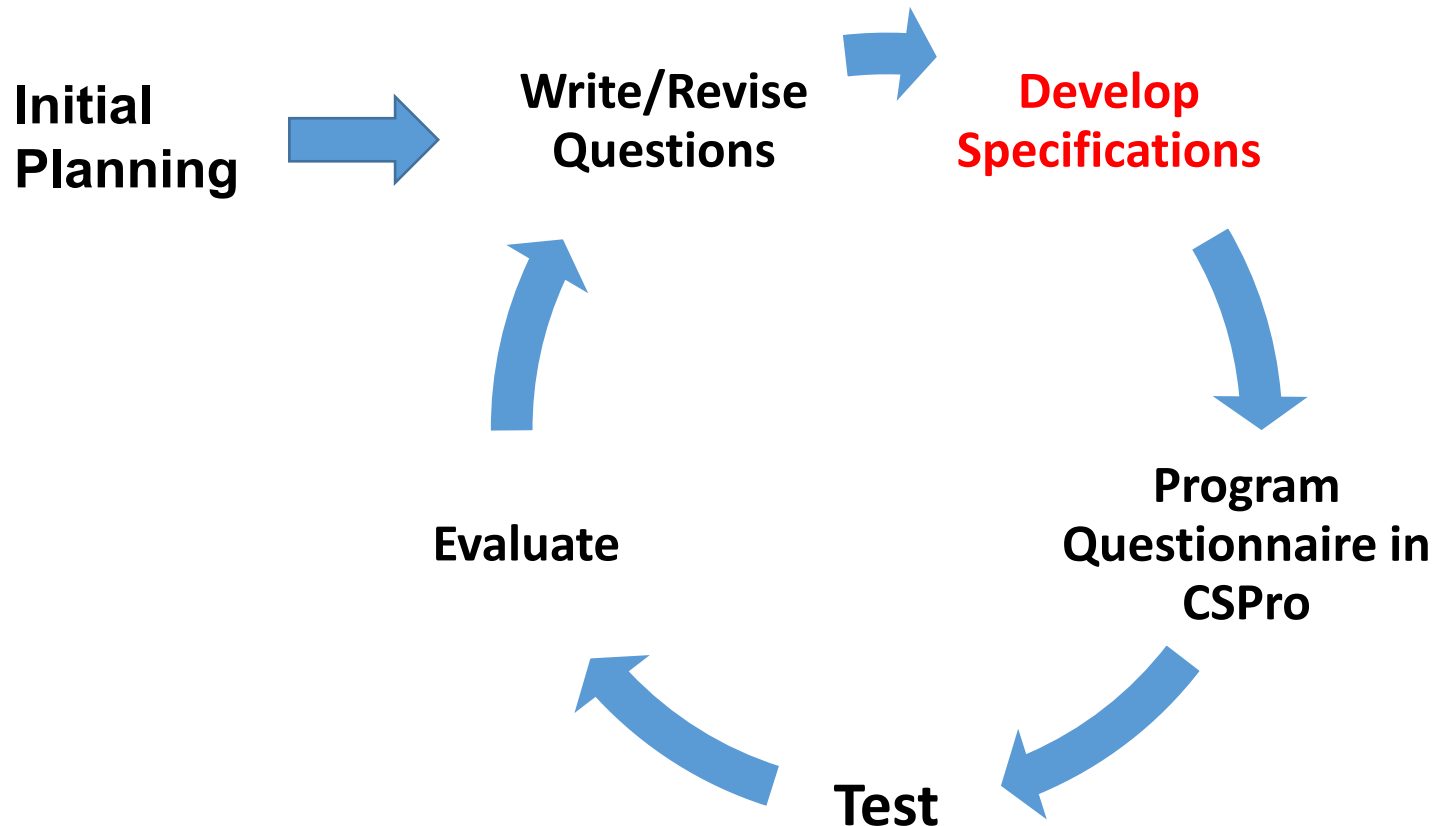# Management Considerations for Transitioning from PAPI to CAPI: CAPI Specifications

Dr. Fabián Romero
International Programs Center
Population Division
September 2022

United States® Census Bureau

USAID FROM THE AMERICAN PEOPLE

Steps in the development of the electronic questionnaire: involve subject matter experts

Initial Planning → Write/Revise Questions → **Develop Specifications** → Program Questionnaire in CSPro → Test → Evaluate →

What are specifications?

- A set of instructions for writing the program for a CAPI instrument.

- Contains everything programmers need to write the questionnaire code
  - Questionnaire wording, with textual questions
  - Answer Options
  - Skip patterns
  - Consistency rules
  - Selection of people for questions (universe)

Why developing specifications?

