Social Survey Data Collection
Challenges and Trends

Gina Cheung
Beth-Ellen Pennell

North American DDI Conference
April 1-2, 2014

© 2014 by the Regents of the University of Michigan
Survey Life Cycle

1. Questionnaire design
2. Questionnaire & SMS programming
3. Interviewer training
4. Field work
5. Data processing; coding
6. Quality assurance
7. Data dissemination

DDI Lifecycle

Concept
- Initial concepts
- Questions and answers
- Grant info

Collection
- Questionnaire
- Coded instrument
- CAI metadata
- Paradata

Processing
- Data specs
- Recodes
- Summary descriptive info

Distribution
- Terms of use
- Citation
- Packaging info

Discovery
- Catalog record
- Indexing
- Related publications

Analysis
- Replication code
- Publications

Archiving
- Preservation metadata
- Confidentiality
- Add’l processing

Repurposing
- Post-hoc harmonization
- Data transformations
Agenda

• Questionnaire Design Challenges
• Survey Management Challenges
• “New” Technology Challenges
Agenda

• Questionnaire Design Challenges
Questionnaire Design

- Traditional Q-list questionnaire
- Word memory list
- Event History Calendar
- Computer assisted self-administered interview
- Neurocognitive tests
- Biomarker data collection and Consent form
- Traditional Web surveys
- Classes Room Observation/Coding/Tagging
How large is large? -- Examples

• Ghana Socioeconomic Panel Survey
  o Sample size of 5009 households, with approximately 18,000 individuals
  o Instrument variables ~ 65,000

• China Family Panel Study (CFPS)
  o Sample size: 13,000~ HHs, 50,000 ~ Individuals
  o 7 instruments total of 40,000 variables

• Mental Health Survey (WMHS)
  o 25+ counties and 30+ languages
  o Complex questionnaire design (World Health Organization’s Composite International Diagnostic Interview CIDI)
## SECTION N: HARVEST - TO BE ASKED OF R1 IN MAIN SURVEY, R2 IN R2 SURVEY

**READ:** Of course, I’d also like to know about the crops you harvested at the end of the farming season. ELECTRONIC VERSION PRELOADS R1 PLOTS IN MAIN SURVEY, R2 PLOTS IN R2 SURVEY PLOT A to J

