Regional Workshop on the 2020 World Programme on Population and Housing Censuses: International standards and contemporary technologies
Ankara, Turkey, 12-15 March 2019

Session 13

Multi-Mode Data Collection

Meryem Demirci

United Nations Statistics Division

Modes of census data collection

- Interviewer-administered data collection
  - PAPI *(paper q w face-to-face interview)*
  - CAPI *(computer-assisted personal interview)*
  - CATI *(computer-assisted telephone interview)*
- Self-administered data collection
  - PASI *(paper q w self-enumeration)*
  - CASI/CAWI *(computer-assisted self interviewing)*
- Multi-mode: use of more than one mode of data collection

Multi-mode data collection

- Benefits of using multi-mode approach:
  - to improve coverage, in a cost effective manner
    - Especially to reach people difficult to enumerate, such as people living alone, living in buildings difficult to access
    - Provide an alternative to people reluctant to participate

- Decision should be given through decision-making process to make sure it is cost- effective

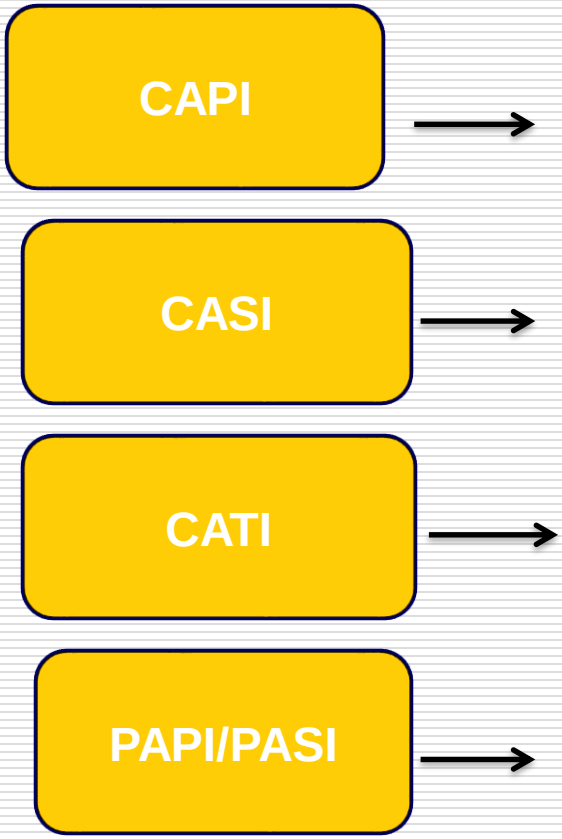
Multi-mode data collection

- **Factors affecting decision:**
  - **Population of interest:** Each method of enumeration provides better access to different population groups
    - The nature and geographical location of the population of interest may have a significant impact on selection of mode
  - **Operational and resource burdens:** Every additional method will create additional operational and resource burdens, including financial costs, human resource requires, data processing, etc.
  - **Time needed:** Planning and designing of each mode may require additional time
  - **Response rate:** Each mode should add significant value on response rate

