

New Census Modalities in Israel

Pnina Zadka*

Central Bureau of Statistics, Jerusalem, Israel pnina@cbs.gov.il

Yael Feinstein

Central Bureau of Statistics, Jerusalem , Israel yaelf@cbs.gov.il

Abstract: The 2008 census in Israel was conducted as an integrated census, using registers, primarily the Israel Population Registry (IPR). Two sample surveys were conducted to adjust for weaknesses in the IPR, mainly undercount and over-count and absence of certain socio-economic variables. The IPR contains records for all permanent residents in the country and is updated continuously by vital events and migration (all domestic movements and immigration). The method used for adjustment was the Dual System Estimation (DSE) method. Two independent sampling frames were used to draw the samples for the surveys that were used to adjust for undercount and over-count in the IPR mainly un-updated addresses and emigrants. The undercount survey drew an area sample (20% of buildup area) based on a computerized Geographic Information System (GIS) covering the whole area of the country, buildings and streets layers. The undercount survey was also used to provide socio-economic information which is not available from the IPR. For the over-count survey the IPR was used as the sampling frame (10% of administrative households). An online computerized query system was incorporated in both surveys to minimize misconduct and other flaws in the data collection procedures. The response rate for the undercount survey was 83% (4% refusals). The response rate for the over count survey (telephone interview) was 78% (6% refusals). Population estimates were adjusted according to the census results. Since the 2008 census, adjustments coefficients are updated annually based on vital events and migration movements registered in the IPR. These updated coefficients correct population estimates each year.

Key words: dual system estimation, GIS based sample survey, integrated census, register based census

1. **Introduction** – The 2008 census in Israel, like censuses in the past, has a primary goal and secondary goals. The primary goal is producing reliable and updated population estimates for all geographic levels up to statistical area (census tract – an informal subdivision of localities, updated for each census, with an average of 3,000 usual residents) and for all localities regardless of size in the country.

There are several secondary goals:

- Producing socio-economic-demographic data at all geographic levels
- Serving as sampling frames for future specified surveys
- Possible linking to other individual information sources
- Socio-economic research

Up to the 2008, the censuses conducted in Israel were traditional censuses. They used paper questionnaires, a short form for 80% of the dwellings and a long form, for 20% of the dwellings.

The cost and the logistic organization of a traditional census on one hand and the availability and coverage of registers which can supply some of the data collected in a traditional census on the other hand, lead to the search for new modalities for the 2008 census. The censuses in Israel are not anchored in any legislation and are therefore not carried out on a fixed cycle. There were other circumstances which encouraged the transition to different modalities

- Declining response rates to surveys

- Fast rate of census data becoming obsolete due to fast changes in the population
- Outdated technologies, , emerging new technologies which cause the previous census technologies to become irrelevant.

As mentioned, Israel maintains a few registers that carry the potential to act as the base for a register based census. Nevertheless, these registers suffer from flaws that require an improvement of the data. This led the Israeli Central Bureau of Statistics (ICBS) to develop a method that will make use of the registers combined with specially designed processes which will complete and improve the register.

The most comprehensive register available in Israel is the Israel Population Registry (IPR). This register contains information on the permanent residents of Israel since the establishment of the state in May 1948. Each person in the register has a unique personal identity number (PIN) which is granted:

- At birth in the country or
- On the first entry to the country with a permanent residency visa or
- Upon changing residency status during a temporary stay in the country.

The data in the IPR contains all the information required in the past censuses (traditional censuses) short form. So it would have been reasonable to anticipate that there won't be any need for the 80% short form questionnaire.

This possibility was checked and found to be insufficient due to three reasons:

- The list of permanent residents in the IPR by itself does not conform with the definition of a "usual resident" as there are almost 12% of the persons in the IPR who do not reside usually in the country (i.e. they live abroad).
- The address in the IPR is not the current address of the resident or the address listed is not a full address (for about 20% of the addresses)
- The information in the IPR relates only to formal family relation and does not depict fully the actual household typology, and there is no information on housing or socio-economic characteristics in the IPR

2. Integrated census principles and methodology

The basic principle of the integrated census is to develop a system which would complete and correct the missing and inaccurate information of the IPR.

Overcoming the IPR drawbacks may be partially accomplished by using additional available registers.

As mentioned the IPR, like most registers, is subject to over registration and under registration - residents who are not registered but do live in the country and vice versa, people who are registered but actually live in another country according to census definitions.

A variety of administrative sources and two large sample surveys are used to complete and correct the IPR.

At the first stage the IPR is corrected using external administrative sources. This process produces the Improved IPR (IIPR)

- Border control files and National Insurance Institute (NII) files are used to delete persons living abroad
- Non residents file is used to add persons residing in the country – these residents are not incorporated into the IIPR and remain as a separate file – as they miss actual place of residency in the country and the surveys can not correct them.

The main flaw in the IPR is the un-updated addresses. These erroneous addresses amount to almost 20% of the usual residents, the proportion of erroneous addresses is quite stable over the years.

