Considerations in Data Harmonization for Multimode Surveys

Hanna Popick, Mina Muller, and Eric Jodts

FedCASIC virtual conference, April 2021

Introduction

- Multimode surveys can take advantage of each mode's strengths
- However, resulting data may have inconsistencies across modes
- What decisions can you make to address these inconsistencies?
- Resolutions affect format and use of final datasets

Agenda

Data design

Data editing for question level inconsistencies

Data editing for survey level inconsistencies

Dataset harmonization

Data design

- Key question that drives the considerations we will discuss:
  - How will the completed data be used in analysis, and what format of final delivery will best support that analysis?
- Your project benefits when sample management, survey modes, instrument development, and plans for data editing are determined up front

Inconsistencies across modes

› Why do inconsistencies arise across modes?

Ability	Available in CAI survey?	Available in paper survey?
Enforce the desired number of responses to a question		
Display only the questions that the respondent should answer		
Apply consistency checks during the survey		

Agenda

Data design

Data editing for question level inconsistencies

Data editing for survey level inconsistencies

Dataset harmonization

Enforcing the number of responses: Example questions

Question 1: What factor is the most important to you when selecting a health care plan?

- ❑Premium
- ❑Deductible
- ❑Doctors in plan

Question 2: What is your highest level of education?

- ❑High school graduate or less
- ❑College graduate
- ❑Post-graduate degree

Data editing for question level inconsistencies

Ability	Available in CAI survey?	Available in paper survey?
Enforce the desired number of responses to a question		

- A computer-aided-interview (CAI) can enforce a single response to a question
- On paper, respondents can and will give more than one response
- For the same question, you have data that will come back in different formats in the two modes. Inconsistent!

Enforcing the number of responses

- How do you approach a survey question when you can constrain responses in one mode but not another?
- For example, you could:
  1. Request and capture multiple response in both modes
  2. Enforce single response in CAI, and on paper:
     a. instruct single response, but keep multiple responses
     b. provide processing rule to keep only a single response
     c. provide processing rule to set multiple responses to missing

Enforcing the number of responses: 1. Non-enforcement

1. Request and capture multiple response in both modes

Question 1: What factors are important to you when selecting a health care plan?

- ☑ Premium (HPF1)
- ☑ Deductible (HPF2)
- ❑ Doctors in plan (HPF3)

CAI or Paper Dataset

ID	HPF1	HPF2	HPF3
10001	1	1	0

Enforcing the number of responses: 1. Non-enforcement

1. Request and capture multiple response in both modes
   - Considerations for this approach:
     - Data is collected in a consistent format in the two modes
     - You capture all responses to learn which factors are important, but do not learn which single factor is most important

Enforcing the number of responses: 2a. Encourage single response

2a. Enforce single response in **CAI**, and on paper: instruct single response, but keep multiple responses

Question 1: What factor is most important to you when selecting a health care plan?

- ☐ Premium (HPF=1)
- ☑ Deductible (HPF=2)
- ☐ Doctors in plan (HPF=3)

CAI Dataset

ID	HPF
10001	2

Enforcing the number of responses: 2a. Encourage single response

2a. Enforce single response in CAI, and on **paper**: instruct single response, but keep multiple responses

Question 1: What factor is most important to you when selecting a health care plan?

- ☐ Premium (HPF1)
- ☑ Deductible (HPF2)
- ☑ Doctors in plan (HPF3)

CAI Dataset

ID	HPF
10001	2

Paper Dataset

ID	HPF1	HPF2	HPF3
10002	0	1	1

Enforcing the number of responses: 2a. Encourage single response

2a. Enforce single response in CAI, and on paper: instruct single response, but keep multiple responses

- Considerations for this approach:
  - More likely to gain information about which single choice is most important consideration
  - Capture multiple responses from paper respondents if they give them
  - Format of the variables from web and paper will be different

Enforcing the number of responses: 2b. Keep single response

2b. Enforce single response in CAI, and on paper: provide processing rule to keep only a single response

Question 1: What factor is most important to you when selecting a health care plan?

- ☑ Premium (HPF=1)
- ☑ Deductible (HPF=2)
- ❑ Doctors in plan (HPF=3)

CAI or Paper Dataset

ID	HPF
10001	?

Enforcing the number of responses: 2b. Keep single response

2b. Enforce single response in CAI, and on paper: provide processing rule to keep only a single response

Question 2: What is your highest level of education?