- Give clear instructions to programmers who may have limited knowledge of the questions

- Avoid possible errors in programming

- Maintain consistency when multiple people are involved in writing and programming of questions

- Useful when testing the questionnaire

- Serve as questionnaire documentation

Consistency rules and field consistency checks

United States® Census Bureau

USAID
FROM THE AMERICAN PEOPLE

Background

- Paper censuses and surveys require that enumerators are trained to:
  - Determine whether a question or section must be answered for a given person or household
  - Fill in valid codes for each response
  - Compare responses with previous or subsequent responses to determine if the response is valid and logically consistent

Background

- Even the best trained enumerators make errors, especially in a project as large in scale as a census

- Most organizations edit data post-capture, either manually by hand and/or automatically with computer programs

- The editing process adds time and complexity to post-capture processing

Background

- An advantage of electronic data capture is that error checks can be programmed into the data capture application so that pollsters can detect and correct errors at the time of the interview, in the field

Why Add Consistency Checks?

- No census or survey data is perfect, but it is important that the data is "clean" enough so that data users do not question the accuracy of the data, or question the data gathering process itself

- For example, publishing data like this may result in suspicion about the census:
  - Name: Yuliia
  - Age: 20 years
  - Years [Yuiliia] has been living continuously in [Kyiv]: 25 years

Why Add Consistency Checks?

- Consistency checks help enumerators identify errors while in the field and allow enumerators to fix problems immediately. This produces:
  - More precise data
 - Savings in terms of time and money during the post-capture/processing stage

Types of Errors

- Coverage Errors
  - Coverage errors arise from omitting or duplicating people during enumeration
  - Coverage errors are difficult to detect by a CAPI system, though one approach is to "list" an EA prior to enumeration and then compare the information from the listing with the information collected during enumeration

- Content Errors
  - Content errors can result from:
    - Respondents giving faulty information
    - Enumerators asking a question incorrectly or misunderstanding the response
    - Enumerators selecting the wrong response option
  - The majority of checks in a CAPI system are to reduce content errors

Goals of Checking

- The ideal CAPI census programs contain enough checks to sufficiently check for grossly inconsistent data while not prohibiting the enumerator from entering data in a timely and unobtrusive manner

- The presence of CAPI census consistency checks does not eliminate the need for post-capture edits, but well-designed checks will reduce the amount of post-capture editing necessary

- Creating consistency checks for a CAPI census should involve:
  - Subject matter specialists
  - Field operation experts
  - Data processors

Types of Consistency Checks

HARD

SOFT
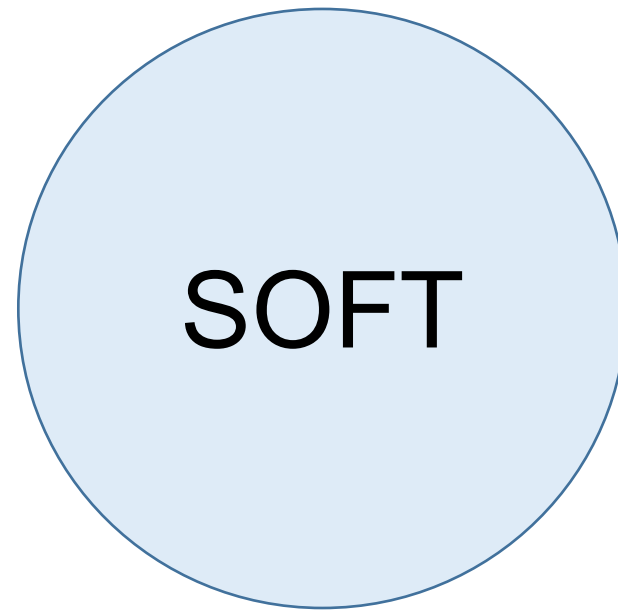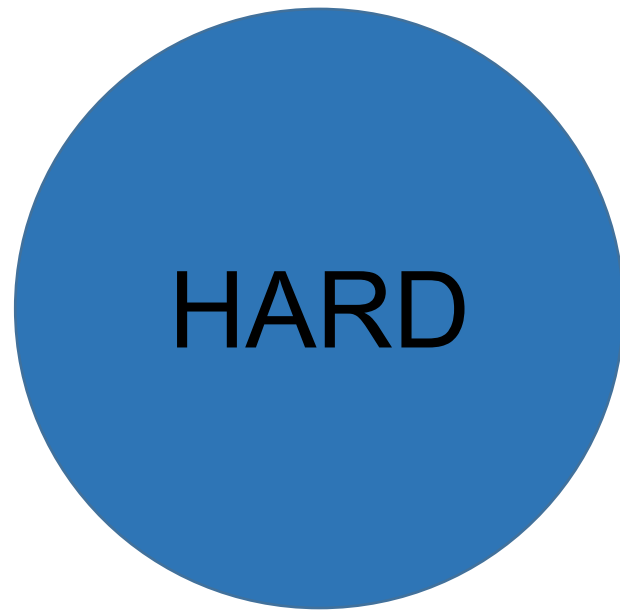
Types of Checks

