SURVEY DATA
DHS & MICS

Aluísio J D Barros
International Center for Equity in Health
Federal University of Pelotas
Household surveys

- Traditional strategy to obtain population representative estimates
- Most countries carry
  - Surveys attached to population census
  - Surveys on income, living standards and employment
  - Surveys on health status and utilization of health services
  - And so and so on
- Representativeness is a strong trait
  - Nationally
  - Geographical regions
  - Provinces
  - Metropolitan areas
  - Rural area
Sample design

• Is the key to representativeness
• Each population group that can have separate estimates is called a “domain”
  • Households in each domain must be selected in a way to provide a representative estimate (back to this soon)

• Precision
  • Sample size calculations needed to guarantee minimum precision
  • Precision is defined as the width of the confidence interval
  • Maximum acceptable CI width (and budget) defines sample size
  • Each domain must meet the sample size requirements
How to select a sample?

- Easiest way = simple random sample
  - Names on paper mixed in a bag
  - In this case everybody has the same probability of selection
  - But usually there is no up-to-date complete list of households!

- To be called a SAMPLE
  - All units must have a known, non-zero probability of selection

- How to select households, then?
Multi-stage cluster samples

• Break down your problem into steps
  • The number of steps may vary
  • Often, two steps

• 1. Within the domain, first, select a cluster
  • Usually a census tract

• For each cluster selected
  • Enumerate all existing households

• 2. From the list
  • Select the desired number of households
Pros and cons

• Advantages
  • Drastically reduce the need to identify households
  • Limit the area to be covered by interviewers

• Disadvantages
  • More complicated to implement
  • Due to similarity within clusters
    • Less information = less precision = more variability
    • Increase the sample size compared to SRS
  • More complicated to analyse

• On balance, reduce the cost of the survey
Stratification

- Population groups have different sizes
  - Men and women are about the same
  - More adults than children or elders
  - More poor than rich
  - More urban than rural (or the opposite in some places)
- If subgroups are proportionally represented in the sample
  - We may end with very few rich households
  - Or very few elders
- Stratification
  - Disproportionate sampling of some groups
  - \[\rightarrow\] reduction in sample size while keeping precision overall
  - Need to know the population proportion of each group!
  - Imply in different weights for households
Analysis must take

• Sample design into account!

• Failing to take clusters into account
  • Will produce variances that are underestimated

• Failing to take stratification into account
  • Will produce variances that are overestimated

• Failing to take sample weights into account
  • Will produce biased point estimates
  • Impossible to predict in what direction
**DHS**

- Demographic Health Surveys
  - Funded by USAID
  - Started in 1985 in El Salvador

**El Salvador: Standard DHS, 1985**

<table>
<thead>
<tr>
<th>DHS Final Reports</th>
<th>FR15</th>
<th>El Salvador 1985 DHS Final Report (Spanish)</th>
<th>PDF, 1938K</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Country:</th>
<th>El Salvador</th>
</tr>
</thead>
<tbody>
<tr>
<td>Contract Phase:</td>
<td>DHS-I</td>
</tr>
<tr>
<td>Recode Structure:</td>
<td>DHS-I</td>
</tr>
<tr>
<td>Implementing:</td>
<td>Asociación Demográfica</td>
</tr>
<tr>
<td>Organization:</td>
<td>Salvadoreña</td>
</tr>
<tr>
<td>Fieldwork:</td>
<td>May 1985 - July 1985</td>
</tr>
<tr>
<td>Status:</td>
<td>Completed</td>
</tr>
</tbody>
</table>

**Respondents**

- Households: Sample Size: 4922
- Female: (All Women)
  - Age: 15 to 49
  - Sample Size: 5207
- Male: No male respondents
- Facilities: N/A