The integrated census methodology was designed to correct these inaccuracies in the IPR and adheres to the two main goals of the census:

- It updates and improves population estimates
- Socio-economic – demographic characteristics for all geographic resolutions

The first goal, updated population estimates is achieved through the dual system Estimation model.

Over-coverage (false captures) - the false inclusion in the classic model of capture/recapture - is represented in the census model by two independent sources; the IIPR and the other, data obtained from the IIPR sample.

The second goal, socio-economic–demographic characteristics - is achieved through the extended questionnaire used in the under-coverage survey (an area sample from the GIS layers). Using the two surveys and the record linkage procedures (Yitzkov 2003), population was divided into three groups:

People enumerated in their registered address

People enumerated in an address different from they are registered

People not enumerated in their registered address.

The relation between the three groups, form the parameters for the estimation formula (Glickman 2004)(Nirel 2004)

$N(i)$ Number of people living in area i (the number to be estimated)

$\hat{U}_{11}(i)$ Number of people that were found in the survey in area i and are registered in area i

$\hat{U}_{21}(i)$ Number of people registered in the ICPR sample in area i and not living there derived from the telephone sample

$\hat{U}_{12}(i)$ Number of people found in area i in the area survey and registered in another address in the ICPR.

$$\hat{P}_{1+}(i) = \frac{\hat{U}_{11}(i)}{\hat{U}_{11}(i) + \hat{U}_{21}(i)}$$

$$\hat{\lambda}(i) = \frac{\hat{U}_{21}(i)}{(\hat{U}_{11}(i) + \hat{U}_{21}(i)) / \hat{P}_{1+}(i)}$$

The census weight formula is therefore: $\hat{w}(i) = \frac{1}{P_{1+}(i) + \lambda(i)}$

3. 2008 census results and quality assessments

Census operations and results were evaluated by several quality assessments; coverage, accuracy, consistency, completeness, reliability and validity de Vries (2002)(Leo 2002) as well as quality assessments recommended by the Conference of European Statisticians (CES-UNECE 2006)

Geographic infrastructure quality - The assessment of geographic coverage combined with updates was achieved with three pre-enumeration specially designed processes

Direct data collection quality (under-count area sample survey) - 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.

Enumeration **coverage** – an average of 5% less dwellings were located in the field process than *a priory* expected number of households in the sampled areas, ranging from +6% to -21% 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%
----------	----

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

e. 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 and the second type was an "error" which prevented the continuation of the interview. The amount of "errors" and "warning" documented in the computerized questionnaires was negligible (less 0.5%)

f. Online assessment of validity and reliability – A computerized system was developed for "online auditing" of census data collection to detect misconduct by interviewers. This procedure 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 (cases of misconduct were re-enumerated)..

g. Matching data collected with the IIPR, and **validating** the results – quality of obtained identification parameters. Matching data collected in the field operation with the IIPR is essential for the integrated census methodology. Matching is *a priori* possible only for Israeli residents, 98.3% of the Israeli residents living in household were successfully matched and 95.8% of the institutionalized Israeli residents were successfully matched. These results prove that the quality of the identification information obtained in the census was adequate.

Data completion survey quality assessment elements (over-count sample survey,- telephone interview) had two validation procedures;

a. The first assessment procedure was listening to a sample of interviews, voice recording, by headquarters personnel, as well as incognito continuous online follow up on a sample of telephone interviewers by professionals from headquarter. for validity and reliability evaluation.

b. Statistical estimate of the proportion of persons and households missed in the field operation. 42,337 of the persons responding in the data completion survey, reported their "census date" address identical to the address under which they are registered in the IIPR, approximately 1% of persons were definitely omitted erroneously in the field operation from the under coverage survey (including persons whose household was interviewed but their name was omitted from the questionnaire, or persons whose entire household was omitted).

4. The Integrated Rolling Census in Israel -2020 round of censuses

The 2008 integrated census (IC), although reducing the cost substantively, still required extensive organizational, logistic and human resources for the implementation of the two large sample surveys.

The outcomes of the 2008 IC initiated the search for ways to modify the IC methodology and to overcome these deficiencies in the upcoming 2020 round of censuses. An organizational decision was made that the resources allocated for the census will be spread throughout a decade, while the total budget will remain the same. Any savings in the census operation will be invested in census supporting operations which will lead eventually to a full register-based census in the future.

4.1 Principles and requirements for the next census (2020 round)

a. Preserving and maintaining the methodology and technology developed for the 2008 census. The methodology of dual system estimation that was developed for the evaluation of total error and undercount in censuses (Hogan & Walter, 1988) was

extended to include the evaluation of over-count for the integrated census. The 2008 census results proved that the flaws in the IPR were still in a magnitude not tolerable in a full register-based census.

The technological tools and applications (for data collection, data integration and so on) developed for the 2008 census incorporated many innovations that proved to be functional for the achievement of most of the goals set ahead of time, in particular those regarding timeliness and accuracy.

b. Preserving experienced manpower and acquired expertise, while minimizing the fluctuations in the organization's manpower and the logistics implications.

c. To minimize the peaks in resources (manpower and budget)

d. Harmonizing sampling frames used for households and individuals in current surveys in the CBS In order to reduce response burden

e. Developing a geo-coded Building and Dwelling Register (BDR).

