the research purpose;
  - any changes in data collection procedures or definitions may affect comparability over time;
  - serious data quality issues may compromise accuracy;
  - metadata is usually not available;
  - access to administrative (micro) data varies by country; and
  - linking data sources is rarely straightforward.

1. Sources of Multidimensional Data

Household Surveys

- Most commonly used data source to study poverty;
- Collect information on a diverse set of topics on a sample representative of the population of interest;
- Areas for improvement:
  - Frequency;
  - Coverage;
  - Dimensional coverage.

2. Household Surveys: Metadata

- Metadata is "data about the data".
- Provides information about the survey sample design, fieldwork activities, questionnaires, structure of the dataset, definitions, coding, etc.

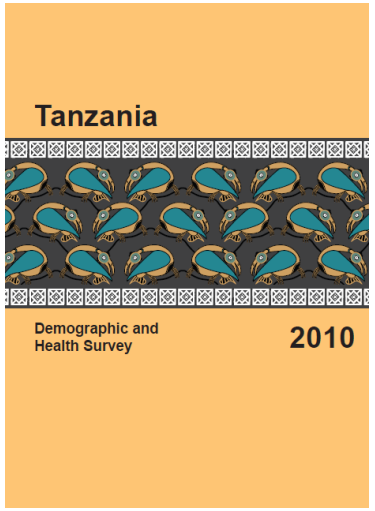
How to use the sample weights?

Who are eligible?

How to interpret the coding?

DHS Country Report

http://www.measuredhs.com/publications/publications-by-type.cfm



CONTENTS

Page

TABLES AND FIGURES ix
FOREWORD xvii
SUMMARY OF FINDINGS xix
MAP OF TANZANIA xvi

CHAPTER 1 INTRODUCTION

1.1 Geography, History, and the Economy 1
1.2 Population 2
1.3 Population, Family Planning, and HIV Policies and Programmes 2
1.4 Objectives and Organisation of the Survey 5

CHAPTER 2 HOUSEHOLD POPULATION AND HOUSING CHARACTERISTICS

2.1 Population by Age and Sex 11
2.2 Household Composition 12
2.3 Children's Living Arrangements and Parental Survival 13
2.4 Education of the Household Population 15
2.4.1 Educational Attainment 15
2.4.2 School Attendance Rates 18
2.5 Household Environment 21
2.5.1 Drinking Water 21
2.5.2 Household Sanitation Facilities 23
2.5.3 Housing Characteristics 24
2.5.4 Household Possessions 25
2.6 Wealth Index 26
2.7 Birth Registration 27
2.8 Household Food Security 29

CHAPTER 3 CHARACTERISTICS OF RESPONDENTS

3.1 Characteristics of Survey Respondents 31
3.2 Education 33
3.2.1 Educational Attainment 33
3.2.2 Literacy 36
3.3 Access to Mass Media 38
3.4 Employment 41
3.4.1 Employment Status 41
3.4.2 Occupation 44

Sampling design and representativeness

Tables and results

Large Survey projects provide plenty of metadata

DHS General Data Manuals: available online

http://www.measuredhs.com/data/Data-Tools-and-Manuals.cfm

Guide to DHS Statistics

Reference to help users who work with DHS survey indicators and datasets to better understand indicator definitions and the calculations used to generate the survey results.

DHS Recode Manual

Describes each data file and the variables contained in them.

DHS Data Editing and Imputation

Presents the methodology used by DHS for the production of edited data files. The paper focuses primarily on the editing of dates of events, and the imputation of incomplete dates.

MICS Metadata: available online

http://www.childinfo.org

Country reports

It includes comprehensive survey results and country survey specificities.

Questionnaires

Flow of questionnaire modules, Household questionnaire, Women's questionnaire, Children under-5 questionnaire, Additional situation specific modules, Optional modules

MICS Manual

Other various documents

ISample Size (Households) Calculation Template, Pictorials for Water and Sanitation Facilities, One-page pictorial on cooking methods using solid fuels. Sample weight calculation.

