Module 6: Data quality monitoring
Video 3 of 6: Data sources

Sharan Sharma
Possible data sources to monitor interviewers

1. GPS coordinates
2. Reinterviews
3. Listening-in
4. Mystery respondents
5. Substantive data
6. Paradata
7. Audio recordings (CARI)
1. Global Positioning System (GPS) data

- Advent of portable/mobile devices to conduct interviews.

- Capture the location of interviewer at different time points.

- Can be automatically collected or require the interviewer to activate.

- Mainly used to detect falsification.
But the CATI interviewer said he did all his calls from his home.

Can supplement basic maps with other information. Are these Coffee-shops?
Geo-fencing

- Can disable the survey program if location is outside.
- Trigger can also cause supervisor to call the interviewer.
2. Reinterviews

- Traditional form of quality control (‘backchecks’). Done for quality control reasons and sometimes to estimate response variance.

- Design:
  - Reinterview done after the interviewer has registered a ‘complete’.
  - Select a sample to reinterview.
  - Call the respondents and ask the survey questions again.
  - Compare reinterview responses to original responses.
Reinterview design considerations

A. How should we sample?

• Random sample e.g. European Social Survey: 10% of each interviewer’s completed interviews and 5% for noncompleted interviews including noncontacts, refusals, and ineligibles.

• Sample based on some characteristics. Examples:
  – All interviews that were completed in less than 10 minutes.
  – Hood and Bushery (1997) find that doing a ‘focused reinterview’ was more efficient than a random sample.
<table>
<thead>
<tr>
<th>Time Span</th>
<th>Interviewers Checked</th>
<th>Suspected Falsification</th>
<th>Cleared</th>
<th>Confirmed</th>
<th>Investigation Pending</th>
</tr>
</thead>
<tbody>
<tr>
<td>Focused 14 Weeks</td>
<td>83</td>
<td>7</td>
<td>1</td>
<td>3 (3.6%)</td>
<td>3</td>
</tr>
<tr>
<td>Random Yearly Avg. 1983-1996</td>
<td>400 +</td>
<td>1.9</td>
<td>0.8</td>
<td>1.1 (0.2%)</td>
<td>0</td>
</tr>
</tbody>
</table>

Four measures used to identify interviewers: Ineligible Unit Rate, Nonminority Screening Rate, Short Interview Rate, No Telephone Number Rate

*Source: Hood and Bushery (1997)*
Reinterview design considerations...

B. Which questions should we re-ask?

• Do not frame the introduction as “I want to check if the [NAME] did a good job” but something like “Thank you for your response when [NAME] visited you. I hope everything was fine. I just wanted to make sure of some questions”.

• Start by verifying if the original interview took place at all or not and whether the incentive (if any) was paid.
In the context of quality control, reinterviews should preferably focus on…

- Factual questions e.g. How many children do you have?
  - Especially useful to detect falsification.
  - In countries where official birth records are relatively new, “How old are you?” may not be a ‘factual’ question.

- Have a history of problematic responses e.g. How many rooms do you have at home?

- Screener questions e.g. Do you own any farmland? [answering “yes” leads to a long list of farm questions]
...and not rely on:

- ‘Soft’ attitude/opinion questions e.g. On a scale of 1 to 5, where 1 is “very low” and 5 is “very high”, how satisfied are you with your life?
- Sensitive questions e.g. Questions on domestic violence.
- Overly time-dependent questions e.g. [to a student] How many minutes did you spend on your homework yesterday?
- Not overly respondent-specific (may not get the same respondent for the reinterview.)
Reinterview questions...

1. If the instrument has a large component of Attitude/Opinion/Behavior (AOB) questions, choose those with relatively ‘harder’ AOB e.g. “Do you think women have a right to the same salary as men?”

2. If must ask time-dependent questions, then these may need reframing e.g. “Do you work from home?” framed as “Did you work from home in May 2020”. Especially important if, e.g., there was a COVID lockdown in force in May 2020.
C. Reinterview design considerations

1. Length of the reinterview
   - Typically not more than 15 minutes.
   - For a panel survey, even more important to not add to respondent burden.

2. Implicit assumption that the ground reality has not changed since the original survey was fielded.
   e.g. a knowledge question, “Is malaria spread via injections?”, Assumption → respondent has not acquired new knowledge.
What should be the gap between the original interview and reinterview?

- Time lag should not be so far that responses are not comparable (need same ground reality).
- Time lag should also not be so small that the respondent merely repeats what was original said + adds to respondent irritation (especially for panel surveys).
- Forsman and Schreiner (2004) report on a pilot study in Sweden:
  - Called back 20 respondents after 4 weeks.
  - None of them remembered their original answers to questions on living conditions.
  - Led to the decision to do the reinterview 3 weeks after the original interview.
- 1 – 3 weeks seems a reasonable lag.
Reinterview design considerations…

3. Reinterviewers:
   – Should be skilled and experienced.
   – Should be persons who are not the original interviewers --- avoid simply having an interviewer be the reinterviewer for another interviewer’s respondents (interviewers can be connected) or even the interviewer’s immediate supervisor.
   – Ideal to match reinterviewer’s gender to the gender of the original interviewer.
   – Unless really required, reinterviewers should not (initially) have access to the original response. Typically not told in advance about assignments.
Reinterview design considerations…

4. May not always get the original respondent for the reinterview.
   – Proxy responses always subject to comparison issues.
   – Reframing may help, e.g., “Did [Name] work from home in May 2020?”. But will not help as much for Attitude/Opinion questions.
Reinterview design considerations…

5. Discrepancies could well be due to a badly worded question, the respondent herself/himself (e.g. motivation, recall capabilities, etc. vary), or even the reinterviewer.