<table>
<thead>
<tr>
<th>Plot</th>
<th>Crop Type</th>
<th>Unit of Measurement</th>
<th>Have you harvested CROP TYPE from PLOT A?</th>
<th>How many UNIT of CROP TYPE has been harvested from PLOT A?</th>
<th>In which month(s) did you harvest the CROP TYPE from PLOT A?</th>
<th>If N1.4=2 or 3 How much do you expect to harvest from PLOT A (IF N1.4=2 do not include what has been harvested already)</th>
<th>What was the crop lost to?</th>
<th>How much of your crop was lost in total?</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Rice</td>
<td>Bag</td>
<td>1 Yes, fully 2 Yes, partially 3 No</td>
<td></td>
<td></td>
<td></td>
<td>1 Yes 2 No (=&gt; NEXT PAGE)</td>
<td>See codes to the right. Select all which apply. (If OTHER, please specify.)</td>
</tr>
<tr>
<td>A</td>
<td>Maize</td>
<td>Bag (dried without cobs and processed)</td>
<td>1 Yes, fully 2 Yes, partially 3 No</td>
<td></td>
<td></td>
<td></td>
<td>1 Yes 2 No</td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>Millet</td>
<td>Bag</td>
<td>1 Yes, fully 2 Yes, partially 3 No</td>
<td></td>
<td></td>
<td></td>
<td>1 Yes 2 No</td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>Sorghum</td>
<td>Bag</td>
<td>1 Yes, fully 2 Yes, partially 3 No</td>
<td></td>
<td></td>
<td></td>
<td>1 Yes 2 No</td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>Groundnut</td>
<td>Bag (without shell)</td>
<td>1 Yes, fully 2 Yes, partially 3 No</td>
<td></td>
<td></td>
<td></td>
<td>1 Yes 2 No</td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>Soybean</td>
<td>Bag (without shell)</td>
<td>1 Yes, fully 2 Yes, partially 3 No</td>
<td></td>
<td></td>
<td></td>
<td>1 Yes 2 No</td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>Cassava</td>
<td>Bag</td>
<td>1 Yes, fully 2 Yes, partially 3 No</td>
<td></td>
<td></td>
<td></td>
<td>1 Yes 2 No</td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>Yam</td>
<td>Bowl (100 tubers)</td>
<td>1 Yes, fully 2 Yes, partially 3 No</td>
<td></td>
<td></td>
<td></td>
<td>1 Yes 2 No</td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>Cowpea</td>
<td>Bag</td>
<td>1 Yes, fully 2 Yes, partially 3 No</td>
<td></td>
<td></td>
<td></td>
<td>1 Yes 2 No</td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>Okra</td>
<td>Bag</td>
<td>1 Yes, fully 2 Yes, partially 3 No</td>
<td></td>
<td></td>
<td></td>
<td>1 Yes 2 No</td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>Pepper</td>
<td>Bag</td>
<td>1 Yes, fully 2 Yes, partially 3 No</td>
<td></td>
<td></td>
<td></td>
<td>1 Yes 2 No</td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>Watermelon</td>
<td>Bowl</td>
<td>1 Yes, fully 2 Yes, partially 3 No</td>
<td></td>
<td></td>
<td></td>
<td>1 Yes 2 No</td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>Mangoes</td>
<td>Box</td>
<td>1 Yes, fully 2 Yes, partially 3 No</td>
<td></td>
<td></td>
<td></td>
<td>1 Yes 2 No</td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>Tomato</td>
<td>Box</td>
<td>1 Yes, fully 2 Yes, partially 3 No</td>
<td></td>
<td></td>
<td></td>
<td>1 Yes 2 No</td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>Salad leaves</td>
<td>Box</td>
<td>1 Yes, fully 2 Yes, partially 3 No</td>
<td></td>
<td></td>
<td></td>
<td>1 Yes 2 No</td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>Cotton</td>
<td>Bag (different from maize bag)</td>
<td>1 Yes, fully 2 Yes, partially 3 No</td>
<td></td>
<td></td>
<td></td>
<td>1 Yes 2 No</td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>Other</td>
<td></td>
<td>1 Yes, fully 2 Yes, partially 3 No</td>
<td></td>
<td></td>
<td></td>
<td>1 Yes 2 No</td>
<td></td>
</tr>
</tbody>
</table>

**Codes for N1 9:**
1. Rotting
2. Disease
3. Fire
4. Flood
5. Drought
6. Birds
7. Insects
8. Ants
9. Bees
10. Cockroaches
12. Grasshoppers
12. Locust
13. Termites
14. Caterpillars
15. Centipedes
16. Saligui
17. Panchotkiri
18. Zunyayas
19. Tambahejigui
20. Rodents
21. Mice
22. Rats
23. Grasscutters
24. Squirrels
25. Monkeys
26. Sheep
27. Goats
28. Cattle
29. Bandicoot
30. Other (SPECIFY)
### Survey Status

**SOConsent Forms**
- S01A: Consent: **Done**
- S01B3: Consent for Under 26: **Done**

**Rosters**
- S01B2: Household Roster: **Done**
- S04: Plot Roster: **Done**
- S05: Non-Farm Enterprise Roster: **Done**
- Person Sections: **Started**
- Plot Sections: **Done**
- Non-Farm Enterprise Sections: **Not Started**

**Household Level Sections**
- S02B: Non-Resident Spouses: **Not Started**
- S02A: Non-Resident Relatives: **Not Started**
- S10C: Social Networking: **Started**
- S10D: Information Seeking: **Not Started**
- S11: Household Consumption: **Not Started**
- S12: Housing Characteristics: **Not Started**
- S04NN: Gathering: **Done**
- S03Ai: Animals: **Done**
- S03Aii: Tools: **Started**
- S03Aiii: Durable Goods: **Started**
- S03Bi: Borrowing: **Not Started**
- S03Bii: Lending: **Done**
- S03Biii: Out-Transfers: **---n/a---**
- S03Biv: In-Transfers: **---n/a---**
- S03Bv: Savings: **Done**
Major Aspects of Design and Implementation

- Questionnaire length
- Question type
- Response options
- Closed vs open-ended
- Use of visuals
- Screen layout
- Progress bar
- Slide bars, drop & drag
PAPI to CAI

- Transition from a well-defined paper & pencil (PAPI) questionnaire to a computer assisted interview (CAI) instrument
  - VERY Complex grid designs
  - No explicit consistency checks
  - Preload previous data collection
  - Question fills
  - Interviewer instructions
  - Question-by-question on-line help
  - Questionnaire translation
Agenda

