Multi-Topic Household Survey Design

TALIP KILIC & ALBERTO ZEZZA
Poverty & Inequality Group
Development Research Group
The World Bank

Multi-Topic Household Surveys
March 6, 2013
Key Survey Design Components

- Data quality, cost efficiency, relevance & timeliness depend on survey design choices
- Documentation is needed throughout
- Session will provide an overview of issues that will be discussed in-depth in subsequent modules
Questionnaire Design: Drivers & Coordination Needs

• Understanding analytical needs of differing complexity crucial
  – Existing policy questions & monitoring needs
  – Emerging policy questions & monitoring needs
  – Research questions, specific program impacts

• Multi-topic approach & need for stakeholder coordination
  – Clients w/in the Bank & the Government
  – Donor Partners
  – International & Local Academic & Research Institutions

• Inclusive questionnaire design process enhances ownership acceptance, & use of the data
Questionnaire Design: First Principles

• History matters:
  – Comparability with the previous series, lessons learned
  – Previous international & domestic survey efforts & associated questionnaires often provide useful reference material

…but don’t get trapped by it!

• Tension between comparability & drive to improve methods
  – Exemplified by recent survey efforts (e.g. Malawi Third Integrated Household Survey (IHS3) 2010/11) that have attempted to use photo aids for better quantification of “size” variations for food item
    non-standard measurement unit combinations
  – Build a way to transition to new, well-tested methods, while providing comparisons
Questionnaire Design: First Principles (Cont’d)

• Tension between comparability & best practice vs. feasibility

• Cross-country comparability vs. contextual customization
  – DHS (standardized) vs. LSMS (customization)
  – Hybrid model for a core set of questionnaire modules?

• Avoid overload:
  – How will the resulting data be used at the analysis stage?
  – Differentiate between core vs. additional modules & rotating modules in panel household surveys (e.g. Indonesia Family Life Survey)
  – “Your questionnaire is too long!
    
    BTW, can you please add these 10 questions for me?”
  – Not all surveys appropriate for all questions!
    • Sensitive information, rare events, etc…
Questionnaire Design: First Principles (Cont’d)

• Importance of systematic metadata collection for data quality control & informing future survey design
  – Date & duration of interviews, questionnaire instruments & modules
  – Supervisor, enumerator & data entry operator names & codes
  – ID Codes for respondents

• Emphasis on inter-linkages & consistency across questionnaire modules crucial to increase the use & relevance of data
  – Example: Beyond individual-level data collection on education, health, labor, anthropometrics, non-cognitive skills & short-term memory, Malawi IHPS 2013 associates over 70 questions (at varying levels of observation) with 2 to 4 roster ID codes
Questionnaire Design: First Principles (Cont’d)

• Emphasis in the context of panel household surveys on linking households, individuals, agricultural plots & other entities across survey rounds, collecting contact & reference info

• Emphasis on integration of household survey data with other third-party data, including geographic information systems, & administrative records
  – Geo-referencing dwelling locations, agricultural plots, community facilities, enterprise locations, etc… proves powerful at the analysis stage but requires careful questionnaire design, field staff training & data quality control
  – Administrative coding of survey enumeration areas with respect to administrative information systems crucial
Typical LSMS Questionnaire Instruments

• Questionnaires integral part of the documentation package

• Household
  – Food Consumption Diary may be integrated

• Agriculture
  – Livestock, Fishery questionnaires may be integrated or on their own

• Community
  – Price, Market questionnaires may be integrated

• Tracking (Applicable to panel household surveys)
  – Pre-printed forms for establishing presence of baseline household members & identifying tracking targets
  – Forms collecting information to locate tracking targets (shifted households/ individuals)
Typical LSMS Questionnaire Instruments

**Household**
- Individual-Level Data on Demographics, Education, Health, Labor & Anthro
- Housing, Durable Assets
- Food & Non-Food Consumption & Expenses
- Income
- Food Security
- Non-Farm Enterprises
- Subjective Welfare

**Agriculture**
- Plot-Level Data on (i) Land Areas, (ii) Labor & Non-Labor Input Applications, (iii) Crop Cultivation & Production
- Crop Sales & Utilization
- Farm Implements
- Extension services
- Livestock
- Fisheries

**Community**
- Demographics
- Services
- Facilities
- Infrastructure
- Governance
- Organizations & Groups
- Price Module
- Market Survey
Methodological Work

