



8th Session of OIC-StatCom

23-25 October 2019
SESRIC Headquarters
Ankara, Turkey

Electronic Data Collection Technologies in Population and Housing Censuses



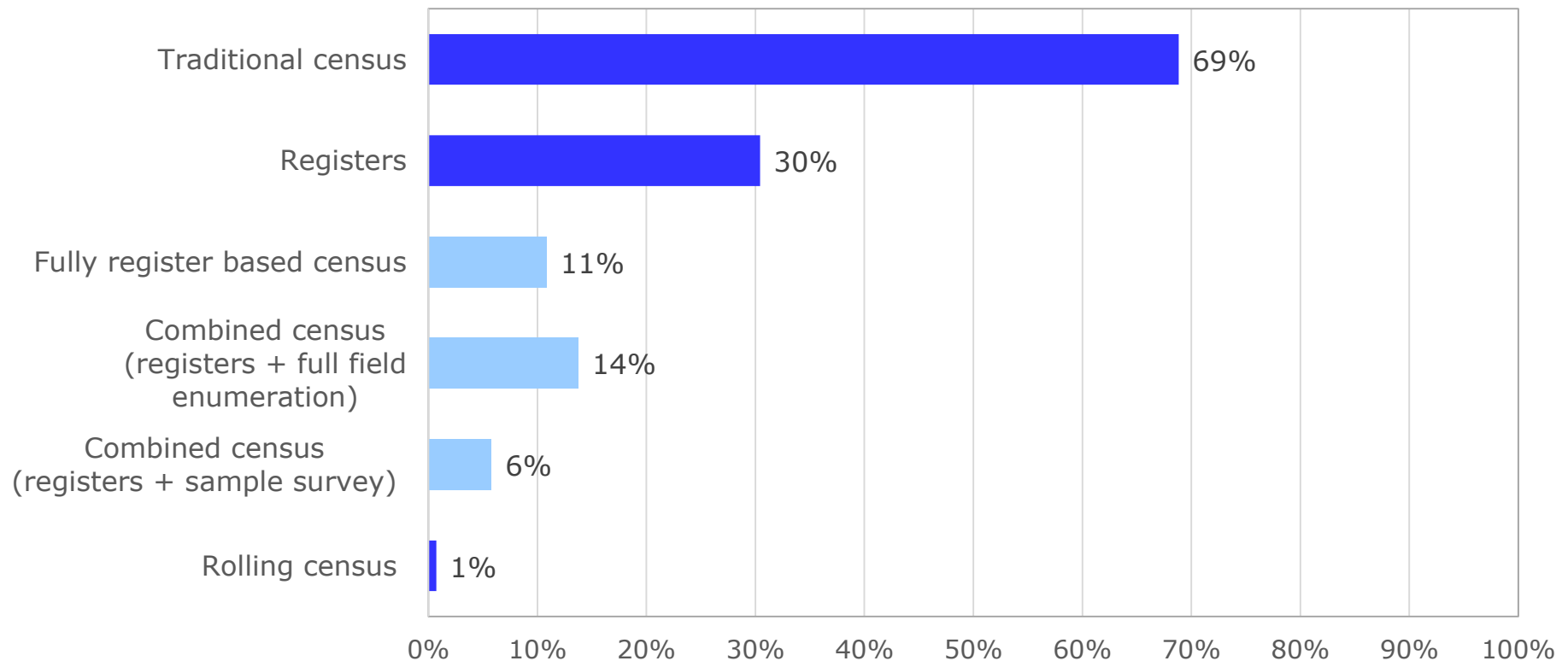
Outline

- ❑ Overview of plans for 2020 census round
- ❑ Planning considerations for designing census processes -
Handheld electronic devices
- ❑ Planning considerations for designing census processes –
Internet
- ❑ Managing multi-mode data collection operations
- ❑ Resources (UNSD methodological publications)