6. Reconciliation
   • Process of determining why a difference exists (if it does).
   • One method to do this:
     – The reinterviewer is initially not given the original response (as mentioned on Slide 15).
     – Reinterviewer conducts the reinterview.
Reinterview design considerations…

6. Reconciliation…

– CATI software flags any differences between the original response and reinterview response. For continuous variables e.g. income, can set a threshold of say, 10% difference for any difference to be flagged.
– The reinterviewer is then presented both responses on screen.
– In some cases, the reinterviewer can detect clear falsification. Else further probes are undertaken to establish why the difference occurred.
  • e.g. for Roster questions, in the case of an extra member included in the re-interview one may ask “When did this person join this household ?”
– Reconciliation also used to identify problems with the question (not the focus of this module) and a third ‘reconciliation interview’ sometimes undertaken for this. Please see Morton et al (2008) for an example.
Reinterview design considerations…

• Look for patterns rather than worry too much about individual data points e.g. isolate interviewers who have more than 33% unmatched answers and then subject their work to a deeper investigation.
Possible data sources to monitor interviewers

1. GPS coordinates
2. Reinterviews
3. Listening-in
4. Mystery respondents
5. Substantive data
6. Paradata
7. Audio recordings (CARI)
3. Listening-in

- Supervisor or monitor uses a parallel connection to *unobtrusively* listen to a live interview → Real-time monitoring. Should not interrupt the interview.

- Historically comes from centralized telephonic facilities. But can be done even in non-centralized contexts with the right technology.

- Especially useful at the initial survey stages and monitoring inexperienced interviewers.
Listening-in: Design issues to consider

1. Which interviewers to monitor?
   • Ideally, every interviewer gets monitored – but at possibly different rates.
   • At least a few cases (e.g. 10) monitored for each interviewer → makes it generalizable; better footing for the monitor to give feedback.
   • Couper et al (1992) divide CATI interviewers into 3 groups based on past performance and experience:
     – Strong: 6+ months experience and above-median performance rating.
     – Average: Remaining 6+ months experience interviewers.
     – New: Less than 6 months experience.
     – Selection ratios → 1: 1.5 : 2
Differential selection strategy seems to work…

<table>
<thead>
<tr>
<th>Component</th>
<th>Strong</th>
<th>Average</th>
<th>New</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. of questions monitored</td>
<td>1349</td>
<td>1040</td>
<td>1361</td>
</tr>
<tr>
<td>Major wording changes</td>
<td>1.3%</td>
<td>0.7%</td>
<td>2.3%</td>
</tr>
<tr>
<td>Minor wording changes</td>
<td>3.3%</td>
<td>6.8%</td>
<td>3.4%</td>
</tr>
<tr>
<td>Incomplete questions</td>
<td>0.2%</td>
<td>0.8%</td>
<td>3.2%</td>
</tr>
<tr>
<td>Directive probing</td>
<td>0.2%</td>
<td>0.4%</td>
<td>1.2%</td>
</tr>
</tbody>
</table>

Adapted from Couper et al 1992
Listening-in: Design issues to consider

2. Which respondents to monitor?

- Rule-based selection. Do not allow complete monitor discretion; some judgement is fine for monitors with good judgment.

- Initial selection can be a mix of respondent profiles for each interviewer. Then adaptively change selection rates for each interviewer based on performance e.g. some interviewers may struggle with respondents of the opposite gender.
Listening-in: Design issues to consider

3. Who should monitor?
   - Very often monitors are experienced interviewers themselves.
   - Tarnai (2007) uses one monitor for every 10 interviewers.

4. What to monitor?
   - Unit refusal avoidance
     - Hard refusals: Sample member explicitly states they do not want to participate
     - Soft refusals: “I’m not available now” or other hints
   - Rapport building
   - Interviewing techniques – not just asking, probing etc. but also voice features (too many fillers e.g. uh, um?)
   - Process needs to be structured, objective, with results well documented. Tarnai (2007) and Couper (1992) provide the actual list/forms used in practice.
<table>
<thead>
<tr>
<th>What Went Well:</th>
<th>Info Given</th>
<th>Accurate</th>
<th>Incomplete</th>
<th>False</th>
<th>N/A</th>
</tr>
</thead>
<tbody>
<tr>
<td>Refusal Prev.</td>
<td>Persuasive</td>
<td>Weak</td>
<td>Fail</td>
<td></td>
<td>N/A</td>
</tr>
<tr>
<td>Pace</td>
<td>Satisfactory</td>
<td>Slower</td>
<td>Faster</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Volume</td>
<td>Satisfactory</td>
<td>Louder</td>
<td>Softer</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Clarity</td>
<td>Satisfactory</td>
<td>Enunciate</td>
<td>Pronounce</td>
<td></td>
<td></td>
</tr>
<tr>
<td>What Could Be Improved:</td>
<td>Courtesy</td>
<td>Satisfactory</td>
<td>Unsatisfactory</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rapport</td>
<td>Satisfactory</td>
<td>Unsatisfactory</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Conclusion</td>
<td>Satisfactory</td>
<td>Unsatisfactory</td>
<td>N/A</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Professionalism</td>
<td>Satisfactory</td>
<td>Too Stiff</td>
<td>Too Casual</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FAQ’s</td>
<td>Satisfactory</td>
<td>Unsatisfactory</td>
<td>N/A</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Extract from Tarnai (2007)
5. Monitoring shifts and sessions