Approaches for implementation of multi-mode data collection

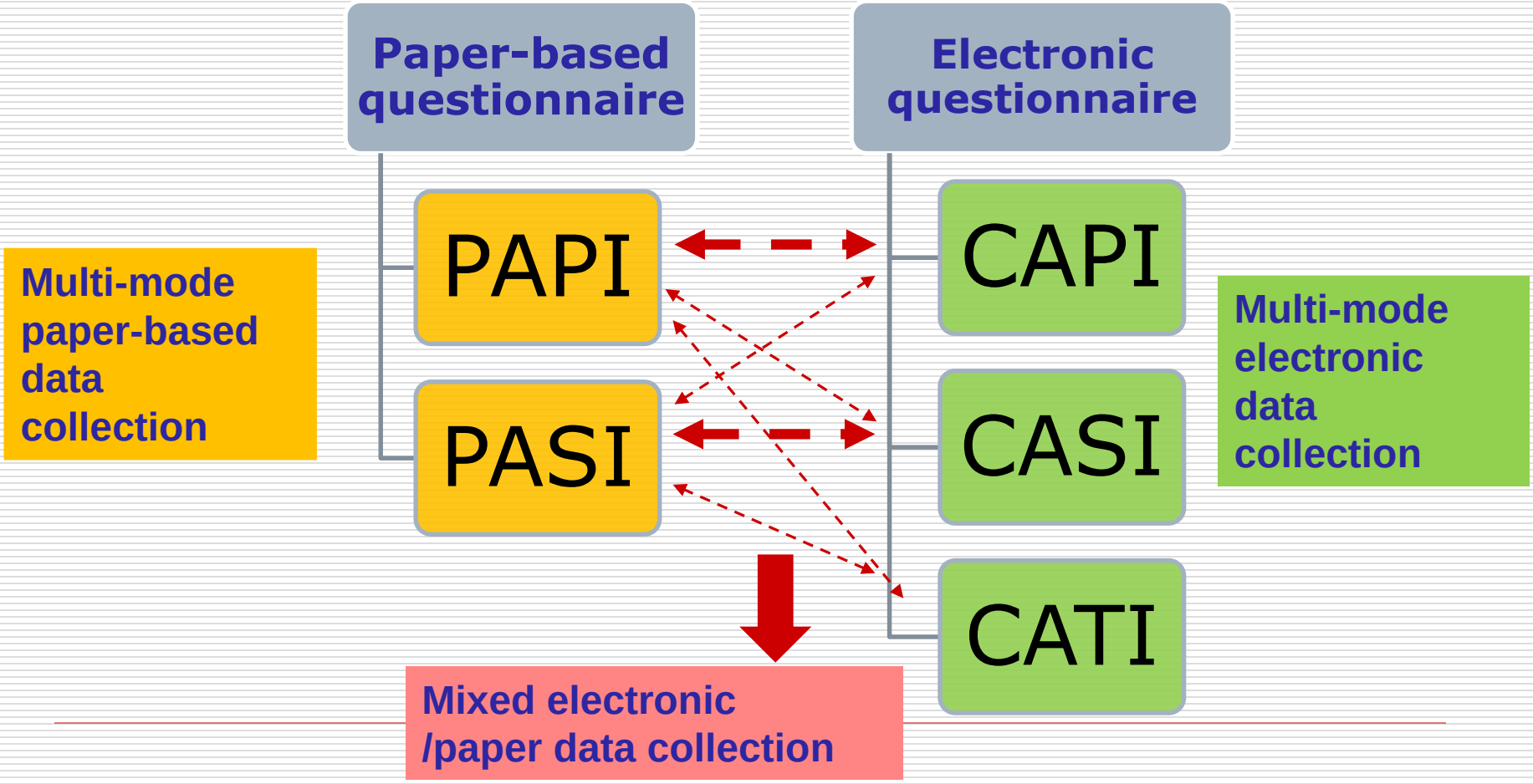
- **Multi-mode data collection can be designed using two different approaches:**
  - **Sequential approach:** All respondents are first requested to provide information in one particular mode only then are offered other modes to increase the response rate. One common practice is to start with the least expensive mode and then progress to more expensive and more persuasive modes
  - **Concurrent approach:** Respondents are offered, at the outset, the choice of one of two or more modes by which they can provide information. An example of a concurrent mixed-mode design is offering respondents the option of completing a paper questionnaire or providing information online via the Internet