- ❑ High school graduate or less (EDU=1)
- ☑ College graduate (EDU=2)
- ☑ Post-graduate degree (EDU=3)

CAI or Paper Dataset

ID	EDU
10001	3

Enforcing the number of responses: 2b. Keep single response

2b. Enforce single response in CAI, and on paper: provide processing rule to keep only a single response

- Considerations for this approach
  - Results in a single response per respondent
  - What biases are introduced by the processing rules to select only a single response from paper?

Enforcing the number of responses: 2c. Set multiple responses to missing

2c. Enforce single response in CAI, and on paper: provide processing rule to set multiple responses to missing

Question 1: What factor is most important to you when selecting a health care plan?

- ☑ Premium (HPF=1)
- ☑ Deductible (HPF=2)
- ❑ Doctors in plan (HPF=3)

CAI or Paper Dataset

ID	HPF
10001	NULL

Enforcing the number of responses: 2c. Set multiple responses to missing

2c. Enforce single response in CAI, and on paper: provide processing rule to set multiple responses to missing

- Considerations for this approach:
  - Results in a single variable per person
  - Not introducing bias by selecting response not directly given
  - Setting the information to missing that the respondent did provide with their multiple responses can also bias

Agenda

Data design

Data editing for question level inconsistencies

Data editing for survey level inconsistencies

Dataset harmonization

Data editing for survey level inconsistencies: Question visibility

Ability	Available in CAI survey?	Available in paper survey?
Display only the questions that the respondent should answer		

- A respondent completing a CAI survey, is only presented the questions they are supposed to answer.
- A respondent on paper can see all questions laid out, and may provide an answer to a question that they should have skipped over.
- For the same question, you have data that may need to be treated differently in the two modes. Inconsistent!

Data editing for survey level inconsistencies: Question visibility

Question 1: Do you have health insurance?

- ❑ No (HI=1)
- ❑ Yes (HI=2)

Question 2: What factor is most important to you when selecting a health care plan?

- ❑ Premium (HPF=1)
- ☑ Deductible (HPF=2)
- ❑ Doctors in plan (HPF=3)

Question 3: (etc)

Keep data as-is

ID	HI	HPF
10001	NULL	2

Clear any subsequent

ID	HI	HPF
10001	NULL	NULL

Upcode gate question

ID	HI	HPF
10001	2	2

Data editing for survey level inconsistencies: Inaccurate responses

Ability	Available in CAI survey?	Available in paper survey?
Apply consistency checks during the survey		

- How do you approach a survey question when you can enforce accuracy in one mode but not another?

Data editing for survey level inconsistencies: Inaccurate responses

Question 1: How many people live in your household?

☐ 2 people (HHS)

Question 2: How many of those people are in each of the following age groups?

☐ 1 Age 5 or younger (HHSLT6)

☐ ☐ Age 6 to 17 (HHSLT18)

☐ 2 Age 18 or older (HHSA)

Keep data as-is

ID	HHS	HHS LT6	HHS LT18	HHSA
10001	2	1	NULL	2

Update the first value

ID	HHS	HHS LT6	HHS LT18	HHSA
10001	3	1	NULL	2

Set to missing

ID	HHS	HHS LT6	HHS LT18	HHSA
10001	NULL	NULL	NULL	NULL

Agenda

Data design

Data editing for question level inconsistencies

Data editing for survey level inconsistencies

Dataset harmonization

Dataset harmonization

CAI Dataset

ID	HPF
10001	3

Paper Dataset

ID	HPF1	HPF2	HPF3
10002	0	1	1

Combined Dataset

ID	Mode	HPF	HPF1	HPF2	HPF3	HPF_P
10001	CAI	3				3
10002	Paper		0	1	1	2

Summary

A focus from the start on how the data will be used drives decisions throughout the survey process, including:

- **Instrument development phase**: How will you capture or resolve multiple responses to a question on a paper survey? Does that affect how you wish to collect data for that question on a CAI survey?
- **Data editing phase**: How will you approach inconsistencies due to skip patterns or logic checks that are possible in programmed surveys but not paper?
- **Data delivery phase:** What format will support your analysis, including variable structure in each mode, and using single or combined datasets?

Thank You

Contact us:

Hanna Popick, HannaPopick@Westat.com

Eric Jodts, EricJodts@Westat.com

With particular thanks to Jane Shepherd and Mike Brick for their development contributions.