



**Economic and Social
Council**

Distr.
GENERAL

ECE/CES/GE.41/2009/12
19 August 2009

Original: ENGLISH

ECONOMIC COMMISSION FOR EUROPE

CONFERENCE OF EUROPEAN STATISTICIANS

Group of Experts on Population and Housing Censuses

Twelfth Meeting

Geneva, 28-30 October 2009

Item 4 of the provisional agenda

CENSUS QUALITY AND DISCLOSURE CONTROL

Quality Assessments of the 2008 integrated census – Israel

Note by the Central Bureau of Statistics, Israel

I. INTRODUCTION

1. Census operation and results were evaluated by several quality assessment dimensions. These included Coverage, Accuracy, Consistency, Completeness, Reliability and Validity^{1,2}. The United Nations Economic Commission for Europe publication "Conference of European Statisticians Recommendations for the 2010 Censuses of Population and Housing" lists some quality elements in the three stages of the census process; design, implementation and evaluation³. The elements relate to: accuracy, timeliness, accessibility, interpretability and

¹ Willem de Vries (2002), Dimensions of Statistical Quality, Inter-agency meeting on coordination of statistical activities, SA/2002/6/add.1.

² Leo, Pipnio, Yang W. Lee and Ricahrd Y., Wang (2002), Data Quality Assessment, Communication of the ACM April 2002/vol. 45, no, 4ve.

³ Conference of European Statisticians Recommendations For The 2010 Censuses od Population and Housing, United Nations Commission for Europe, 2006.

coherence. Israel conducted an integrated census in 2008 with no post enumeration phase since it was assumed that the integrated census with its two samples would provide the necessary information to evaluate the census operations. The evaluation process also took advantage of the fact that the census was an integrated one, which incorporates administrative sources and field operations. The census data collection operation was fully computerized with the use of laptops for data collection and daily data transmission to a central database of collected data. These features made it possible to monitor the field operation with specific tools which took advantage of the technology. To date, the full evaluation process has not yet been completed; it will be completed by the end of 2009. Nevertheless, most of the evaluation elements are completed.

II. INTEGRATED CENSUS METHODOLOGY AND PROCEDURES (IN A NUTSHELL)

2. The integrated census relies heavily on administrative sources, mainly on the Population Registry with the addition of other administrative sources to update the census for "inactive" persons and as an aid for constructing "administrative families". The quality of these data is subject to the needs, procedures, regulations and laws which the administrative data is collected and maintained. These do not necessarily comply with census definitions and needs. Two sample surveys were designed to adjust for the flaws of the administrative sources. The surveys' main purposes were:

- (a) Estimating the quality of the main administrative sources and adjusting accordingly population estimates of the "usual residents" for the country and their geographic distributions;
- (b) Collect socioeconomic-demographic characteristics of population and households for "statistical areas" (SA) and small localities. The Population Registry is the backbone of the integrated census⁴. Two independent sample frames were used, an area sample drawn from a geographic database created specifically for the census, which divided the country into 40,000 cells (which covered the total area of the country) with an average of 50 households in each cell (ranging from 30 to 70 households) to estimate under coverage of the Population Registry and a sample drawn from the Population Registry to estimate over coverage of the registry. The institutional population was enumerated fully, in a classic census procedure.

III. FIELD WORK OPERATIONS

3. A computerized questionnaire administered with laptops (CAPI) was used for the area sample. The Population Registry sample was interviewed by telephone (CATI). For the area sample survey, the country was divided into 12 regions. The field operation lasted 3 months, the first month being devoted to training, orientation and marking the sampled buildings and dwellings. The actual data collection phase lasted 7 weeks. The "census date" was 27 December

⁴ http://www.cbs.gov.il/www/mifkad/census2008_e.pdf

2008 and the data collection started on 28 December. Interviewers were recruited for a three-month period (from December 2008 to February 2009) and were paid differentially, according to quality of data collected and difficulty of data collection. The national minimum wage was paid as the lower limit.

IV. IMPROVED POPULATION REGISTRY

4. The Population Registry holds records of persons who were ever citizens or "permanent residents" in the country since the creation of the registry in November 1948, shortly after the establishment of the state of Israel. People who, definitely, no longer reside permanently receive a legal "inactive" status (mainly applies to people who have died or people who have officially notified the Ministry of Interior that they live in another country). People of a very old age (over 110 years) and without any updates in the registry for a long period are assigned an "inactive" flag which means they are most probably absent from the population. Records of people who are suspected of being absent from the country for a long, continuous period with no update in the registry for the time elapsed from their last open departure from the country in Population Registry are flagged as "possibly inactive" by removing their "residential address" from the Population Registry. The flags mentioned above are assigned administratively by the Ministry of Interior. All types of "inactive" residents are removed from the "voting lists" for parliament elections and municipal elections. Supporting administrative sources that were incorporated into the Improved Population registry (IPR) were:

- (a) National Insurance files (also responsible for national health insurance) which specify eligibility for health insurance and national social allowances such as: child allowance, disability/handicap allowances etc.;
- (b) List of electricity meters;
- (c) Vehicle registration.