- Questionnaire design Challenges
- Survey Management Challenges
Survey Data Collection “Mode”

• Computer Assisted Telephone Interview (CATI)
• Computer Assisted Personal Interview (CAPI)
• Computer Assisted Web Interview (CAWI)
• Computer Assisted Self-administrated Interview (CASI)
• Computer Assisted Data Entry (CADE)
• Paper Pencil Survey
• Mail Survey
• Group Administered Survey (either by paper or by computer)
Survey Management System (SMS)

- Survey Management System differs between modes

- Major Common Functions are:
  - Sample assignment
  - Delivery of sample to interviewers/respondents
  - Launch survey data collection software
  - Administrate sample status and the outcome
  - Send interview data to central database
  - Merge all the individual interviewer’s data files to a master data file
Context – Mixed Modes of Collection

“One of the most important challenges to survey researchers is deciding which data collection method or mix of methods is optimal...”

Pressures to use Mixed Modes of Collection

• Declining response rates
• Complex human measurements
• Increasing effort to collect surveys
• Increasing burden on respondents

→ Management information to inform decision making while fielding a survey; multi-mode or single-mode
Definition: Mixed Mode

The use of multiple ways to access, obtain self-reports, collect observations, or measure attributes, within the same survey effort.

Mixed-mode designs can use multiple modes concurrently or sequentially on the same and different sample units.
Survey Design Modes Example

Data collection with multiple modes (sequential or concurrent) or single mode:

- **Single Mode**
  - Mode #1

- **Sequential Mixed Modes**
  - Mode #1 ➔ Mode #2 ➔ Mode #n

- **Concurrent Mixed Modes**
  - Mode #1 ➔ Mode #2 ➔ Mode #n
Mixing Modes

- Data collection often involves trade-off between the stronger and weaker points of each mode and method.

- Mixed modes survey are appealing but have risks and inherent issues:
  - measurement error
  - cost considerations
  - bias
Survey Management Considerations for Mixed Mode

- Survey Design
  - Multiple sample frames
  - Types of contact and modes
  - Sequence of modes
  - Switching modes
  - Propensity models and responsive design
  - Staffing and resource management

- Sample delivery
  - Parameter/rules-based
  - Often link sample to mode of collection
  - Sample element only available to one ‘location’ at a time
Agenda

• Questionnaire Design Challenges
• Survey Management Challenges
• “New” Technology Challenges
The Trends

• Rising smartphone and internet usage creates a viable mode for survey data collection and needs formal investigation (Buskirk and Andrus, 2012)

• Recent study found 23% respondents completed the internet survey via mobile, even though an attempt was made to redirected Rs (Wells, Bailey, & Link, 2012)

• The Pew Research Center Report (Smith, 2012)
  – Smartphone ownership grew 11% in just nine months to 46%
  – 17% of all adult mobile phone owners mostly access the internet via their device only
  – For 10%, their phone is their only option for online access
  – 31% of American adults own a tablet computer
More bad news than good news

- Optimizing design of web surveys for so many devices, OS versions, and browsers
- Usability of the survey instrument
- Connectivity (and efficiency)
- Mobile app programming
- Survey sample management
- Data transmission and security
- Survey preload and paradata collection
- Quality assurance procedures
- Optimizing other mobile components to enhance data collection
- Methodological implications of using mobile technology
Questions to ask us

• Will off-the-shelf “iCAPI” /”iCollector” type of survey development software provide capability to design effective, tailored instruments?

• Does the depreciation of the mobile devices present a cost-prohibitive driver for expanded use?

• Overcome all the usability's concerns for the field data collectors?
Social Media (Twitter, Facebook...)

• Purpose: Service for building & reflecting social connections & communications

• Current some uses in Survey Research:
  ✓ Locating respondents
  ✓ Question testing
  ✓ Focus group recruitment
  ✓ Study “Groups”

• “Big Data” is very hot topic!!!
Final Comments

• Rapid and continuous change: new technologies and new approaches to collect data making dramatic changes in our survey designs (multiple and mixed mode data collection)

• Face some old issues: COVERAGE, SAMPLING, MEASUREMENT ERROR, NONRESPONSE, DIFFERENTIAL NONRESPONSE

• New opportunities & challenges for social survey researchers
Thank you!

EMAIL: QIANYANG@UMICH.EDU