Collecting Rich Paradata to Monitor Data Collection Quality in Challenging Contexts

Beth-Ellen Pennell, Gina-Qian Cheung, Patty Maher, Zeina Mneimneh, Jennifer Kelley, Jay Lin

University of Michigan
Outline

• Paradata collection

• Examples all highlight rich paradata available in Blaise

• Three projects in three countries

• Issues and challenges
Paradata includes...

- Interviewer (experience, training grades, historical performance)
- Sample segments (PSU, stratum, observations)
- Address (probability of selection, observations, # contacts, status)
- Screener contacts (call #, interviewer, time, date, informant behavior, outcome)
- Household (composition, informant behavior, sample respondent characteristics)
- Main interview contacts (call #, interviewer, time, date, informant behavior, outcomes)
Paradata also includes...

- Audit trails
  - Screener and survey interview (keystrokes, timings, functions, consistency checks, suspensions)
  - Sample management system (log and timing of actions)
- Digital photos
- Fingerprints
- GPS (Global Positioning System)
- Digital recordings
- Collection of various anthropometric data using digital devices
Three Countries; Three Projects

• China: China Family Panel Study and beyond

• Ghana: Ghana Socioeconomic Panel Survey

• Nepal: Nepal Household Registry Project
China: Peking University’s Institute for Social Science Survey (iSSS)
2010-2020 China Family Panel Study

- Omnibus panel survey covering: social well-being; economic well-being; demographic outcomes; education; health

- National in scope:
  - 16,000 households; 56,000 individuals
  - All household members 10 years of age and older
  - More than 88,000 interviews completed in first wave
  - First large scale CAPI study conducted in China
China Family Panel Interview
**Blaise Paradata**

### 主要工作
- **8.1** 公安干警、交通干警
  - **8.1.811** 公安、交警机构局级及以上级别的干部

### 主要职业
1. 国家机关、党群组织、企业、事业单位负责人
2. 专业技术人员
3. 办事人员和有关人员
4. 商业工作人员
5. 服务性工作人员
6. 农、林、牧、渔、水利业生产人员
7. 生产工人、运输工人和有关人员
8. 警察及军人

**刘建新 男 32岁 未婚 无**

<table>
<thead>
<tr>
<th>姓名</th>
<th>性别</th>
<th>年龄</th>
<th>婚姻状况</th>
<th>学历</th>
</tr>
</thead>
<tbody>
<tr>
<td>刘建新</td>
<td>男</td>
<td>32岁</td>
<td>未婚</td>
<td>无</td>
</tr>
</tbody>
</table>

**代码：8.1.811**

© 2016 by the Regents of the University of Michigan
China Health and Retirement Longitudinal Survey
China National Mental Health Study

• National probability sample
  – 30,000 interviews
• Interviews are audio-recorded
• Paradata used to stratify quality control – optimize resources
  – 15% validation; 25% evaluation
• Also includes physical measurements as well as saliva and blood collection
Quality Control Optimization