**Survey Characteristics**

- Micronutrients
- Social Marketing

**Survey Datasets**

- Data Available

**HIV Testing**

- Not Collected

**GPS Datasets**

- Not Collected

**SPA Datasets**

- Not Applicable
DHS phases

- DHS I – 1985-89
- DHS II – 1990-94
- DHS III – 1995-99
- DHS IV – 2000-04
- DHS V – 2005-09
- DHS VI – 2010-14

Most common topics in the surveys

- Anthropometry: 200
- HIV Knowledge: 182
- HIV Behavior: 149
- Maternal Mortality: 126
- Reproductive Calendar: 123
- Micronutrients: 121
- Tobacco Use: 103
- Women's Status: 102
- Iodine salt test: 99
- Birth Registration: 90
- Anemia Testing: 87
- Social Marketing: 85
- Vitamin A Questions: 82
- Malaria/Bednet Questions: 57
- Malaria Module: 55
DHS resources

- Measure DHS site
  - Info on surveys – manuals, questionnaires
  - Survey lists – by country, year, region, etc.
  - Statcompiler – custom tables/graphs of indicators, possible to obtain national estimates, and some stratifications
MICS

- Multiple indicator cluster surveys
  - Funded & managed by Unicef
  - Started in 1995:
    - MICS1 (1995)
    - MICS2 (2000)
    - MICS3 (2005-06)
    - MICS4 (2009-11)
    - MICS5 (2012-14)
MICS resources

- Childinfo site
  - Information on surveys – manuals, questionnaires, reports
  - Survey lists
  - MICScompiler - separate site where you can build tables, graphs and maps
### Available surveys

<table>
<thead>
<tr>
<th>Country</th>
<th>Countdown</th>
<th>Counted year</th>
<th>Year</th>
</tr>
</thead>
<tbody>
<tr>
<td>Afghanistan</td>
<td>y</td>
<td>2000(a)</td>
<td>2003(a)</td>
</tr>
<tr>
<td>Ethiopia</td>
<td>y</td>
<td>2000</td>
<td>2005</td>
</tr>
<tr>
<td>Lebanon</td>
<td>n</td>
<td>2000(a)</td>
<td></td>
</tr>
<tr>
<td>Malawi</td>
<td>y</td>
<td>1992(a)</td>
<td>1995(a)</td>
</tr>
<tr>
<td>Pakistan</td>
<td>y</td>
<td>1990(a)</td>
<td></td>
</tr>
<tr>
<td>Pakistan (Balochistan)</td>
<td>y</td>
<td>2010</td>
<td></td>
</tr>
<tr>
<td>Peru</td>
<td>y</td>
<td>1986(a)</td>
<td>1991(a)</td>
</tr>
<tr>
<td>Rwanda</td>
<td>y</td>
<td>1992(a)</td>
<td>2000</td>
</tr>
<tr>
<td>Tanzania</td>
<td>y</td>
<td>1991(a)</td>
<td>1996</td>
</tr>
</tbody>
</table>

Legend:
- MICS
- DHS standard
- DHS continuous
- DHS special or interim
- Non-standard DHS
- a = old survey for which data are available but were not analyzed (MIC51, MIC52, DHS1, DHS2)
- b = old survey for which data are not available or have restricted access (MIC51, MIC52, DHS1, DHS2)
- n = recent survey for which data are not available (MIC5 onwards, DHS3 onwards)
- z = survey that was recently made available - yet to be analyzed
ICEH - original objectives

• Analyze all DHS or MICS starting with DHS3 and MICS3 and update analyses as new surveys become available
  • coverage of a set of core Countdown maternal, neonatal and child health interventions + WHO selected indicators
  • the composite coverage indicator
  • the co-coverage index for preventive interventions
• Do original analysis & write up papers
• Maintain an equity database
Coverage indicators

- Contraceptive prevalence / demand for family planning
- Skilled attendant at delivery
- Antenatal care (1+ visit with skilled provider)
- Antenatal care (4+ visits)
- C-section rate
- Early initiation of breastfeeding
- Postnatal visit for baby
- DPT3, measles, polio vaccine
- Oral rehydration therapy for children with diarrhea
- Careseeking for pneumonia
- Insecticide-treated bednets (child slept under last night)
- Vitamin A supplementation
- Improved drinking water source
- Improved sanitation
Stratifiers

- Sex
- Place of residence
- Wealth quintiles
  - Q5/Q1 ratio & difference
  - Slope index of inequality (absolute inequality)
  - Concentration index (relative inequality)
- Maternal education
- Region of the country
- In 2013 adding maternal age
What we have done so far