Overview of plans for 2020 census round

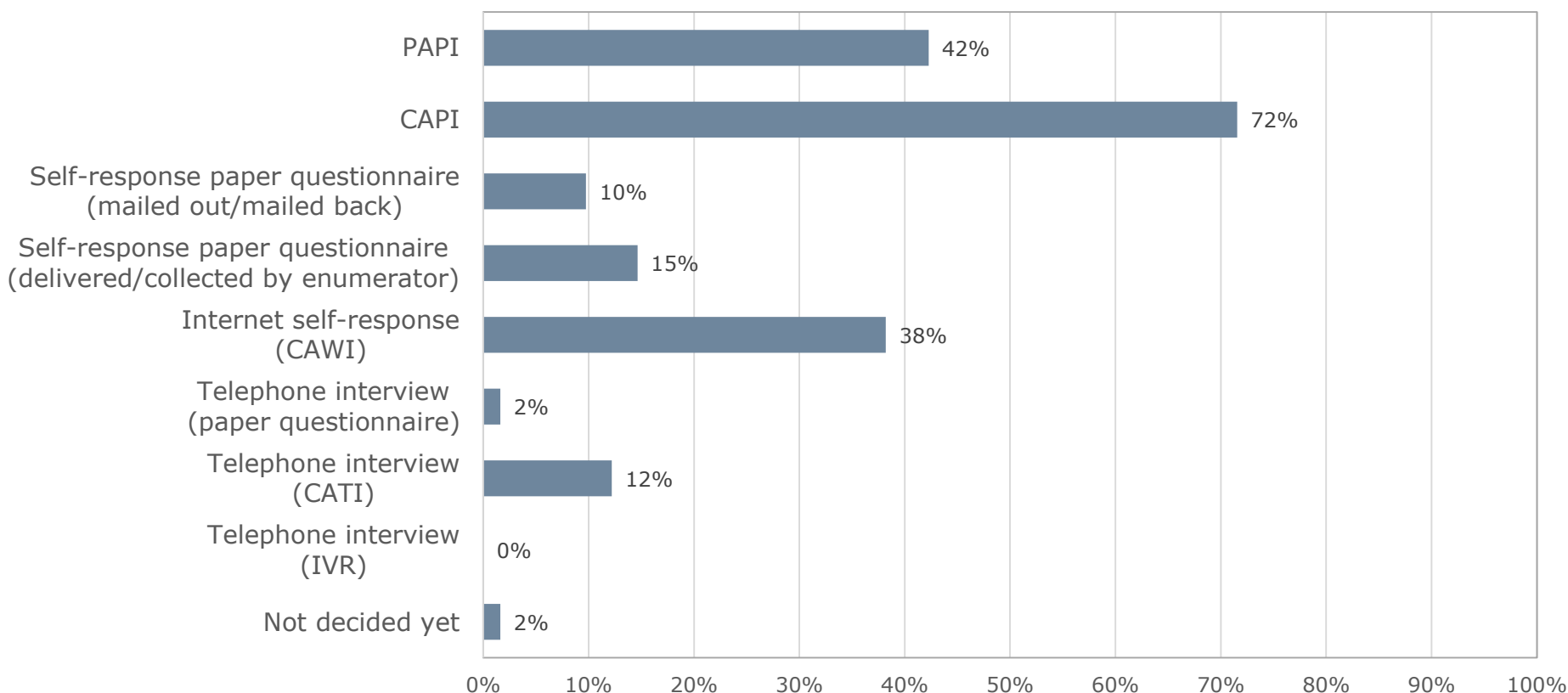
Main census method





Overview of plans for 2020 census round

Enumeration methods for field-based data collection



Source: UNSD survey on 2020 round PHCs

Response: 138 countries

Note: 123 countries conduct field-based data collection; 15 fully register-based census



Use of Handheld Electronic Devices in Population and Housing Censuses



Handheld Electronic Devices in PHC

Advantages:

- ☐ Reduced data entry errors -> Improved data quality
- ☐ Reduced time (interview, post-enumeration editing)
 - Data validation checks (range, logical/consistency checks)
 - Automated routing/skips
 - Automated coding
 - Preloading
 - Customization of questions
- ☐ Improved field operations management
 - Ability to collect paradata/operational info (GPS, date/time stamp)
- ☐ Easier handling (vs paper), positive perception



Handheld Electronic Devices in PHC

Challenges:

- ☐ High cost of equipment (high upfront investment)
- ☐ More time needed for preparation, testing
- ☐ Infrastructure constraints (connectivity, electricity)
- ☐ Sufficient technical expertise
- ☐ Assuring data confidentiality
- ☐ Physical security (devices)



Handheld Electronic Devices in PHC

❑ Critical factors in planning

- Census timetable
- Budget
- Technical skills and capacity development
- Infrastructure considerations (electricity, connectivity)
- Electronic questionnaire development (specification, testing)
- Systems and software design (testing: pre-field, field, pilot census)
- Integration with geospatial tools/digital maps
- Data transfer, Data security
- Monitoring during field enumeration
- Logistics (distribution/return)
- Use policy (authorized uses, proper handling)
- Re-use (benefits)/disposal of devices (damaged devices, data purging, environ. requirem.)