2. Household Surveys: Survey Design

Usually household surveys follow a complex sampling design in two stages:

1. Clusters (e.g. PSU) are selected from within each strata (e.g. Region+urban/rural)

2. Households are selected from household listings within each cluster (listing are generated during census operations)

The result is a representative and yet efficient sample that reduces cost and increases quality of data collection

Sample weights

Weights are computed as the inverse probability of selection:

- probability of selecting the cluster;
- probability of selecting the household within the cluster;
- they may also adjust by response rate and/or by the demographic structure of the population (Yansanhe, 2005).

Ignoring the weights would produce significantly biased results

Samples and subsamples

Some data can be particularly more difficult and expensive to collect, either because it takes longer (e.g. revisits) or it requires enumerators with more expertise (hence supervision is more difficult).

Check the metadata for subsamples and how to undertake analysis with it!

What geographical level can you decompose?

Are all ethnic group represented well in the sample?

Survey representativeness depends on the sample design, and will limit how far one can undertake decomposition analysis.

For example in Tanzania:

"The 2010 TDHS sample was designed to provide estimates for the entire country, for urban and rural areas in the Mainland, and for Zanzibar.

For specific indicators such as contraceptive use the sample design allowed the estimation of indicators for each of the then 26 regions. To estimate geographic differentials for certain demographic indicators, the regions of mainland Tanzania were collapsed into seven geographic zones. Although these are not official administrative zones, this classification is used by the Reproductive and Child Health Section of the MoHSW. Zones were used in each geographic area in order to have a relatively large number of cases and a reduced sampling error."

3. Indicators' Design – Unit-Level Indicator Accuracy

- **Unit of identification:** entity who is identified as poor or non-poor – usually the individual or the household.

- HH surveys are usually designed to create indicators that are representative of achievements of some population subgroups.

Unit Level Indicator Accuracy

- Indicators collected with short reference periods and are judged to be accurate 'on average'. Examples:
  - consumption in the last seven days,
  - illness in the last two weeks, and
  - time use in the past 24 hours.

- However achievements may not be accurate at the individual level. And what if…
  - last seven days' consumption included a family wedding,
  - the respondent had a rare and brief bout of the flu,
  - the last 24 hours was a major public holiday.

Unit Level Indicator Accuracy

- Indicators used for targeting are always required to be accurate at the individual level.

- Multidimensional measures require the joint distribution of deprivations to be accurate on average.
  - Selected indicators ideally balance indicator precision and unit-level accuracy.

- When tracking changes over time in poverty, indicators should reflect individual achievement levels across the relevant period. No distortions due to seasonal effects, or short-term shocks.

Indicators Transformation to Match Unit of Identification

- Relevant data may be available at the individual, household, and community levels.
- So, we may need to transform indicators such that they reflect deprivations of the chosen unit of identification.
- Suppose a child poverty measure with children, household and village level data.
  - How do we construct the $n \times d$ achievement matrix?
  - What is the implicit assumption?

4. Applicable Population

- Applicable population: group of people for which the achievement is relevant; namely,
  - it can be measured – it is conceptually applicable – **and**
  - it has been effectively measured – data is available.

- Some achievements relevant for poverty measurement are either conceptually or empirically applicable only for certain population groups.

Examples?

4. Applicable Population

- The achievement may be conceptually relevant only for certain groups:
  - Income ,
  - Vaccinations,
  - Employment status.

- The achievements may be conceptually applicable to the whole population but data is only collected for some groups…
  - Anthropometric indicators

How to deal with this?

4. Applicable Population

- To restrict consideration to universally applicable achievements.
  - Narrows the set of indicators.

- To construct group-specific poverty measures.
  - Discriminating by groups may not eliminate applicability issues.
  - Not possible to track national poverty or target households.
  - May miss the overlaps of disadvantaged groups.

- To combine achievements that are not universally applicable (e.g. Global MPI).

5. Combined Measures

- Approach followed when constructing the MPI.