- Methodological survey experiments to inform questionnaire and fieldwork design in key domains of data collection

<table>
<thead>
<tr>
<th>LSMS-IV</th>
<th>LSMS-ISA</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Finance</td>
<td></td>
</tr>
<tr>
<td>• Consumption</td>
<td></td>
</tr>
<tr>
<td>• Labor</td>
<td></td>
</tr>
<tr>
<td>• Income</td>
<td></td>
</tr>
<tr>
<td>• Migration</td>
<td></td>
</tr>
<tr>
<td>• Subjective welfare</td>
<td></td>
</tr>
<tr>
<td>• Gifting</td>
<td></td>
</tr>
<tr>
<td>• Land Area</td>
<td></td>
</tr>
<tr>
<td>• Soil Fertility</td>
<td></td>
</tr>
<tr>
<td>• Crop Production</td>
<td></td>
</tr>
<tr>
<td>• Agricultural Labor</td>
<td></td>
</tr>
<tr>
<td>• Skill Testing</td>
<td></td>
</tr>
<tr>
<td>• Sampling for Nomads</td>
<td></td>
</tr>
<tr>
<td>• Livestock Production/Pastoralists</td>
<td></td>
</tr>
</tbody>
</table>
LSMS Publications

• Household Surveys:
  – Designing Household Survey Questionnaires for Developing Countries: Lessons from 15 years of the Living Standards Measurement Study

• Tracking:
  – Tracking in Longitudinal Household Surveys

• Climate Change:
  – Improving Household Survey Instruments for Understanding Agricultural Household Adaptation to Climate Change: Water Stress and Variability
  – Understanding Agricultural Households’ Adaptation to Climate Change and Implications for Mitigation: Land Management and Investment Options

• Fisheries:
  – Design and Implementation of Fishery Modules in Integrated Household Surveys in Developing Countries

• Computer-Assisted Personal Interviewing (CAPI)
  – Comparative Assessment of CAPI Software Packages
Human Subject Research & Ethics

• LSMS surveys implemented by national statistical agencies immune to developing country institutional review board (IRB) clearance

• The Bank currently does not have an IRB, although different initiative within the Bank may require a review of the research by an in-country, an academic or a private IRB

• Efforts underway to formulate institutional guidelines & provide training to help integrate ethics considerations into human subjects research at the Bank
  – Consent, data security, obligations for follow-up key issues
Sampling Design

• Need for a **specialist** & **comprehensive documentation** on enumeration area (EA) & household selection processes
  – Permits making inferences from the sample to the entire population with known margins of error

• Underlined by theory: can verify whether done right

• Direct impact on field organization, survey costs

• Trade-offs: Sample size/representativeness & quality control
  – Population size has almost no influence on the sample size necessary to achieve a given precision
  – Increasing the sample size reduces sampling errors while increasing the prospects for non-sampling errors
Sampling Design (Cont’d)

• Optimal sample size: Balancing accuracy needs with logistical, timing & resource Constraints

• National household survey samples not simple random samples, but instead stratified (by Region, Urban/Rural, Treatment/Control) & selected in multiple stages (Eas & HHs within EAs)

• Why stratification?
  – In order to potentially improve precision, by gaining control of the composition of the sample
  – In order to provide estimates for subgroups that would otherwise be poorly represented (small regions, women-headed HHs, etc.)
Sampling Design (Con’d)

• Why two stages?
  – An updated list of all households generally unavailable
  – A single-stage sample likely too scattered, therefore costly, to cover

• While stratification works towards improving precision, clustering comes at the cost of precision

• Most stratified samples select households with unequal probabilities as such survey needs to be analyzed with weights

• Parts of the country may need to be excluded due to accessibility, security, or other reasons
  – OK if properly documented & results are not extrapolated later to the whole country […]
Sampling Design (Cont’d)

- EAs in Eastern Ethiopia excluded from the EA universe consulted for the selection of Ethiopia Rural Socio-Economic Survey (ERSS) 2011/12 EAs
Sampling Design (Cont’d)

• Key Inputs:
  – Population-based sampling frames
    ▪ Census enumeration area (EA) listing (i.e. household counts)
  – Area-based sampling frames
  – Past survey data on key outcomes variables driving power calculations
  – Updated sampled EA household listings
    ▪ Need for close supervision
    ▪ The time and budget of household listing are
      o Small enough to be considered a marginal part of the overall data collection effort
      o Large enough to be a headache if they are forgotten or underestimated
Fieldwork Preparation & Implementation

