Computer-Assisted Personal Interviewing

ARTHUR SHAW
Consultant
Development Research Group
The World Bank

Multi-Topic Household Surveys
March 6, 2014
The evolution of data capture

PAPI

CAFE

CAPI
What does CAPI do?

• Everything a paper questionnaire does…
  – Provides a structured interview script
  – Offers means of recording answers

• …and many things paper does not
  – Data entry—as part of data capture
  – Error correction—at moment of data collection
  – Facilitates data transfer—as part of an integrated system
Automated routing

1. IS Gaddafi ANSWERING FOR HIMSELF OR HERSELF? [Yes [1]]
2. WHAT IS THE NAME OF THE PERSON RESPONDING FOR Gaddafi?
3. Can Gaddafi read and write with understanding in any language? [able to read and write [1]]
4. Has Gaddafi ever attended any formal school? [Never attended [1]]
5. Why has Gaddafi not attended school? 
6. What was the highest grade/class that Gaddafi completed?
7. What was the main reason that Gaddafi left school?
8. What grade/class was Gaddafi attending in the last complete
Non-linear navigation

We would like to make a complete list of Household members in the last 12 months including guests who slept here last night and those that left permanently.

<table>
<thead>
<tr>
<th>Wave1</th>
<th>Name</th>
<th>Gender</th>
<th>Residential Status</th>
<th>Age</th>
<th>Moved</th>
<th>Died</th>
<th>Died</th>
<th>Demographics</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Mugalu Mohammed</td>
<td>Male [1]</td>
<td>Usual member present [1]</td>
<td>3</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>![a]</td>
</tr>
<tr>
<td>3</td>
<td>Kimbowa Martin</td>
<td>Male [1]</td>
<td>Usual member present [1]</td>
<td>19</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>![a]</td>
</tr>
</tbody>
</table>
Consistency checks

1. IS Mugalu ANSWERING FOR HIMSELF OR HERSelf? Yes [2]
2. WHAT IS THE NAME OF THE PERSON RESPONDING FOR Mugalu? Kimbowa Martin
3. During the past 30 days, did Mugalu suffer from any illness or injury? Yes [1]
4. For how many days did Mugalu suffer due to illness or injury during the past 30 days? 4
5. For how many days did Mugalu have to stop doing Mugalu's usual activities due to illness o...
6. Can you describe the symptoms that Mugalu primarily suffered due to Diarrhoea (acute) [1]
7. Can you describe the symptoms that Mugalu primarily suffered due to
8. Was anyone consulted (e.g. a doctor, nurse, pharmacist or traditional healer) for the major...

Member Health

<table>
<thead>
<tr>
<th>Wave 1</th>
<th>Name</th>
<th>Age</th>
<th>Health Info</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Mugalu Mohammed</td>
<td>3</td>
<td>![a. Health]</td>
</tr>
<tr>
<td>2</td>
<td>Gadaffi Ismail</td>
<td>25</td>
<td>![a. Health]</td>
</tr>
<tr>
<td>3</td>
<td>Kimbowa Martin</td>
<td>19</td>
<td>![a. Health]</td>
</tr>
</tbody>
</table>

LSMS
Living Standards Measurement Study
Consistency checks (Cont’d)
# Use of media for better quantification

**T6AE Food consumption item**

**105 Sweet potatoes (fresh)**

**Household ID 1021000402**

**Over the last 7 days**

1. How many days in the past 7 days was Sweet potatoes (fresh) consumed?

2. Unit of quantity

<table>
<thead>
<tr>
<th></th>
<th>Quantity</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.</td>
<td>Amount of Sweet potatoes (fresh) consumed in the HOUSEHOLD over the last 7 days which came from PURCHASE...</td>
<td></td>
</tr>
<tr>
<td>4.</td>
<td>Amount of Sweet potatoes (fresh) consumed AWAY FROM HOME over the last 7 days which came from PURCHASES...</td>
<td></td>
</tr>
<tr>
<td>5.</td>
<td>Amount of Sweet potatoes (fresh) consumed over the last 7 days which came from HOME PRODUCTION?</td>
<td></td>
</tr>
<tr>
<td>6.</td>
<td>Amount of Sweet potatoes (fresh) consumed over the last 7 days which came from GIFTS/IN-KIND sources?</td>
<td></td>
</tr>
</tbody>
</table>