All Completed Cases

Data Assessment

Flagged

Suspicious cases

Evaluation failed

Call back verification

Random selective cases

Call back verification failed

Evaluation

Alternative cases
Ghana: Ghana Socioeconomic Panel Survey

• Sponsored by Yale University and carried out by the Institute for Statistical, Social and Economic Research (ISSER) at the University of Ghana
• First wave was completed on paper between October 2009 and February 2010
• Sample size of 5009 households, with approximately 18,000 individuals
• 334 enumeration areas country-wide
• 55,000+ variables; 8 hours of survey content
• Programmed in Blaise
• The plan is to revisit these households at 3-4 year intervals for 20 years
## 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>N1.1</th>
<th>N1.2</th>
<th>N1.3</th>
<th>N1.4</th>
<th>N1.5</th>
<th>N1.6</th>
<th>N1.7</th>
<th>N1.8</th>
<th>N1.9</th>
<th>N1.10</th>
</tr>
</thead>
<tbody>
<tr>
<td>Plot</td>
<td>Crop Type</td>
<td>Unit of measurement</td>
<td>Have you harvested CROP TYPE from PLOT A?</td>
<td>How many UNIT of CROP TYPE has been harvested from PLOT A?</td>
<td>In which month(s) did you harvest the CROP TYPE from PLOT A?</td>
<td>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)</td>
<td>While still in the field, was any of your CROP TYPE from PLOT A damaged by pests, rotting, or any other reason?</td>
<td>What was the crop lost to?</td>
<td>How much of your crop was lost in total?</td>
</tr>
<tr>
<td>------</td>
<td>---------</td>
<td>----------------------</td>
<td>------------------</td>
<td>------------------</td>
<td>------------------</td>
<td>--------------------------</td>
<td>------------------</td>
<td>----------------</td>
<td>------------------</td>
</tr>
<tr>
<td>A</td>
<td>Rice</td>
<td>Bag</td>
<td>2 Yes, fully 2 Yes, partially 3 No</td>
<td>2 Yes, fully</td>
<td>1 Yes, fully (no CROP TYPE is remaining on PLOT A)</td>
<td>Blaise preload months</td>
<td>Use units from N1.3</td>
<td>1 Yes</td>
<td>Enter Percentage (%)</td>
</tr>
<tr>
<td>A</td>
<td>Maize</td>
<td>Bag (dried without cobs and processed)</td>
<td>2 Yes, fully</td>
<td>1 Yes, fully</td>
<td>1 Yes, fully (no CROP TYPE is remaining on PLOT A)</td>
<td>Blaise preload months</td>
<td>Use units from N1.3</td>
<td>1 Yes</td>
<td>Enter Percentage (%)</td>
</tr>
<tr>
<td>A</td>
<td>Miller</td>
<td>Bag</td>
<td>2 Yes, fully</td>
<td>1 Yes, fully</td>
<td>1 Yes, fully (no CROP TYPE is remaining on PLOT A)</td>
<td>Blaise preload months</td>
<td>Use units from N1.3</td>
<td>1 Yes</td>
<td>Enter Percentage (%)</td>
</tr>
<tr>
<td>A</td>
<td>Sorghum</td>
<td>Bag</td>
<td>2 Yes, fully</td>
<td>1 Yes, fully</td>
<td>1 Yes, fully (no CROP TYPE is remaining on PLOT A)</td>
<td>Blaise preload months</td>
<td>Use units from N1.3</td>
<td>1 Yes</td>
<td>Enter Percentage (%)</td>
</tr>
<tr>
<td>A</td>
<td>Groundnut</td>
<td>Bag (without shell)</td>
<td>2 Yes, fully</td>
<td>1 Yes, fully</td>
<td>1 Yes, fully (no CROP TYPE is remaining on PLOT A)</td>
<td>Blaise preload months</td>
<td>Use units from N1.3</td>
<td>1 Yes</td>
<td>Enter Percentage (%)</td>
</tr>
<tr>
<td>A</td>
<td>Soya bean</td>
<td>Bag (without shell)</td>
<td>2 Yes, fully</td>
<td>1 Yes, fully</td>
<td>1 Yes, fully (no CROP TYPE is remaining on PLOT A)</td>