- Monitoring “shift” → Continuous period of time that a monitor is scheduled to monitor interviewers.
  - Decide on the number of shifts in a day. If monitors are also performing other survey tasks, spread out shifts for a monitor within a day.

- Pre-assign monitors to shifts.

- A monitor shift could have varying number of monitoring “sessions” → Periods of time that a monitor listens to a specific interaction between interviewer and respondent.
Listening-in: Design issues to consider…

5. Monitoring shifts and sessions

- Introduce monitor rotation to reduce bias
  - Different times of the day (also weekday/weekend)
  - Different interviewers listened-into.
  - If shifts are not in place (e.g. decentralized interviewers working at their own pace), at least make sure to rotate monitors across interviewers.
Listening-in: Design issues to consider…

6. Is it necessary to monitor the entire interview?

- Some empirical evidence shows that monitoring partial interviews at least as good as monitoring the full interview.

- In fact, monitor fatigue can occur. Tarnai (2007) finds:
  - Partial interview: 1 error for every 7 questions
  - Full interview: 1 error for every 12 questions
# How do CATI interviewers feel about being listened-into?

<table>
<thead>
<tr>
<th>Question</th>
<th>% Positive</th>
<th>% Neutral Not Sure</th>
<th>% Negative</th>
</tr>
</thead>
<tbody>
<tr>
<td>What are your overall feelings about this monitoring session?</td>
<td>93%</td>
<td>5%</td>
<td>2%</td>
</tr>
<tr>
<td>How helpful was the Monitor’s feedback?</td>
<td>92%</td>
<td>1%</td>
<td>7%</td>
</tr>
<tr>
<td>How professional or unprofessional would you say the Monitor was in giving you feedback about your performance?</td>
<td>98%</td>
<td>0%</td>
<td>2%</td>
</tr>
<tr>
<td>Will the Monitor’s feedback help you to improve your interviewing skills?</td>
<td>89%</td>
<td>3%</td>
<td>8%</td>
</tr>
<tr>
<td>How strongly do you agree or disagree that monitoring is necessary to improve telephone interview performance?</td>
<td>92%</td>
<td>0%</td>
<td>8%</td>
</tr>
<tr>
<td>What is your opinion of being monitored in general?</td>
<td>63%</td>
<td>23%</td>
<td>14%</td>
</tr>
</tbody>
</table>

Source: Tarnai 2007
Listening-in: Design issues to consider…

7. Monitoring the monitors

To make sure that monitoring is: a) Constructive b) Bias-free c) Result-oriented. Check:

a) **Capturing of interviewing behavior.**

Survey managers should independently listen to some cases (especially that of interviewers at the extremes of the performance distribution) or get some cases listened-in to by more than one monitor and compare (similar to inter-reliability studies).

a) **Communication of feedback to the interviewer.**

Survey managers should check the quality of feedback provided and give monitors feedback, e.g., a tendency for some monitors to make definitive judgmental statements based on partial listening.

- Encourage monitors to discuss uncertain cases.
Possible data sources to monitor interviewers

1. GPS coordinates
2. Reinterviews
3. Listening-in
4. Mystery respondents
5. Substantive data
6. Paradata
7. Audio recordings (CARI)
4. Mystery respondents

• Include monitors as sample IDs in the CATI call scheduler.

• Monitors will then act as respondents, unknown to the interviewer → Real-time monitoring.

• Best to audio record the interview and then review. Ensures that the monitor is fully focused on playing the respondent role.

• Can also pre-plan responses so that the questionnaire takes a known routing.
Mystery respondents…

• Advantages:
  – The monitor can create situations to test specific interviewing skills in a live setting e.g. give an uncodable initial response that should prompt the interviewer to probe.
  
  But Important not to get carried away and harass the interviewer - the idea is not to be a nasty/extreme respondent.

• Drawbacks:
  – A monitor can typically be interviewed only once by a specific interviewer.
  – Monitor’s voice should be unrecognizable by the interviewer.
  – The monitor can be identified by her/his number if that is a personal telephone to which the call is being made.
Possible data sources to monitor interviewers

1. GPS coordinates
2. Reinterviews
3. Listening-in
4. Mystery respondents
5. Substantive data
6. Paradata
7. Audio recordings (CARI)
END OF VIDEO 3