Handheld Electronic Devices in PHC

❑ Important features for selecting handheld devices

- Processor performance / Operating system / Storage capacity
- Keyboard / Screen / Battery
- Connectivity options
- Portability / Ruggedness
- Sensors and peripherals

❑ Evaluating requirements for selection of device

- Security criteria / Manageability criteria
- Productivity criteria / Performance criteria

❑ Acquiring the device

- Outright purchase / Renting / Bring your own device (BYOD)
- Sharing/pooling among agencies/countries



Handheld Electronic Devices in PHC

❑ Essential features of data collection application (CAPI)

- Interface for field users
- Intuitive navigation
- Automatic routing (skipping)
- Precoding
- Customising of questions
- Data quality control (validation/consistency checks)
- Case management
- Data management
- Technical support and documentation
- Support the collection of operational/paradata information



Handheld Electronic Devices in PHC

□ Electronic questionnaire design/development

- Questionnaire specifications should be written by subject matter specialists and programmers working together
- Data validation checks
 - Hard edits (age, sex, relationship to reference person)
 - Soft edits
- Questionnaire layout and design
 - Not just adapting/replicating a paper questionnaire (eg. one question per screen)
 - Exploit the features of CAPI/device, and take into account how interviewer interacts with device (eg. on-screen help, error message, language capability)
 - Where possible, preload with admin data/geocodes
- Testing (pre-field, field, pilot census)



Handheld Electronic Devices in PHC

□ Data transfer - Means of transmission

- Networking (Cellular 2G, 3G, 4G / WiFi / Bluetooth)
- Synchronization options (Online / Off-line)
- Servers (Web-based / FTP / Cloud-based service)

□ Data transfer - Technical considerations

- Estimating volume of data flow (for cellular)
- Managing mobile devices over cellular networks, taking into account:
 - Network performance constraints (network coverage/bandwidth/ and reliability, network imposed latency)
 - Application performance (how data collection application handles connection and response delays)
- Formulating strategies for mitigating risk in data transmission



Handheld Electronic Devices in PHC

❑ Security - Major security threats and vulnerabilities

- Data stored in the device
 - Physical, disclosure, application/web-based, poor authorization/authentication, use of location services, insecure storage, Bluetooth attack
- Data in transit
 - Insufficient transport layer protection, network exploits, wifi sniffing, phishing attacks, man-in-the-middle-attacks
- Aggregated data on the server
 - Insecure data storage, server side controls
- Other security-related challenges
 - Enumerators fabricating data, authentication not being possible where no cellular coverage, low budgets, not fully developed infrastructure



Handheld Electronic Devices in PHC

□ Security - Measures to secure data

- Authentication
- Encryption of stored data/data in transit
- Device use policy
- Use of MDM (mobile device management) technologies
- Development of security policies



Handheld Electronic Devices in PHC

□ Field workforce

- Review roles and responsibilities of field staff
 - Regional managers, Supervisors, Enumerators
 - Technology coordinator and IT support teams
 - Installation of applications on device
 - Checking all accessories
 - Assisting in case of hardware/software problems
 - Ensuring that devices are synchronized
 - Additional training of supervisors and enumerators, as needed
- Recruitment and training (content and technical)
- Technical support to staff in the field
 - IT support team, Call Centre



Handheld Electronic Devices in PHC

□ Field operation management and monitoring systems

(important to ensure complete enumeration)

- **Field operation management tools**
 - Supervisor's module (assignment, monitoring)
 - Enumerator's module (case management)
- **Management information system for operation control**
(tools for managers at HQ and in Regional offices)
 - Performance indicators (daily, cumulative, etc)
 - Supervision alerts
 - Dashboards, Geospatial tools