• Allow adequate time for piloting, staff recruitment, training, procurement of equipment & organization of survey vehicle fleet, finalization & printing of questionnaires

• Human resource management is of paramount importance
  – Training: In-class & field practice, usually 3-5 Weeks, depending on complexity of survey design
  – Team leader verification, data entry feedback, headquarters field visits crucial for monitoring & improving interviewer performance
  – Need for timely data quality control & feedback provision to rectify enumeration mistakes & avert future ones

• Documentation of fieldwork implementation issues to alert analysts on potential pitfalls & inform future survey design
Piloting

• Needs to be scheduled 2 to 4 weeks prior to the training
• Essential for identification of gaps, duplications, inconsistencies
• Flow of questionnaire, understanding of questions, applicability of response categories are key areas of focus
• Possible reformulation of questions & training methods conditional on interactions with the respondents
Field Staff Recruitment & Training

• Systematic & documented tests to identify qualified candidates
• NSO statistical clerks/enumerator corps complemented with contracted enumerators
• Factors in staff selection, team organization & assignment
  – Gender
  – Ethnicity
  – Languages
  – Previous experience with similar household surveys or with prior survey rounds in the context of panel surveys
  – Special skills (depending on the survey design)
Timing, Single vs. Multiple Visits

• Timing of fieldwork
  – Comparability with earlier rounds, project implementation cycle in the context of impact evaluations, the NSO workload, festivities (e.g. Ramadan), agricultural calendar, elections & weather often important parameters to take into account & plan for

• Organization: Mobile teams vs. resident enumerators

• LSMS survey interviews traditionally administered in a single visit, distributed randomly over a 12-month period to accommodate seasonality in consumption outcomes

• Increasing use of multiple visits
  – Synchronization with agricultural season to reduce recall
  – Dividing workload across visits vs. repeated observations
  – Cost implications
Multiple Visit Case Study: Malawi Third Integrated Household Survey (IHS3) 2010/11

• IHS3 designed to be representative at the national-, regional-, urban/rural-& district-level, covering 768 EAs across Malawi, over a period of 12-months to capture seasonality in household consumption

• Sampling design & 12-month implementation approach comparable to the IHS2 2004/05 set up

• 204 IHS3 EAs designated as “Panel” prior to the IHS3 field work, for tracking and re-interview as part of the Integrated Household Panel Survey (IHPS) 2013

• 564 IHS3 EAs designated as “Cross-Sectional”
Multiple Visit Case Study (Cont’d)

• Panel EAs: Visited twice during the IHS3 field work, in sync with the 2009/10 rainy agricultural season (November-June) primarily to reduce recall in agricultural data collection
  – 1st (Post-Planting/Pre-Harvest) Visit Period: April-June 2010
  – 2nd (Post-Harvest) Visit Period: July-September 2010
  – To be revisited in 2013 as part of the IHPS, in accordance with the IHS3 interview timeline

• Cross-Sectional EAs: Visited once during the IHS3 field work, provided information on last completed agricultural season (2008/09 or 2009/10) depending on the date of the interview
  – Comparable to the IHS2 2004/05 set up
# Multiple Visit Case Study (Cont’d)

<table>
<thead>
<tr>
<th>EA CLASSIFICATION</th>
<th>VISIT 1 WORKLOAD</th>
<th>VISIT 2 WORKLOAD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Panel A (102 EAs)</td>
<td>HH Qx – Full&lt;br&gt;AG Qx – Visit 1 (If Applicable)&lt;br&gt;Fishery Qx (If Applicable)&lt;br&gt;Community Qx</td>
<td>HH Qx – HH Roster Update Only&lt;br&gt;AG Qx – Visit 2 (If Applicable)</td>
</tr>
<tr>
<td>Panel B (102 EAs)</td>
<td>HH Qx – HH Roster Only&lt;br&gt;AG Qx – Visit 1 (If Applicable)</td>
<td>HH Qx – Full&lt;br&gt;AG Qx – Visit 2 (If Applicable)&lt;br&gt;Fishery Qx (If Applicable)&lt;br&gt;Community Qx</td>
</tr>
<tr>
<td>Cross-Sectional (564 EAs)</td>
<td>HH Qx – Full&lt;br&gt;AG Qx – Full (If Applicable)&lt;br&gt;Fishery Qx (If Applicable)&lt;br&gt;Community Qx</td>
<td>NOT APPLICABLE</td>
</tr>
</tbody>
</table>
Multiple Visit Case Study… (Cont’d)