During the developing stages of the 2008 IC one of the registers that was lacking was the BDR. This register was found to be crucial for future censuses

f. Making use of all data available to increase robustness of census estimates-

Data obtained from administrative sources and from sample surveys are to complete and enrich census data.

g. Annual population estimates will be updated by census results- annual population estimates were based, in the past, on censuses as a baseline and on the demographic changes registered in the IPR (births, deaths, marriages and divorces - vital statistics events - as well as local and international migration). This process distorted the accuracy of the estimates, as flaws in the census data were kept until the next census corrected them, as well as flaws in the IPR were introduced into the estimates.. Another distortion was caused by the lack of updated institutional population.

h. Incorporating GPS technology in census operation in order to correct for flaws in addresses specifically missing full addresses.

i. Sample size -Statistical tests made on the 2008 census data proved that there is no significant increase in the variances of census estimates if a 10% sample is drawn for each survey (compared to approximately 20% in each survey in the 2008 census).

j. supporting registers - Three major registers will support the future census: BDR, Education Register, and Income Register. Another register that will be used is the Institutions Register.

4.3. Census of Institutions (collective dwellings) - A traditional census of institutionalized population will be carried out in two years cycles.

4.4. The future censuses - The type of census chosen to meet the terms set for the upcoming census was an **Integrated Rolling Census (IRC)**. The main principles of the 2008 census are to serve the IRC. Yet, the operation of the census is to be applied in a ten years period.

4.5. The sampling methodology in the IRC

The sampling is based on the scheme developed for the 2008 IC with only few modifications:

Under-coverage estimation survey- Hierarchical annual three-stage sampling process of municipalities, 10% of the Statistical Areas (SA) within municipalities, and 10% of the dwelling units or buildings in each sampled SA (wherever possible).

Over-coverage estimation survey – an improved and geo-coded IPR is used as the sampling frame. Sampling 10% of the administrative HHs registered in the sampled SA, amounts to 1% of the total number of HHs each year. Another 1% is sampled each year for the over-count survey from all non sampled SAs and they are used for the over-count estimates on a municipality level.

4.6. The Labor Force Survey (LFS) as a supporting survey- The LFS, a current monthly survey, conducted by the CBS will be used to support the census coverage and reliability. For this purpose census questions were added to the LFS questionnaire, these will be used for the estimation of the undercount on the municipality level.

4.7. Current stage

a. Currently, the CBS is engaged in the data processing and assessment of results phase of the 2011 IRC that was launched (as the first cycle) in February 2012, The reference day is December 31, 2011. The IPR and the BDR were used as the sampling frames for the surveys. The first round of the Institutions census was completed and data was incorporated in the 2011 population estimates. Annual population estimates for the SAs are to be adjusted based on the updated census results and will be calibrated to the municipality adjusted estimates.

5. Discussion

The rolling integrated census (IRC) is designed as the beginning of a process aiming toward a full register-based census. It will also serve as an ongoing evaluation survey, to facilitate quality control of the available administrative sources and registers to support a fully register based census.

The first steps in the implementation of the IRC seem to meet major goals set for the 2020 census round. It proves to be fully integrated in the organizational operations and mode of management. It reduces the peaks in organizational operations, and saves budget and time used for training inexperienced personnel. The preliminary findings, generated by monitoring the data collection process, show improved quality of data collected, achieved through better and more extensive training of interviewers by experienced staff. The management plans and the data collection technology are implemented in a short period of time, and are running smoothly with no need for a pilot test.

Careful risks management will be required. Some public criticism has already been directed to the IRC: Municipalities and the research community that are used to the data produced in a traditional census or the IC would like the next census to provide the very same information. These issues should be solved partially by the foreseen advancement in use of statistical registers such as the occupation and industry register (on the individual level), as well as those that are already available

References:

- Glickman, H., R Nirel and D. Ben-Hur (2004), "False Capture in Capture-Recapture Experiments with Application to Census Adjustments" *Bulletin of International Statistical Institute* 54th, session. Vol LX 413-414
- Nirel, R. H. Glickman and D. Ben-Hur (2004), A Strategy for System of Coverage Sample for an Integrated Census " *Proceedings of Statistics Canada Symposium 2003: Challenges in Survey Taking for the Next Decade*
- Willem de Vries(2002), "Dimensions of Statistical Quality", *inter-agency meeting on coordination of statistical activities*, SA/2002/6/add.1,
- Leo, Pipnio, Yang W. Lee and Richard Y. Wang (2002), Data Quality Assessment, *Communication of the ACM* April 2002/vol. 45, no.4ve
- Conference of European Statisticians Recommendations For The 2010 Censuses of Population and Housing, United Nations Commission for Europe, 2006
- http://www.cbs.gov.il/www/mifkad/census2008_e.pdf