<td>Blaise preload months</td>
<td>Use units from N1.3</td>
<td>1 Yes</td>
<td>Enter Percentage (%)</td>
</tr>
<tr>
<td>A</td>
<td>Cassava</td>
<td>Bag</td>
<td>2 Yes, fully</td>
<td>1 Yes, fully</td>
<td>1 Yes, fully (no CROP TYPE is remaining on PLOT A)</td>
<td>Blaise preload months</td>
<td>Use units from N1.3</td>
<td>1 Yes</td>
<td>Enter Percentage (%)</td>
</tr>
<tr>
<td>A</td>
<td>Yam</td>
<td>Bowl (100 rubers)</td>
<td>2 Yes, fully</td>
<td>1 Yes, fully</td>
<td>1 Yes, fully (no CROP TYPE is remaining on PLOT A)</td>
<td>Blaise preload months</td>
<td>Use units from N1.3</td>
<td>1 Yes</td>
<td>Enter Percentage (%)</td>
</tr>
<tr>
<td>A</td>
<td>Cowpea</td>
<td>Bag</td>
<td>2 Yes, fully</td>
<td>1 Yes, fully</td>
<td>1 Yes, fully (no CROP TYPE is remaining on PLOT A)</td>
<td>Blaise preload months</td>
<td>Use units from N1.3</td>
<td>1 Yes</td>
<td>Enter Percentage (%)</td>
</tr>
<tr>
<td>A</td>
<td>Okra</td>
<td>Bag</td>
<td>2 Yes, fully</td>
<td>1 Yes, fully</td>
<td>1 Yes, fully (no CROP TYPE is remaining on PLOT A)</td>
<td>Blaise preload months</td>
<td>Use units from N1.3</td>
<td>1 Yes</td>
<td>Enter Percentage (%)</td>
</tr>
<tr>
<td>A</td>
<td>Pepper</td>
<td>Bag</td>
<td>2 Yes, fully</td>
<td>1 Yes, fully</td>
<td>1 Yes, fully (no CROP TYPE is remaining on PLOT A)</td>
<td>Blaise preload months</td>
<td>Use units from N1.3</td>
<td>1 Yes</td>
<td>Enter Percentage (%)</td>
</tr>
<tr>
<td>A</td>
<td>Watermelon</td>
<td>Bowl</td>
<td>2 Yes, fully</td>
<td>1 Yes, fully</td>
<td>1 Yes, fully (no CROP TYPE is remaining on PLOT A)</td>
<td>Blaise preload months</td>
<td>Use units from N1.3</td>
<td>1 Yes</td>
<td>Enter Percentage (%)</td>
</tr>
<tr>
<td>A</td>
<td>Mangoes</td>
<td>Box</td>
<td>2 Yes, fully</td>
<td>1 Yes, fully</td>
<td>1 Yes, fully (no CROP TYPE is remaining on PLOT A)</td>
<td>Blaise preload months</td>
<td>Use units from N1.3</td>
<td>1 Yes</td>
<td>Enter Percentage (%)</td>
</tr>
<tr>
<td>A</td>
<td>Tomato</td>
<td>Box</td>
<td>2 Yes, fully</td>
<td>1 Yes, fully</td>
<td>1 Yes, fully (no CROP TYPE is remaining on PLOT A)</td>
<td>Blaise preload months</td>
<td>Use units from N1.3</td>
<td>1 Yes</td>
<td>Enter Percentage (%)</td>
</tr>
<tr>
<td>A</td>
<td>Salad</td>
<td>Box</td>
<td>2 Yes, fully</td>
<td>1 Yes, fully</td>
<td>1 Yes, fully (no CROP TYPE is remaining on PLOT A)</td>
<td>Blaise preload months</td>
<td>Use units from N1.3</td>
<td>1 Yes</td>
<td>Enter Percentage (%)</td>
</tr>
<tr>
<td>A</td>
<td>Cotton</td>
<td>Bag (different from maize bag)</td>
<td>2 Yes, fully</td>
<td>1 Yes, fully</td>
<td>1 Yes, fully (no CROP TYPE is remaining on PLOT A)</td>
<td>Blaise preload months</td>
<td>Use units from N1.3</td>
<td>1 Yes</td>
<td>Enter Percentage (%)</td>
</tr>
<tr>
<td>A</td>
<td>Other</td>
<td>2 Yes, fully</td>
<td>1 Yes</td>
<td>Enter Percentage (%)</td>
<td></td>
<td></td>
<td></td>
<td></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
- 11 Grasshoppers
- 12 Locust
- 13 Termites
- 14 Caterpillars
- 15 Centipedes
- 16 Salig
- 17 Pandocirisi
- 18 Zanax
- 19 Tambepeglu
- 20 Rodents
- 21 Rice
- 22 Rats
- 23 Grasscutters
- 24 Squirrels
- 25 Monkeys
- 26 Sheep
- 27 Goats
- 28 Cattle
- 29 Bandicoot
- 30 Other
- (SPECIFY)
Benefits of this Design