- Re-analysis of 207 DHS & MICS surveys
- Indicators estimated
  - All coverage estimates relevant to Countdown and GHO
  - Under-5 mortality and all components
  - Age-specific fertility rates
  - Nutritional scores, % of deficit and % children obese
- All estimates stratified by
  - Wealth quintiles, maternal education, urban/rural, geographic area, sex of the child
What our analyses added

• From compilers
  • Not all indicators are directly available (e.g. SBA)
  • Stratifications vary by indicator

• All our estimates have
  • N’s (important to assess validity)
  • SE’s (direct estimation or jackknife for mortality & fertility)

• Composite indicators
  • Composite coverage index
  • Co-coverage (only estimable from microdata)
Individual vs batch analysis

• From start, our idea was to build a platform allowing for standardized analyses of a large number of surveys

• Advantages
  • Quick response to new indicators or stratifiers
  • Comparable estimates, independent of local adaptations or political decisions
  • Ability to produce specific indicators such as the CCI and co-coverage
The analysis process

**Prep**
- Dataset names
- Check variables
- List relevant variables & codes
- → Control dataset

**Analysis**
- Use individual, child and household datasets
- Merge wealth index and nutritional scores
- → Estimates

**Verification**
- All estimates and stratifications produced?
- Results comparable to published?
- → Final results
Equity datasets

- Our main statistical package = Stata
- Results in 6 Stata datasets
  - 1 for national estimates = “all.dta”
  - 1 for each stratifier
    - Wealth quintiles = “wiq.dta”
    - Maternal education = “meduc.dta”
    - Area of residence (urban/rural) = “area.dta”
    - Region of the country = “region.dta”
    - Sex of the child = “sex.dta”
- Also, results in a single Excel spreadsheet
  - Each original Stata dataset in one separate sheet
  - Tabs have the same names as the Stata datasets

Plenty of opportunity to work with the datasets in the practical sessions
Full equity profile

Each country gets
- A big table
- 4 graphs

They sit at
www.countdown2015mnch.org/reports-and-articles/equity

There is also an Excel version for easier data access.
The table rows

- Stratifiers in the rows
  - 5 in total
    - Wealth quintiles
      - 1 to 5
    - Maternal education
      - No education
      - Primary
      - Secondary +
    - Child’s sex
    - Urban/rural
    - Geographic region
      - Variable by country

Azerbaijan (DHS 2006)

<table>
<thead>
<tr>
<th></th>
<th>Family planning needs satisfied</th>
<th>Contraceptive prevalence rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>National</td>
<td>69.4</td>
<td>51.1</td>
</tr>
<tr>
<td>Wealth quintiles</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q1</td>
<td>71.3</td>
<td>55.2</td>
</tr>
<tr>
<td>Q2</td>
<td>65.5</td>
<td>48.0</td>
</tr>
<tr>
<td>Q3</td>
<td>65.9</td>
<td>47.0</td>
</tr>
<tr>
<td>Q4</td>
<td>65.6</td>
<td>48.4</td>
</tr>
<tr>
<td>Q5</td>
<td>78.4</td>
<td>57.2</td>
</tr>
<tr>
<td>Equity indicators</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q5/Q1 ratio</td>
<td>1.1</td>
<td>1.0</td>
</tr>
<tr>
<td>Q5-Q1 difference</td>
<td>7.1</td>
<td>2.0</td>
</tr>
<tr>
<td>Concentration index</td>
<td>2.1</td>
<td>1.3</td>
</tr>
<tr>
<td>SII</td>
<td>7.5</td>
<td>2.8</td>
</tr>
<tr>
<td>Maternal education</td>
<td></td>
<td></td>
</tr>
<tr>
<td>none</td>
<td>67.6</td>
<td>51.4</td>
</tr>
<tr>
<td>primary</td>
<td>39.8</td>
<td>25.0</td>
</tr>
<tr>
<td>secondary +</td>
<td>69.6</td>
<td>51.4</td>
</tr>
</tbody>
</table>
The table columns