5. The supporting files were used to update addresses and create "administrative families". The quality of the Improved Population Register and specifically the quality of the administrative families is still under assessment.

V. GEOGRAPHIC INFRASTRUCTURE QUALITY

6. The assessment of geographic coverage combined with updates was achieved with three pre-enumeration processes:

- (a) Improved orientation tools were developed for localities without registered addresses (mainly Arab villages without street names). It was carried out 12 months prior to the enumeration stage, and land marks were added to the computerized "building and street" layer;
- (b) Another assessment of building coverage was conducted a month prior to the enumeration process. Land marks were added manually to printed paper maps by

Geographic Coordinators, recruited specifically for update and orientation assistance mission. 3,694 land marks were added in 50 localities to the computerized layer and about 900 were added manually to paper maps, the latter are currently being incorporated into the digital geographic database;

- (c) The building layer needed updating for new buildings in the sampled cells. The sample was drawn from a digital geographic database, which was frozen 15 months prior to field enumeration. The computerized data related to the previous year database. New buildings which were not included in the database at the sampling stage were added at the pre-enumeration stage when enumerators listed all the dwellings in the sampled buildings and added buildings that were missing in the paper maps – 9,032 buildings (5 per cent) were added to the sampled cells (a total of 181,001 buildings in the original list). The number of new buildings was within the expected range of new buildings in the time frame (over two years) from aerial photographic images taken until the actual enumeration phase. This lag was a major concern and another task which the "Geographic Coordinators" executed was manually updating the paper maps produced for all tasks performed by the field personnel.

VI. THE QUALITY OF THE DIRECT DATA COLLECTION QUALITY (UNDER-COUNT AREA SAMPLE SURVEY)

7. The quality assessment of the integrated census is presented in the five dimensions of the quality assessment and by the various tools used in the process.

- (a) Enumeration coverage – an average of 5 per cent less dwellings were located in the field process than *a priori* expected number of households in the sampled areas, ranging from +6 to -21 per cent of dwelling, in different localities;

Enumeration Results

Households interviewed	83%
Full questionnaire	78%
Partial questionnaire	5%
Dwellings with no response (closed, not occupied or business)	9%
Refusals	4%

Each dwelling was visited at least 5 times before the final status was determined.

- (b) Re-enumeration by phone to assess reliability (short verification questions) by interviewer's superior – approximately 5 per cent of the sample was re-contacted to verify that the household was interviewed in person. Cases of misconducts were rarely detected;

- (c) Online consistency tests - The computerized questionnaires (BLAISE software) included integrated consistency tests. There were two types of consistency parameters; the first was a "warning" - which enabled continuing the interview after answering the warning remark, such as "Spouse same sex as interviewee, please verify" or "verify child's age" the second type was an "error" which prevented the continuation of the interview such as "immigration year earlier than birth year" check your entries. The amount of "errors" and "warning" documented in the computerized questionnaires has not yet been estimated;
- (d) Online assessment of validity and reliability – a computerized system was developed for "online auditing" of census data collection. The need for such a system emerged during the final pilot census (dress rehearsal in 2006) when it was suspected that some interviewers had identified ways of reducing their workload while getting fully paid. The daily computerized data transmission enabled almost online auditing of the transmitted data and interviewers performance, in additions to the incorporated consistency tests in the questionnaires. This procedure detected misconduct and enabled immediate feedback to superiors in the field and at headquarters. A set of pre-defined queries was run each morning on the previous night data transmissions. The queries were not tested prior to the census as the need and possibilities were defined after the final pilot test. New queries were added throughout the process.