• Interviewers have a high-level of autonomy with respect to interview navigation
• Interviewers are able to switch respondents easily
• Interviewers are able to jump to any section of questionnaire quickly
• Need for questionnaire “Dashboard” to show the status of all the questionnaire sections and all the respondents within the household
Questions about the design...

• How does instrument design affect instrument navigation?

• How does interview navigation affect interview length?
  – Order of interview initiation
  – Movements between blocks
Most Common Block Moves All Types

Edge Weight >= 500
Movement within sections dominates
Exceptions are rosters and Personal to Household
Nepal: Household Registry Project

- Longitudinal project conducted by the Institute for Social and Environmental Research (ISER) in Nepal since 1996
- Life histories for more than 10,000 individuals
- Continuous measurement of community change both within the original study neighborhoods and with respondents
- Track those who also move out of those neighborhoods (now across more than 50 countries)
Nepal: PAPI vs. CAI Data Collection
<table>
<thead>
<tr>
<th>S.N.</th>
<th>ID #</th>
<th>Living</th>
<th>Marital Status</th>
<th>Pregnancy Status</th>
<th>Education</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Age</td>
<td>Jan</td>
<td>Feb</td>
<td>March</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

***For interviewers***:

- See the last month information and ask the appropriate questions:
- **D** = Died
- **X** = Unmarried
- **S** = Separated due to quarrel
- **N** = No Pregnant
- **M** = Miscarriage
- **B** = Birth
- **A** = Abortion
- **F** = Pregnant
- **W** = Widow
- **T** = Not in school
- **I** = In school

Last time we updated your household information in .......... month .......... Year. This time, I am here to update your household information since that time. I would also like to add if there is any member of your household we left out in the household listing last time, who may have been away from home at that time. First let me start with people who were in your household last time and I will ask about those who were left out time:

A. At that time .......... was living in this household, is he/she still living or not?
B. Last time .......... was away from home is he/she still living away or started living here in your household?
C. Besides these people, is there anybody came to live in your household?
D. Now, are there any member of your household not included above? If there are please tell me their names: Arjuna, Bala, Devi.

***For interviewers***:

- Ask the questions to above 12 years old respondent:
- A. Last month .......... was unmarried/living (together/ separate) with his/her spouse, widow, what about now?
- B. Besides these is there anybody who got married, living together/ separate, widow, divorced or eloped?

Now let’s talk about schooling. By in school we mean currently going to school, in vacation, between school, adult education & training and preparing for exam. Using this criteria, how about .......... schooling? Is .......... in school or not in school?

***For interviewers***:

- Ask the questions to 5 to 34 years old respondent:
- A. Last month .......... was pregnant, not pregnant, What about now?
- B. Besides these is there anybody in your household get pregnant, gave birth, still born, miscarriage or abortion?

2. Now, I will read the information to you I record above. Please tell me if I have not recorded correctly.

3. Finally, did I left any member of your household to record here?

Thank you very much for your precious time and co-operation. Information given by you is very important for us. We will come again for more information. O.K. bye, Namaskar!
### Household Member Status:

<table>
<thead>
<tr>
<th>SN</th>
<th>Name</th>
<th>Registry</th>
<th>Living</th>
<th>Marital</th>
<th>Pregnancy</th>
<th>Education</th>
</tr>
</thead>
<tbody>
<tr>
<td>001</td>
<td>सुनै गहरले</td>
<td>Eligible</td>
<td>Eligible</td>
<td>Eligible</td>
<td></td>
<td></td>
</tr>
<tr>
<td>002</td>
<td>सन्त कृमारी महरले</td>
<td>Eligible</td>
<td>Eligible</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>003</td>
<td>प्रदेव महरले</td>
<td>Eligible</td>
<td>Eligible</td>
<td></td>
<td>Eligible</td>
<td></td>
</tr>
<tr>
<td>004</td>
<td>प्रमिला महरले</td>
<td>Eligible</td>
<td>Eligible</td>
<td></td>
<td></td>
<td>Eligible</td>
</tr>
<tr>
<td>005</td>
<td>संहिसा महरले</td>
<td>Eligible</td>
<td>Eligible</td>
<td></td>
<td>Eligible</td>
<td></td>
</tr>
<tr>
<td>006</td>
<td>विकस महरले</td>
<td>Eligible</td>
<td>Eligible</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>007</td>
<td>सुजिता महरले</td>
<td>Eligible</td>
<td>Eligible</td>
<td></td>
<td>Eligible</td>
<td></td>
</tr>
<tr>
<td>008</td>
<td>स्मृति महरले</td>
<td>Eligible</td>
<td>Eligible</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>009</td>
<td>शाकिन्त महरले</td>
<td>Eligible</td>
<td>Eligible</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>010</td>
<td>प्रकाश महरले</td>
<td>Eligible</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- **Household Type:** Intact
- **Household Location Status:** Eligible
- **Outstanding!** The Registry tab has been completed, please finish any items that are not completed.
Migration

53 countries

Destination

59 districts
Challenges: 1. Survey (re)design

• Transition from a well-defined paper & pencil questionnaire to a computer assisted interview (CAI) instrument
  ▪ Complex grid structures
  ▪ Preload previous data collection/sample frame data
  ▪ Question fills (greater complexity with multiple languages)