How to decide multi-modes

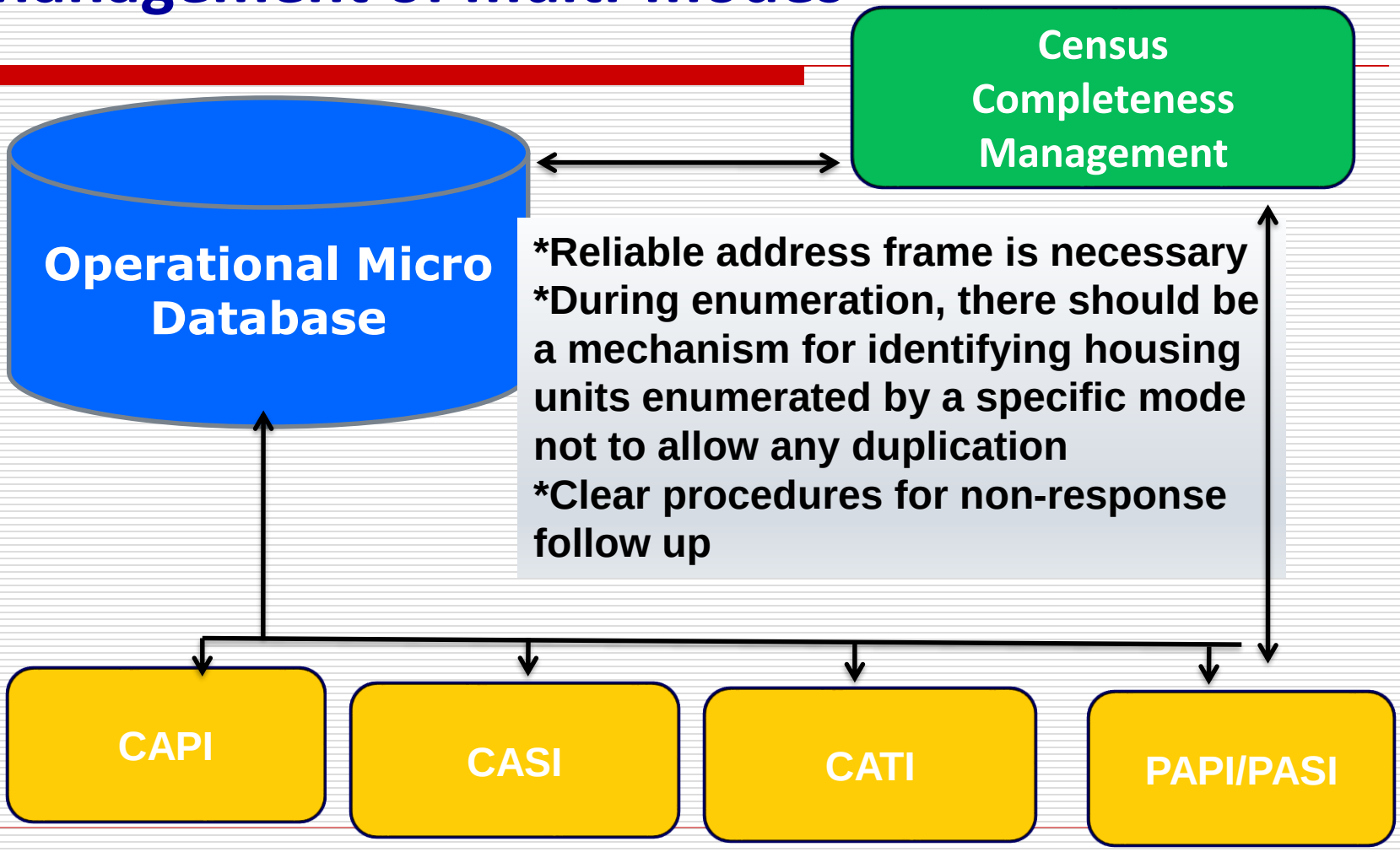


- What geographical areas/population groups are targeted with additional mode?
- Sequential or concurrent approach will be used
- What is the expected proportion of population who will be enumerated with specific method?
- What are additional burdens on operation?
- Is this approach cost-effective ?
- What extend response rate will increase
- Other ……???

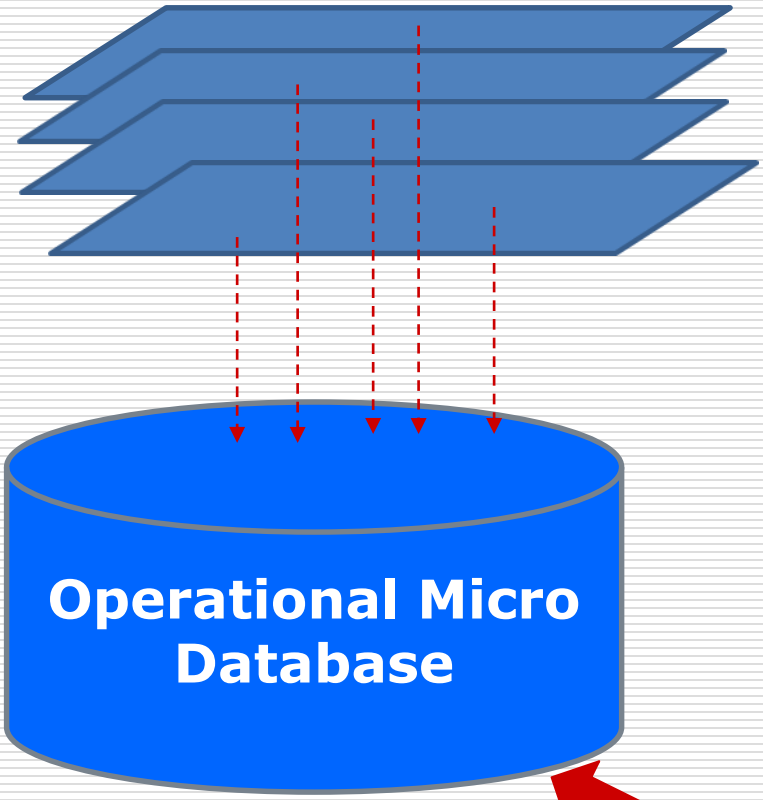
Possible options for multi-mode data collection