**Totals**

<table>
<thead>
<tr>
<th></th>
<th>0.0</th>
</tr>
</thead>
</table>

*If none, enter 0*
Use of media for better quantification
Use of external sensors

Water quality
Weight
Soil quality
How does CAPI compare to PAPI?

- **Cost** of data collection
- **Timeliness** of data delivery
- **Quality** of data
Costs of CAPI

No longer pay

- Printing
  - Paper (as much)
  - Packing questionnaires

- Data entry
  - Training a team
  - Time to type (twice)

- Transporting questionnaires
  - Fuel and vehicles
  - Drivers

Now pay

- Hardware
- Software
- Survey programmer time
- Data transfer costs
- Extra (CAPI) training time
- Extra (pilot) testing time
Cost, in other words

**No longer pay**

- Disposable goods
  - Printed surveys

- One-time services
  - Printing
  - Data entry
  - Transportation for questionnaires

**Now pay**

- Fixed costs, with benefits
  - Survey programmer time
  - Extra training and testing time

- Investments
  - Hardware
  - Software
  - (Survey programmer time)
Costs, hidden

• Costs of learning
  – How to be a critical consumer of CAPI software, hardware
  – How to develop the a CAPI questionnaire (as a team)
  – How to test a CAPI questionnaire

• Costs of doing
  – Pain of making (and checking) last-minute questionnaire changes
Costs, other hidden

• **Replacement**
  – If machines break
  – If machines are stolen

• **Maintenance**
  – Hardware
  – Software
Data collection → Data entry → Data cleaning → Data processing → Data analysis → Policy decisions
Quality

• Limited quantitative empirical evidence that CAPI improves data quality
  – Caeyers et al. (2012). Reduces or eliminates routing errors as well as unlikely or impossible entries.
  – Fafchamps et al. (2012). Reduces number of missing values, and slightly increased autocorrelation of measures expected to be correlated across time.
Quality

• But reasons to believe that CAPI may improve quality
  – Data validation occurs at the time of entry, when corrections can rely on better information (i.e., the respondent’s rather than that of survey staff)
  – Data passes through fewer steps during which errors could be made (e.g., data editing, data entry)
Quality

• Caveats
  – Depends critically on the quality of programming
    • Poor programming can make CAPI no better than PAPI (e.g., if a check is not programmed)
    • Or considerably worse (e.g., wrong skip jumps past questions the enumerator should ask)
  – CAPI is a tool, not a solution
    • CAPI surveys are still surveys
    • As such, they need the same type of oversight and quality control. If anything, might need more oversight at the beginning, to make sure programmed properly.
So you’re thinking about a CAPI survey…

• What the LSMS experience has been
• At least three things to consider
  – Staff capacity
  – Hardware
  – Software
LSMS Experience

• LSMS operations marked by a gradual transition to CAPI
  – 2003 - CAPI survey experiment (~200 households) (Albania)
    • Application developed in CSProX
  – 2007 - CAPI survey (~500 hholds) (Ngara District, Tanzania)
    • Application developed in CWEST
  – 2010 - Kagera Health and Development Survey (KHDS) (Tanzania)
    • Application developed in CWEST
    • Supported by LSMS-ISA, implemented by Bureau of Statistics
    • Partial transition to CAPI in 2010/11 (in CWEST); CAPI transition completed in 2011/12 (on-going; in CWEST & Surveybe); Next round 2013/14
  – 2011-12 - Ethiopia Rural Socioeconomic Survey (ERSS) (2011-2014)
    • Supported by LSMS-ISA, implemented by Central Statistical Agency
    • CAPI application developed in Surveybe for the Ag Questionnaire, implemented in a subset of EAs in 2011/12
• 2014 – pilots with World Bank CAPI system
Uganda National Panel Survey (UNPS) CAPI Experience