• CAI testing

• CAI training (new data collection procedures)
Challenges: 2. Technology

- Mobile phone based technology (bandwidth can be very narrow)
- Modifying matured and tested systems to fit each country’s project specific requirements
  - Off-line transfer of cases capabilities
  - Off-line distribution of sample
  - Able to switch respondents easily (multiple respondents)
  - Able to jump to any section of questionnaire quickly
  - Development of a questionnaire “dashboard” to show the status of all the questionnaire sections and all the respondents within the household
- Adapt country-specific equipment (e.g., laptops) and systems (e.g., operating systems)
- Logistically difficult to carry out system integration testing
- Development of remote production support
Challenges: 3. Internationalization

• Languages
• Terminology
• Time zones
• Conversational interviewing techniques, group/team interviews
• Staff skill and experience
• Training – practice – training – production (more training)
• Data collection monitoring shifts from reliance on monitoring survey data to including analysis of paradata once we shift from PAPI to CAPI
Lessons Learned

• Survey (re)design is time consuming and significant time needs to be allotted
  – but time is gained on back end and marked increase in quality
• Need to balance between standardization and localization of approach and methods - last mile variation
• Multiple trainings needed throughout lifecycle
• Understanding differences and reacting quickly
  – Multiple types of communication channels: landline/cell phone calls, E-mail, Skype, Viber, WhatsApp, Text Messaging, and Facebook Messenger
Welcome to the Cross-cultural Survey Guidelines!

These Guidelines were developed as part of the Comparative Survey Design and Implementation (CSDI) Guidelines Initiative. The aim of the Initiative was to promote internationally recognized guidelines that highlight best practice for the conduct of multinational, multicultural, or multiregional surveys, which we refer to as "SMC" surveys. The intended audience is researchers and survey practitioners planning or engaged in comparative survey research across cultures or countries.

See Chapters to access the guidelines, which cover all aspects of the survey lifecycle.

The CCSG Guidelines draw upon and are based on: (1) general good practice survey methodology, as well as cross-cultural and comparative literature on survey methodology; (2) available study-specific manuals and documentation; and (3) the experiences and lessons learned by authors, reviewers, and editors that have emerged through their work on and with numerous comparative surveys. Best practices are dynamic and can be expected to evolve over time.

We welcome your feedback. New methodological research and additional documentation about comparative surveys will inform future versions of the CCSG Guidelines. You can greatly help us by providing comments and suggestions, or simply alerting us about a topic we need to address. Please Contact Us.

Janet A. Harkness Student Paper Award

We dedicate these guidelines to Dr. Janet A. Harkness. Dr. Harkness passed away in 2012. She initiated the International Workshop on Comparative Survey Design and Implementation were the development of these Guidelines was launched. Dr. Harkness not only contributed to the overall framework and content of the guidelines but she also authored three of the key chapters: Questionnaire Design, Adaptation and Translation. She inspired this work through her steadfast conviction that resources must be made available to researchers and survey practitioners if we are to improve comparative survey research methods, dissemination and analysis.

The World Association for Public Opinion Research (WAPOR) and the American Association for Public Opinion Research (AAPOR) jointly support Dr. Harkness's work by sponsoring the Janet A. Harkness Student Paper Award each year. This award recognizes "emerging young scholars in the study of multi-national/multi-cultural/multilingual survey research (aka SMC survey research) through support of the winner's participation in the WAPOR Conference and a cash prize." Please consider...
The Cross-cultural Survey Guidelines (CCSG) have been developed to cover all aspects of the lifecycle of multinational, multicultural, or multiregional (3MC) surveys, as shown in the figure below. Navigate to the chapters with guidelines by clicking on a stage in the survey lifecycle in the figure or by using the menu to the left. The lifecycle begins with establishing aspects of the study design and organizational structure (Study Design and Organizational Structure) and ends with data dissemination (Data Dissemination). In some study designs, the lifecycle may be completely or partially repeated. There might also be iteration within a production process. The order in which survey production processes are shown in the lifecycle does not represent a strict order to their actual implementation, and some processes may be simultaneous and interlocked (e.g., sample design and contractual work). Quality and ethical considerations are relevant to all processes throughout the survey production lifecycle.