- Hard Checks
  - Hard checks prohibit the enumerator from moving to the next question until a problem is resolved
  - Hard checks must only be employed when there is certainty about whether the problem is an absolute error
  - The worst case scenario with a hard check is that the check is programmed incorrectly and an enumerator is not able to proceed with a correct response; the enumerator may then input incorrect data in order to continue the interview

Types of Checks

- Soft Checks
  - Soft checks warn the enumerator that the data may not be correct
  - The enumerator must review the response and determine whether to proceed with the interview or whether to fix the response
  - Soft checks are ideal for circumstances in which the check is generally valid but when there are occasionally exceptions

Types of Checks

If in doubt, add a soft check, not a hard check!

Timing of Checks

- In general, you want to check the validity of the data as soon as possible so that the enumerator and the respondent can quickly correct the error.
    - Unitary checks
    - In-occurrence checks
    - Across-section checks

Timing of Checks (cont.)

Unitary Checks

- Unitary checks depend only on the response to a given question
- Unitary checks are generally simple range checks
  - For example: Is the response in the valid list of codes?
- Unitary checks can occur as soon as the response is given

Timing of Checks (cont.)

In-Occurrence Checks

- In-occurrence checks depend on other responses given in a section for one occurrence

  - For example: Is the person's age at first marriage less than or equal to the person's age?

- In-occurrence section checks can occur once all dependent data has been entered, or after all data for the occurrence has been entered

Timing of Checks (cont.)

Across-Section Checks

- Across-section checks depend on other responses given in a section across all occurrences
  - For example: Is one person specified as a head of household for the household?
- Across-section checks can occur once the dependent data for all occurrences of a section has been entered, or after all data for all occurrences has been entered

Unknown and Partially Known Responses

- It is also important to determine whether a "missing" or "unknown" response should be provided for questions

- Alternatively, some questions can have "partially known" response categories
  - For example, instead of an "Age Unknown" category, it may be useful to have categories such as:
    - Age unknown but a child not yet in school
    - Age unknown but a working-age adult
    - Age unknown but likely 50+

Unknown and Partially Known Responses (cont.)

- Partially known responses allow enumerators to collect some data rather than leaving a response blank

- Partially known responses can be assigned actual responses during post-collection editing and imputation

CSPro

Scope of Checks

- It is possible to add **too many** checks to a program

- It may be easier to fix some problems post-capture rather than during data collection
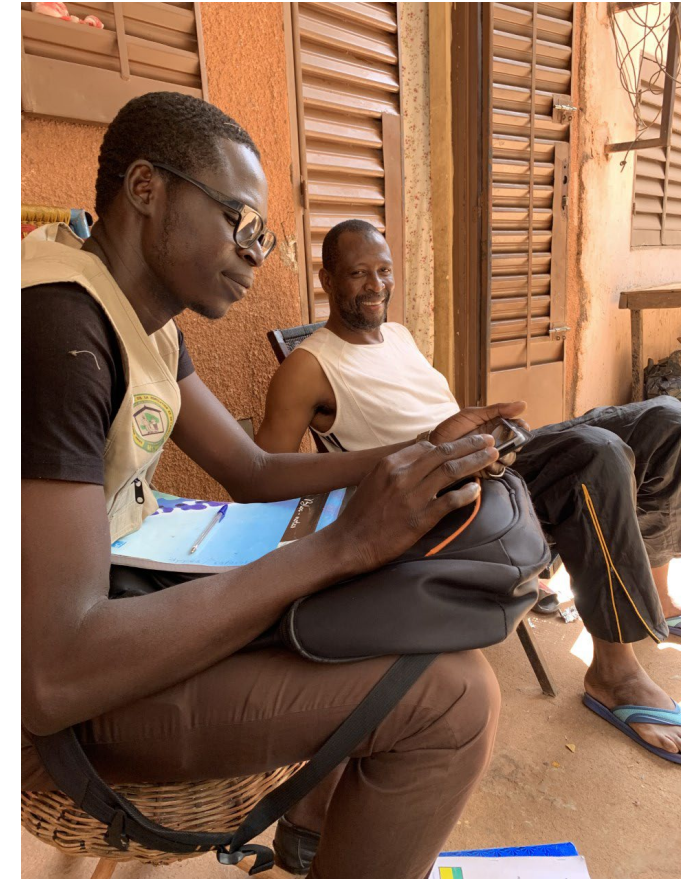
Mali, 2020

Scope of Checks

- The results of a questionnaire pretest and pilot census allow you to:
  - Analyze what checks were triggered the most
  - Check if there is any glaringly inconsistent data that should be checked for in future versions of the CAPI instrument
  - Ask the enumerators if the checks got in the way of a smooth data collection process
  - Check how much data passed through soft checks and determine whether the soft check should exist at all, or if the check should be changed to a hard check

Documenting Checks

It is important to have a document that describes all of the checks conducted by the CAPI census applications

- Programmers can use the document as they design the instrument
- Subject matter specialists can review the checks for accuracy and completeness
- Post-capture data users can understand the parameters that went into data collection

What consistency checks would you include for these questions?