Use of Internet in Population and Housing Censuses



Internet in PHC

❑ Advantages

- Improved data quality (programmed checks)
- Improved timeliness
- Potential cost saving
- Improved coverage
- Greater convenience and lower response burden
- Protecting privacy (no interviewer involvement)
- Other benefits (environmentally friendly, generating a large amount of metadata)



Internet in PHC

❑ Challenges

- Providing reliable linkage between household and address/location
- Developing robust systems/infrastructure – portal may crash/respondent data may be corrupted or lost
- Protecting data security
- Extended collection period (hard to capture picture as at census day)
- Essentially a multi-mode collection
 - Challenges associated with multi-mode collection including mode effect and data comparability issues



Internet in PHC

❑ Requirements

- High literacy rate, and computer literacy
- High level of access to computers and the Internet
- Availability of address/building/dwelling list
- Legal authority
- Public trust and acceptance of Internet for official business
- Making reasonable assumptions about take-up rate
- Questionnaire design for mobile devices (multiple formats)
- Robust IT infrastructure (able to serve multiple concurrent users)
- System for tracking response status and for following-up on non-response



Internet in PHC - Planning considerations

❑ Contact and communication strategy

- Multi-phase contact approach (initial invitation, reminder)
- Use appropriate communication approaches (for different population groups)
- One of the strategies: offering paper questionnaire
 - at initial contact or;
 - during the reminder phase or;
 - on request



Internet in PHC - Planning considerations

❑ Identification and authentication of respondents

Benefits

- Properly link responses to households (address of dwellings)
- Reduced risk of duplicate responses
- Reduced risk of impersonation
- Better security



Internet in PHC - Planning considerations

□ Design/development of data collection application and portal

- Simple and intuitive Log-in screen (user experience starts here)
- Screen design - intuitive and easy to navigate
- Implementing:
 - Programmed logical/consistency checks
 - Validation messages
 - Automated skip patterns
- Well designed response options and menus
- Ability to progress through/save/submit questionnaire
- Mobile friendly, responsive design (multiple browsers, OS)
- Embedded support/help for respondents



Internet in PHC - Planning considerations

☐ Support for respondents

- Clear instructions
- Embedded help text in the online questionnaire
- Frequently asked questions
- Questionnaire guide
- Online help (including through chats)
- Census helpline (Call Centre)



Internet in PHC - Planning considerations

❑ Management and monitoring of Internet response (critical to ensure complete enumeration)

During enumeration, monitor:

- Daily returns, cumulative returns, compare with predicted number of returns
- Monitor number of saved/abandoned/incomplete returns
- Monitor number of concurrent users, IT infrastructure stability and capacity

For later analysis, monitor:

- Number of edits, item non-response rate, average completion time
- Use of help pages, use of foreign language options
- Different devices used to access (socio-demographic variables of HH)

Manage non-response follow-up:

- Central and integrated data collection operation control system and procedures are needed to manage non-response follow-up (to send reminder, paper questionnaire; initiate other mode, send enumerator, etc.)



Internet in PHC - Planning considerations

❑ Security

To maintain security of personal information:

- Secure log-in
- Ensure Internet application has zero footprint on respondent computer
- Timing-out after period of inactivity
- Encryption
- Powerful firewalls, intrusion detection
- Strong access control procedures
- Contingency plans for temporary service interruptions
- Communication strategies to assure respondents



Multi-mode data collection in Population and Housing Censuses



Multi-mode data collection in PHC

Objectives for using multi-mode data collection:

- ☐ Need to reduce high costs and administrative burden associated with field-based data collection
- ☐ Improve coverage
 - Reach respondents hard to reach – those living in gated communities/buildings with restricted access, young and mobile persons not found at home
 - Reach respondents reluctant to participate due to concerns about privacy by providing alternatives
- ☐ Reduce respondent burden by offering modes suited to different groups
- ☐ As more and more people become familiar with contemporary technologies, they demand modern solutions



Multi-mode data collection in PHC

Modes of 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*)
 - MOMB
 - DOPU
 - CASI/CAWI (*computer-assisted self interviewing*)
 - Web-based electronic questionnaire (*validation, skip*)
 - Download/upload form
- Multi-mode: use of more than one mode of collection