- Assumption of negative externalities. All household members are deprived:
  - If has at least one child or women undernourished;
  - If at least one child in the household died;
  - If at least a child of school age is not attending school.

- Assumption of positive externalities. All household members are considered non-deprived:
  - If at least one person has five years of schooling.

5. Combined Measures

- How to deal with households where not even one person qualifies for the achievement under consideration?
  - Drop these households from the sample.
    - ➢ That would bias the estimates…
  - Drop the indicator and re-weight the remaining indicators
    - ➢ That would violate dimensional breakdown…
  - Consider them as non-deprived (deprived) in that indicator
    - ➢ Need to scrutinize this assumption…

5. Combined Measures

- Suppose survey has not collected information from all applicable members. How to deal with households where there is no data for any member?

  - Consider them as non-deprived (deprived) in that indicator.

  - Considering them as non-deprived could be seen as a 'conservative' approach, and will lead to a `lower bound' poverty estimate.

Assessing Combined Measures

- Potential household composition effect.

- Inclusion of indicators referring specific groups can be made provided that:
  - Not all indicators refer to a particular specific group;
  - An important proportion of households have at least one member for whom the achievement is relevant;
  - Empirical test of the impact of the household composition.

Assessing Combined Measures

- Alkire and Santos (2014)

  - Tests of differences in means between MPI-poor and non-poor households in terms of size, number of children under 5, number of females, number of members 50 years or older, proportion of female-headed households, and proportion of school-aged children.

  - Decompose country's MPI by age and gender and compare the rankings, correlations and proportion of robust pairwise comparisons.

6. Handling missing values

Missing value: a variable that should have a response, but because of interview errors the question was not asked.

Inconsistent: This code is generally used by people in the secondary editing group, when a value or code is not plausible.

"Don't know" responses: These codes are normally pre-coded in the questionnaires, but they are consistently used throughout the recode file.

How should we treat missing values?

6. Handling missing values

Ways to deal with missing values:

1. Use rule to assign value for the missing data. E.g. Global MPI:
   - Household non-deprived if at least one member has 5+ years of education.
   - Household deprived if we have information for at least 2/3 of the household members and none of them has at least 5 years of education.

2. Drop the observation from the sample. E.g. Global MPI:
   - Household with missing information in any of the relevant indicators are dropped from the sample

How should we treat a missing value when computing the Adjusted Headcount Ratio?

Suppose the following matrix, and a poverty line of k>=1:

$$\begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 0 \\ . & 0 & 1 \\ 0 & 0 & 0 \\ . & . & 0 \\ 1 & 0 & 0 \end{bmatrix}$$

How do we compute the average deprivation with missing information?

Is this individual poor?

In practice we reduce the sample to only cases with information in all indicators, having a consequent "sample drop" due to missing information

Sample Drop and Bias Analysis

Problem: sample drop may lead to biased estimates.

Bias analysis: group with missing values is compared to rest, using the indicators for which values are present for both groups.

- Series of hypothesis test for difference of means or proportions.

Results:

- No significant differences: we can use the reduced sample.
- Significant differences: we can still use the reduced sample but should explicitly the direction of the bias.

Sample Drop and Bias Analysis

The sample drop may also…

Affect the representativeness of the sample.

- Need to check the proportion of missing values for each indicator and analyze the proportion of total sample drop.

Affect the population share when regions are decomposed.

- Need to check how the share of each region changes before and after sample drop – Is there a bias towards a particular region?

Sample Drop and Bias Analysis

- **What about using imputation?**
  - Estimate a model with the achievement as the dependent variable against a set of explanatory variables.
  - Use estimated parameters to predict achievements for cases with missing values.

- **Limitations**:
  - The estimated model needs to be accurate.
  - We would have to specify a model that could predict a vector of deprivations.
  - Cannot solve problem of non-applicable populations.

Need further research!


Tabita, Kenya
Rabiya, India
Stephanie, Madagascar
Agatha, Madagascar
Dalmia, Kenya
Ann-Sophia, Kenya
Valérie, Madagascar

Thank you!