Management of multi-modes

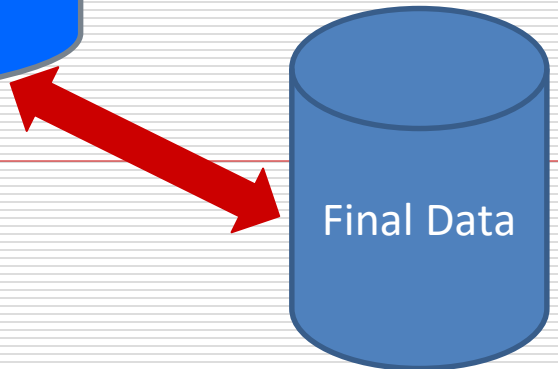


Final Record Generation



Data from different modes

- Integration with Census Frame
- Validation during the field enumeration (non-response, missing, inconsistency)
- If PAPI/PASI used, data capture, coding, editing
- Imputation
- Data validation/analysis



Mode effect in multi-mode data collection

- Multi-mode data collection has implications for the quality of the collected data, particularly for data comparability
- **'Mode effect'** means the bias caused by the mode of the data collection due to the delivery of different results as a consequence of using different means of collection
  - Mode effect creates artificial differences in the population
  - Mode effect varies depending on the type of multi mode system

Multi-mode data collection- Mode effect

- Four factors concerned with data quality and associated with mode-effect include:

  Coverage, Response rate, Item non-response and Measurement differences

a. **Coverage:** Coverage differs depending on the method of data collection
   - It is higher for face-to-face interview compared to self-enumeration method
   - A respondent who fills in the questionnaire may create a different type of error in listing members of the households due to misinterpretation of census questions or instructions

Multi-mode data collection- Mode effect

b. Response rate: Response rates may vary with mode of contact and mode of data collection

- Face-to-face interviews are more effective at securing high levels of participation, with a generally equal cooperation rate across different population groups, compared to the other modes (telephone interviews and self-response via postal mail or Internet)
- Modes of data collection are more or less likely to encourage different members of the population to participate

Multi-mode data collection- Mode effect

- **Item non-response:** Item non-response varies with mode of data collection
  - It is widely recognised that the use of electronic data collection will reduce significantly item non-response compared to paper-based data collection, due to the automatic control of non-response and consistency checks during questionnaire completion
  - Item non-response rate tends to be lower with the use of electronic data collection modes including self- response via Internet, face-to-face interview with hand-held devices and telephone interview, compared to paper-based data collection methods

Multi-mode data collection- Mode effect

- **Measurement differences:** One of the major challenges mixed-mode data collection presents is the fact that people tend to give different answers to questions (especially of a sensitive nature) depending on the mode in which they are asked
  - Measurement differences typically result from either the design of the questionnaire and the wording of particular questions being asked of the respondents, or – in the case of face-to-face and telephone interviewing – the competence of the interviewer

Multi-mode data collection- Mode effect

- **Minimizing mode effect on data quality**
  - **Optimize design**
    - *Sequential design- one mode is used as* main data collection method supported by additional data collection method for non-response follow up
      - Main data collection should be used to its maximum potential to enumerate the majority of population
    - *Concurrent design* - the different methods are equally important and respondents are given a choice
      - There is a risk of not using a mode to its fullest potential. For example, the minimal use of editing and validation in electronic questionnaires
      - ✓ *In general, sequential design has less pronounced mode-effect as compared to concurrent design*

Multi-mode data collection- Mode effect

- **Minimizing mode effect on data quality**
  - **Conduct empirical studies**- understanding causes of mode effects and assessing the magnitude of the mode effect
    - This kind of experimental study can be conducted with pre-tests and pilot censuses to understand mode effects on the data quality especially on item-non-response and measurement error
    - Findings from such experimental studies can be used to develop editing and imputation strategies in a way of decreasing mode effect on the data quality.
    - Where experiments are not possible, matching studies is another option to assess mode effects

Collection Response Rates - Canada

	2011 Census	2016 Census	
Collection Rate	**Actual**	**Planned**	**Actual**
Census Collection rate	98.1%	98%	**98.4%**
Internet	53.8%	65%	**68.3%**
Paper	31.3%	20%	**20.5%**
Self-Response	85.2%	85%	**88.8%**
NRFU	12.9%	13%	**9.7%**
Workload at start of NRFU	4.8 M	4.5M	**3.7 M**

Source : Statistics Canada Statistique Canada

THANK YOU…