<table>
<thead>
<tr>
<th>National Wealth quintiles</th>
<th>Family planning needs satisfied</th>
<th>Contraceptive prevalence rate</th>
<th>Antenatal care (1+ visits, skilled provider)</th>
<th>Antenatal care (4+ visits, any provider)</th>
<th>Skilled attendant at delivery</th>
<th>C-section rate</th>
<th>Early initiation of breastfeeding</th>
<th>Postnatal care for babies born at home</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>69.4</td>
<td>51.1</td>
<td>76.9</td>
<td>45.2</td>
<td>88.6</td>
<td>4.7</td>
<td>31.9</td>
<td>10.8</td>
</tr>
</tbody>
</table>

- **Indicators in the columns**
  - FPS – family planning need satisfied
  - Contraceptive prevalence
  - ANC1s, ANC4
  - SBA
  - C-section
  - Early BF
  - Postnatal care for babies (born @ home & all)

- **Indicators in the columns**
  - BCG, DTP3, Measles
  - Vitamin A
  - ORT
  - Careseek for diarrhea
  - ITN by children
  - Improved drink water
  - Cocoverage
  - CCI – composite coverage index
Reading the table

- Each cell shows coverage
- For the whole country
- For wealth quintiles
- And so on…
  - Not getting into the definition of each one

- Oops
  - Not all rows are stratifiers!

Azerbaijan (DHS 2006)

Table - Coverage of Countdown indicators

<table>
<thead>
<tr>
<th></th>
<th>Family planning needs satisfied</th>
<th>Contraceptive prevalence rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>National</td>
<td>69.4</td>
<td>51.1</td>
</tr>
<tr>
<td>Wealth quintiles</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q1</td>
<td>71.3</td>
<td>55.2</td>
</tr>
<tr>
<td>Q2</td>
<td>65.5</td>
<td>48.0</td>
</tr>
<tr>
<td>Q3</td>
<td>65.9</td>
<td>47.0</td>
</tr>
<tr>
<td>Q4</td>
<td>65.6</td>
<td>48.4</td>
</tr>
<tr>
<td>Q5</td>
<td>78.4</td>
<td>57.2</td>
</tr>
<tr>
<td>Equity indicators</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q5/Q1 ratio</td>
<td>1.1</td>
<td>1.0</td>
</tr>
<tr>
<td>Q5-Q1 difference</td>
<td>7.1</td>
<td>2.0</td>
</tr>
<tr>
<td>Concentration index</td>
<td>2.1</td>
<td>1.3</td>
</tr>
<tr>
<td>SII</td>
<td>7.5</td>
<td>2.8</td>
</tr>
</tbody>
</table>

Maternal education

- none: 67.6, 51.4
- primary: 39.8, 25.0
- secondary+: 69.6, 51.4
Equity indicators

• Just after the coverage by wealth quintiles
  • Some equity indicators:
    • Richest / poorest ratio
    • Richest – poorest difference
    • Concentration index (CIX)
    • Slope index of inequality (SII)

We will discuss equity indicators in detail in the next equity lectures.

<table>
<thead>
<tr>
<th>Equity indicators</th>
<th>FPS</th>
<th>CPR</th>
<th>ANC1s</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q5/Q1 ratio</td>
<td>1.1</td>
<td>1.0</td>
<td>1.8</td>
</tr>
<tr>
<td>Q5-Q1 difference</td>
<td>7.1</td>
<td>2.0</td>
<td>42.1</td>
</tr>
<tr>
<td>Concentration index</td>
<td>2.1</td>
<td>1.3</td>
<td>11.4</td>
</tr>
<tr>
<td>SII</td>
<td>7.5</td>
<td>2.8</td>
<td>53.0</td>
</tr>
</tbody>
</table>
Issues to bear in mind

• When using survey-based indicators

1. Measurement error

• Results only as good as the quality of info collected
• Questionnaires rely on understanding and recall
  • Complex indicators are less reliable
  • Some definitions are loose – ex. pneumonia based on cough and difficulty breathing
• Anthropometry is difficult!
  • Especially measurement of height/length
Issues continued

2. Sampling error
   - Coverage and prevalence measured to a given precision
   - Subgroups have smaller Ns, and less precision
   - Need to take precision into account
     - Time trends, group comparisons
     - Evident differences may not be so evident!

More issues will be covered in another lecture