<table>
<thead>
<tr>
<th>EA CLASSIFICATION</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>TOTAL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Panel A</td>
<td>34</td>
<td>34</td>
<td>34</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>102</td>
</tr>
<tr>
<td>Panel B</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>34</td>
<td>34</td>
<td>34</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>102</td>
</tr>
<tr>
<td>X-Section</td>
<td>30</td>
<td>30</td>
<td>30</td>
<td>30</td>
<td>30</td>
<td>64</td>
<td>64</td>
<td>64</td>
<td>64</td>
<td>64</td>
<td>64</td>
<td>64</td>
<td>564</td>
</tr>
<tr>
<td>TOTAL</td>
<td>64</td>
<td>64</td>
<td>64</td>
<td>64</td>
<td>64</td>
<td>64</td>
<td>64</td>
<td>64</td>
<td>64</td>
<td>64</td>
<td>64</td>
<td>64</td>
<td>768</td>
</tr>
</tbody>
</table>
Tracking in Panel Household Surveys

- Dimensions: Shifted (intact) households, shifted individuals, agricultural plots, other entities…
- Preparation of baseline data for updating
- Identification of eligible tracking targets
  - Criteria: age, relationship to head, etc.?
- Completion & data entry of tracking forms on eligible targets
- Local vs. distant tracking
  - Organization of workload, information sharing across teams
- Complementary tracking survey teams, field staff incentives?
- Designation of a team member to organize & follow up on tracking specifically
Data Entry & Processing

- Centralized Data Entry
- Computer-Assisted Field Entry (CAFE)
- Computer-Assisted Personal Interviewing (CAPI)
- CAFE & CAPI, if implemented right, yields better results in staff learning & addressing enumeration errors no/limited lag
- Second Data Entry & Verification recommended to be combined with Centralized Data Entry or CAFE to rectify data entry errors
- Designation of a well-supervised, small team tasked with data cleaning, provided with clear protocols
Data Entry & Processing (Cont’d)

• Systematic documentation of changes to raw data & anonymization of unit record data prior to dissemination

• LSMS fully committed to open data
  – LSMS surveys supported by the team without direct financial contribution have been documented & made publically available (a few exceptions, particularly from the earlier era, remain)
  – LSMS surveys directly financed under the program documented & made publically available within 1 year of completion of field work
  – Public data dissemination & timeline needs to be an integral part of the agreements with the respective governments
  – Contracted firms must provide non-anonymized raw data; needs to be part of the firm contract, cleaning routines documented & fully replicable
Data Entry & Processing (Cont’d)

Centralized

CAFE

CAPI
### Sample Schedule of Activities for a Cross-Sectional LSMS Survey

| Activity groups                  | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 | 25 | 26 | 27 | 28 | 29 | 30 |
|---------------------------------|---|---|---|---|---|---|---|---|---|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|
| 1 Logistical Planning           |   |   |   |   |  |   |   |   |   |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |
| 2 Questionnaire Design          |   |   |   |   |   |   |   |  |  |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |
| 3 Sampling Design               |   |   |   |   |   |   |   |   |  |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |
| 4 Staffing and Training         |   |   |   |   |   |   |   |   |   |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |
| 5 Data Entry Platform Development|   |   |   |   |   |   |   |   |   |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |
| 6 Data Entry & Processing       |   |   |   |   |   |   |   |   |   |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |
| 7 Fieldwork                     |   |   |   |   |   |   |   |   |   |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |
| 8 Data Analysis                 |   |   |   |   |   |   |   |   |   |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |
| 9 Documentation                 |   |   |   |   |   |   |   |   |   |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |
Key Messages

• Manage the survey as an integrated project
• **Documentation** is a MUST at every step
• Stakeholder **coordination** to ensure ownership & use of data
• Hire a **sampling expert**, fully document in-office procedures & supervise **household listing operations**
• Organize fieldwork on the basis of **teams**
• Ensure sufficient **training & field practice** prior to fieldwork
• Establish strong **supervision** procedures & **in-field presence**
• Integrate computerized **data quality control** to fieldwork
Multi-Topic Household Survey Design

TALIP KILIC & ALBERTO ZEZZA
Poverty & Inequality Group
Development Research Group
The World Bank

Multi-Topic Household Surveys
March 6, 2013