How old is Peter?

- Range (0-100)
    - Is this a soft or hard consistency check?

Cannot be blank

- Is this a soft or hard consistency check?

If the month and year of birth were recorded, do a consistency check verifying that the estimated age is equal to the reported age.

- Is this a smooth or hard consistency check?

What consistency checks would you include for these questions? (cont.)

- What is the highest educational level Peter attended or attends?
  - Preschool
  - Primary education
  - Secondary education

- Should hard checks be added?

Answer	Age	Check
Preschool	3-5 >=6	Correct Soft check
Primary Education	6-16 <6 >16	Correct Soft check Soft check
Secondary education	12-20 <12 >20	Correct Soft check Soft check

Specifications Development

Specifications Template

- Number
- Variable
- Label
- Question Text
- Interviewer Instructions
- Fills
- Universe
- Responses

- Routing
- Checks
- Error messages
- Programmer instructions
- Help menu
- Change log

Specifications Template (cont.)

Number
- Question number that matches the paper questionnaire

Variable
- Specify the name of the variable for the CAPI data file (e.g. P01_NAME, P02_SEX, …)

Label
- Specify the phrase that will appear on the top of the tablet screen

Question Text
- The exact wording of the question to be read out by the census taker

Specifications Template (cont.)

Interviewer Instructions

- Any special instructions for the interviewer

Fills

- Specify how an item that will be preloaded will be written, e.g. [NAME]

Universe

- Specify for whom or what does this question apply. (e.g. Residents aged 3 years and older; What type of fuel is used for cooking in the housing unit?)

Response Format

- Format for the responses, e.g. radio button, check boxes, text boxes, etc.

Specifications Template (cont.)

Routing

- Specify what question should be asked next and for whom. Make sure you include specific routing for answers such as "Other: specify"

Checks

- For each question, specify any consistency or data completeness check that is required. For each response option, indicate:
  - if a check is required
  - soft vs. hard checks
  - error message if the check is violated
  - routing if check is violated

Specifications Template (cont.)

Error messages
- Specify what message should be issued for a given consistency or range check.

Programmer instructions
- Describe any special instructions to the programmers

Help menu
- If including a help menu in the program, the menu content will also need to be specified

Change log
- Summarize changes made. Include date and name of programmer making the changes.

Sample Specifications for Relationship to Head of Household Question

Specification	Sample
Number	A03
Variable Name	A03_RELATIONSHIP
Label	Relationship to head of household.
Question Text	What is [NAME]'s relationship to the head of the household?
Fills	A01_NAME
Data Type	Numeric
Capture Type	Radio Button.
Field Width	2
Interviewer Instructions	This field should not be blank.
Universe	All household members listed.
Responses	00- Head 01- Spouse/Partner 02- Son/Daughter 03- Son/Daughter-in-law 04-Step Child 05- Grandchild 06-Parent 07- Father/Mother-in-law 08 -Grand Parent 09- Brother/Sister 10- Nephew/Niece 11- Other relative 12- Not Related
Routing	If error message appears, (E1) please go back to A01_NAME and put the correct head of household. Otherwise go to A04_SEX.
Consistency Checks	1. Each household must have exactly 1 head of household. 2. Check if the first person listed is the head of household.
Error Message	E1: The first person listed MUST always be the head of household.
Programmer Notes	Soft check.
Help Menu	1. Head of household is a person who is at least 12 years old and is considered by other members to be head of household. In case head of household is less than 10 years old, should be allowed, in those rare cases. 2. It is possible to have one household head with more than one spouse.
Change Log	Misspelling on Question Text field corrected on 3/25/2018 by John Doe. Consistency check added on 5/28/2019 by Dr. Jane Doe.

Source: U.S. Census Bureau.

United States® Census Bureau

Summary

- Subject matter specialists have a key role to play in questionnaire design and CAPI application development

- Paper questionnaires often needs to be adjusted to be used in a CAPI environment

- Questionnaire specifications are key! Contain instructions for compiling the CAPI application

- Test, test and test again! Errors in jump patterns or consistency checks are harder to detect, as they may not be visible on the screen. Testing is critical in a CAPI environment.

Questions?

fabian.romero@census.gov