Examples of the on Line query system

Type of query	Parameter	Verification type	Operational definition
Purposely reducing household members	Number of household members	Comparing household size to the average household size in the neighborhood	More than one SD from the mean HH size
	Frequent update in the household members	Frequent deletion of persons from household list	2%-5% suspicious >5% unlikely
Information not obtained from interviewee	Duplicated ID numbers	More than one record with the same ID number in all previous records for the same interviewer	5%-10% suspicious >10% unlikely
	Legitimate wrong ID number	ID not in the population register or belonging to an inactive person	<5% suspicious >5% unlikely
	Dates of birth with irregular distribution	Distribution of each component of date of birth	30%-40% suspicious >40% unlikely
Skipping questions	Frequency of "don't study" or	Number of persons with such responses	10%-20% suspicious >20% unlikely

Type of query	Parameter	Verification type	Operational definition
	"never studied"		
	Frequency of "don't know" or "refuse"	Number of questionnaires with more than 4 such responses	10%-20% suspicious >20% unlikely
Missing critical variables	Missing first name	Percent of persons with missing variable	1%-5% suspicious >5% unlikely
	Missing last name	Percent of persons with missing variable	1%-5% suspicious >5% unlikely
	Missing age/date of birth	Percent of persons with missing variable	1%-5% suspicious >5% unlikely

A report listing: name of interviewer, type of misconduct, level of each case of misconduct, number of questionnaires and number of records, was delivered, electronically, to the region chiefs and to headquarters each day. The system was not perfect but adequate and very flexible and, as mentioned, queries were added and amended along the way;

- (e) Matching data collected with the Improved IPR results, and validating the results – quality of obtained identification parameters. The results of the matching procedure are presented in Chart. Matching data collected in the field operation with the IPR is essential for the integrated census methodology. Matching is *a priori* possible only for Israeli residents, 98.3 per cent of the Israeli residents living in household were successfully matched and 95.8 per cent of the institutionalized Israeli residents were successfully matched. These results prove that the quality of the identification information obtained in the census was adequate.

VII. DATA COMPLETION SURVEY (OVER-COUNT SAMPLE SURVEY – TELEPHONE INTERVIEW) QUALITY ASSESSMENT ELEMENTS

8. The quality assessment in relation to the data completion survey was conducted using different procedures:

- (a) The first assessment procedure was, listening to a sample of interviews, voice recording, by headquarters personnel;
- (b) Incognito continuous follow up on a sample of telephone interviewers by professionals from headquarter for validity and reliability evaluation. The results of these quality assessments are not completed yet;
- (c) The third, was estimating the proportion of persons and households missed in the field operation. Number of persons found in their registered address in the telephone interview phase who were not enumerated in the field survey. The

results are an estimate of persons omitted from the field operation. 42,337 of the persons responding in the data completion survey, (telephone interview), reported their "census date" address as identical to the address under which they are registered in the Population Registry approximately 1 per cent of persons were definitely omitted in the field operation. These include persons whose household was interviewed but whose name was omitted from the questionnaire, or persons whose entire household was omitted.

VIII. AUTOMATED PROCEDURES

9. The integrated census uses automated procedures, only residuals are treated manually. Editing and imputation processes are fully automatic. The automated procedures were developed and implemented in order to achieve timeliness and consistency.

- (a) Answers assigned to "other" categories with text entered according to interviewee's verbal response to the questionnaire – a special automated procedure based on textual strings verifies that there was no specific category that could be assigned. 15 per cent of such "other" categories were assigned a specific category by the process;
- (b) Automated coding of economic industries and main occupation was used. 75 per cent of the records were coded automatically, of which 5 per cent were re-coded manually, to estimate the quality of the automated coding. The residual 25 per cent are coded manually in a computer assisted coding system (25 coders over a period of 8 months), 10 per cent of the manually coded records are randomly re-coded by different coders. Incompatible codes are re-coded by expert coders. The results of the quality assessment of the automated and manual coding are expected in November 2009;
- (c) Automated editing and imputation using NIM and CANCEIS are currently performed on socioeconomic and demographic information. A paper describing the rules and procedures will be presented at the UNECE meeting in December 2009.

IX. CONCLUSION

10. The results of the matching process (not yet fully completed as the manual matching efforts are still going on) for those with *a-priory* matching potential are 98.3 per cent for the household survey and 95.8 per cent for the institution survey. These results for the field operation are higher than the expected lower limit of 95 per cent match. The matching results for the "hard-to-contact" sub-population groups (based on the results of the previous census in 1995) are also in the range of 90 per cent matching. These results are assumed to be due to an extensive census promotion campaign specifically designed for each sub-population group. The integrated census which was based on two large samples (approximately 20 per cent of the population) and

full enumeration in institutions enabled better and more concentrated efforts in the sampled household and reduced the response burden on the population.

10. The final results due to be disseminated a year after census, is due to the extensive use of automated procedures at all stages of the census operations. The dual-system estimation enables the assessment of the census quality without the use of a "post enumeration survey".

Chart. Matching Results from the Under-Count Sample Survey