Multi-mode data collection in PHC

Modes differ in characteristics *(potential strengths and limitations)*

☐ **Ability to access different population groups**

- Each mode provides better/lesser access to different population groups depending on the nature and geographical location of the population of interest

☐ **Administrative and resource burdens**

- In terms of relative costs, each mode can be ranked, with face-to-face interview being the most expensive, and self-enumeration via mail or Internet offering more economical solutions

☐ **Time needed**

- Planning, designing and implementing of each mode may require more or less time (for testing, field work, etc)

☐ **Response rate**

- Each mode offers a different response rate



Multi-mode data collection in PHC

Design of multi-mode data collection

- ❑ **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 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



Multi-mode data collection in PHC

Challenges - Operational management

- ☐ Reliable census frame for linking households to addresses
- ☐ Mechanism for tracking contact/response status of households to avoid duplication
- ☐ Procedures for non-response follow up
- ☐ Allocating and assigning Enumerators
- ☐ Strategies for contact and communication (if self-enumeration modes offered)



Multi-mode data collection in PHC

Challenges – Data integration

- ❑ Prior to data collection (questionnaire design stage)
 - Ensure variable construction (in terms of variable name, type, length; and question wording, response options) is consistent across modes
- ❑ During data collection
 - Ensure instructions are clear for self-response (esp. paper) to minimize missing values, inconsistencies and outliers (as programmed checks and edits not applicable)
- ❑ Post data collection (duplicate cases can occur across modes)
 - Identify duplicate cases and reconcile data records; establish priority order for modes (eg. electronic collection mode subjected to programmed checks and edits given priority over self-response with paper)
 - Combine multiple input data files into a single output file (create master record)



Multi-mode data collection in PHC

Challenges - Mode effect

- ❑ **Mode-effect** is the bias caused by the mode of the data collection used – i.e., the use of one mode could potentially obtain result that is different from that of another mode → leading to data comparability (data quality) issues
 - Mode effect creates artificial differences in the population (i.e., differences observed are due to how the data are collected rather than real differences in the population)
 - Mode effect varies depending on the type of data collection design chosen



Multi-mode data collection in PHC

Four factors contribute to mode effect

- ☐ **Coverage** - differs depending on the mode selected (usually, higher for face-to-face interview compared to self-enumeration methods; for example, coverage errors made during listing of household members could easily be resolved thru follow-up questions during a face-to-face interview, while such possibility does not exist for self-respondent)
- ☐ **Response rate** - varies 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)
- ☐ **Item non-response** - varies with mode of data collection (electronic data collection will reduce significantly item non-response compared to paper-based data collection)
- ☐ **Measurement differences** - the tendency for people to give different answers to questions, especially of a sensitive nature, depending on the mode in which they are asked



Multi-mode data collection in PHC

Minimizing mode effect

❑ Optimize design

- In general, *sequential design* has less pronounced mode-effect as compared to concurrent design, especially, if the primary mode is used to its maximum potential to enumerate the majority of population
- In *concurrent design*, there is a risk of not using a mode to its fullest potential (for example, the potential benefit from use of validation checks in electronic questionnaires)

❑ **Conduct empirical studies** to understand causes of mode effects on coverage, response, item-non-response & measurement errors with a view to providing information to help optimize design of data collection and minimize mode-effect

- This kind of experimental study can be conducted with pre-tests and pilot censuses
- Findings from such studies can also be used to develop editing and imputation strategies in a way that decreases mode effect



UNSD Methodological Framework

Principles and Recommendations for
PHC, Rev 3 (2015)

Handbook on the Management of PHC,
Rev 2 (2016)

Technical
Report on
Post
Enumeration
Surveys
(2010)

Handbook
on
Measuring
International
Migration
through
censuses
(2017)

Guidelines on
the Use of
Registers and
Administrative
Data for PHC
(ECE, 2018)

Guidelines
on the Use
of Electronic
Data
Collection
Technologies
in PHC
(2019)

Handbook
on Editing
for PHC
Rev 2
(2019)

Handbook
on GIS in
Support of
Census
Activities
(2009)
Under
review

Handbook
on
Measuring
Economic
Activity
(2010)
Revised
under
E-learning
course