- Team leaders quick to adapt, instrumental in training & knowledge sharing

- Required change in institutional thinking on HH survey: Greater up-front work (& costs) with respect to PAPI with CAFE
  - Prep of Wave I (PAPI) data uploaded onto Wave II (CAPI) application
  - Hundreds of intra/inter-module consistency checks, in addition to range & default checks for missing values
  - Programming of rules on generation of household & individual identifiers for new additions to the sample
  - Training of UBoS Headquarters staff on case management suite
UNPS CAPI Experience (Cont’d)

• In-country procurement problems
  – Lags assoc. with operating within Government systems/unreliable suppliers
  – US procurement by the LSMS-ISA project: Not straightforward either

• Anti-virus software critical to maintaining the hardware integrity

• Application glitches even after piloting three times: Need for more intensive testing in comparison to PAPI with CAFE

• CAPI application platform based on multiple software packages: CWEST & CSPro (in 2010/11); CWEST & Surveybe (in 2011/12)
  • Dependence on the CWEST application developer for adjustments
  • Continued reliance on multiple software packages necessitated by lack of case management features on Surveybe

• Timely communication of bugs that might compromise the integrity of incoming data critical: No paper questionnaires to re-enter
UNPS CAPI Experience (Cont’d)

• Continuing improvements to the CAPI application on a rolling basis throughout the field work

• Even with internet dongles, slow internet speeds & lack of service in certain areas
  – Affects timely headquarters review of data sent from the field
  – Receipt of application updates by the survey teams not always timely

• Regular backup of interview files in the field & at the HQ crucial

• Lags associated with Surveybe data export

• Still need a paper questionnaire for dissemination purposes: CAPI application dictionary is not more than a linear questionnaire report
Staff capacity

- Capacity assessment from the field to the corner office
  - Enumerators – computer literate?
  - Support staff – got programmers? What kind?
  - Survey developers – understand technology?
  - Management – understand and support shift?
Hardware

Asus Eee PC T101MT (UNPS; ERSS) $450-500

Samsung Q1b Ultra (KHDS; UNPS) $650-700

Google Nexus
10": $335-372
7": $229-269
Hardware

• What you should prefer
  – Low price
  – Rugged design
  – Proven reliability
  – Long battery life
  – Largest (affordable) screen
  – Redundant systems for data entry
Hardware

• **Hardware choice is software choice**
  – Windows devices ➔ Windows software
  – Android devices ➔ Android software

• **Hardware does not stop at devices**
  – Protective covers
  – Powering options
  – Connectivity devices
Software

• Three critical questions:
  1. What do you need?
  2. What can you pay?
  3. How much time are you willing to invest?
Software

• Three tradeoffs in software selection

1. Difficulty and price
   • Easiest to use is (often) the most expensive
   • Hardest have the steepest learning curves but lowest prices

2. Openness and support
   • Most open-source software has little or scattered resources
   • Most proprietary (and thus fee-based) software has excellent support and documentation

3. Ease of use and ease of re-use
   • Menu-based programming
   • Command line programming
Software

• **Need help selecting a software?**

• **Where to look**
  
  – LSMS-ISA > Tools > Sourcebooks
  
  – Comparative Assessment of Computer-Assisted Interview (CAPI) Software Packages

• **What it contains**
  
  – Brief overview of each packages
  
  – Comparative assessment
  
  – Detailed evaluation of each package
Software

• Demonstration of the World Bank’s CAPI software
  – Designer
  – Interviewer
  – Supervisor
  – Headquarters

• Where to learn more
  – www.worldbank.org/capi