



Economic Commission for Europe

Conference of European Statisticians

Group of Experts on Population and Housing Censuses**Eighteenth Meeting**

Geneva, 28 - 30 September 2016

Item 4 of the provisional agenda

Alternative methods for counting of population, in particular hard-to-count population groups**Coverage challenges in a multi-frame integrated census****Note by the Central Bureau of Statistics, Israel¹***Summary*

Sampling area cells in the under-coverage survey and sampling individuals from an adjusted population register in the over-coverage survey, present coverage challenges in addition to those usually confronted in censuses.

On the global level, generating the total population estimate became almost external to the census. The core source is the NPR, adjusted to include Israeli citizens who belong to the census population, and to exclude emigrants whose usual residence is abroad. Legal foreigners (mainly foreign workers) are added to the total estimate.

On the local level, the coverage challenge is embedded in ascribing the right statistical area to each person belonging to the census population. However, the combination of geographic frames and adjusted population frames creates unique cases of potential under or over-coverage.

A perfect census entity's profile in an integrated census should include an individual record in the adjusted population register, of a person who belongs to the census population, who lives in the census area, in a recognized locality, in a residential statistical area. However, reality is more complex and many records lack one or more of the above attributes, and therefore constitutes a potential under or over-coverage.

In this paper, we would like to identify these gaps in the integrated census, and contemplate possible solutions to narrow them down in the next census.

¹ Prepared by Olivia Blum and Yael Feinstein.

I. Introduction

1. The 2020 population and housing census in Israel, as the previous census (2008), will be an integrated census, based on the National Population Register (NPR). The NPR cannot serve as the sole source of geo-demographic information, since not all NPR records are of people belonging to the census population, not all people belonging to the census population are registered, and those who are registered, do not always have the right census address in their records.

2. Therefore, coverage evaluation surveys are carried out; an over-coverage evaluation survey uses an improved NPR (IAF – Integrated Administrative File) as the sampling frame for individuals; and an under-coverage evaluation survey uses the census area as the sampling frame for area-cells to be sampled from. These area cells are classified according to identified uses and only those area cells classified as residential or suspected to be residential, are sampled. The census estimates are based on the administrative records calibrated by the results of these surveys, using the expanded dual system estimates model (Glickman et al, 2003).

3. As a result, there are four, not necessarily nested, frames: NPR, IAF, census area, residential area-cells. This multi-frame basis of the integrated census presents coverage challenges of special populations to be addressed. Some can be answered by designated work processes and others are to be addressed by assigned statistical methodology.

II. Population coverage in the administrative registers

4. Ideally, building the IAF by improving the NPR would have meant omitting those persons who do not belong to the census population, adding those persons who belong, and assigning the right statistical area to each person. However, the available administrative files do not support all the required processes.

5. The NPR was created upon the establishment of the state of Israel, in its first census of 1948. During the last 68 years, the register accumulated records of people who migrated abroad, usually without notifying the authorities about their migration status. Some of them have already died. ICBS developed an Emigrants' Model in order to classify NPR records as representing people who do not belong to the census population. The model combines data from three sources; *The NPR* contains records of all people with permanent or temporary residential status, with their basic demographic information; *The Borders Control* adds history of entrances and departures to and from the country. The model is based on patterns of entry and exit that define place of usual residence either in Israel or abroad. The third data source is *the National Insurance Institute* that identifies people whose place of usual residence is not in Israel for at least five years, and people who declared, for taxation purposes, that their "center of life" is not in Israel. The National Insurance Institute also provides information on a small, hard to identify, population group; senior citizens who do not claim their old age pension. This usually implies that these people either are not alive – probably died abroad many years ago, or do not live in Israel and have no motivation to report it (See also Sabah, 2012).

6. The formal definition of an emigrant describes a person who left the country for a year or more. The operational definition allows for visits of up to 90 days during the last year. In practice, more than 80 percent of the emigrants, as defined by the model, are individuals who have emigrated for five years or more. The model is evaluated regularly using field surveys performed by ICBS. The evaluation operation validates the identification of "certain emigrants". Yet, there are still cases of indecisive travel patterns. Since Israelis can hold more than one passport, at times, they use their foreign passport

when crossing borders, and therefore are not recorded as Israelis, thereby adding noise to the model's input.

7. This model has been used continuously since 2008, and is being improved and refined. In recent years there were no significant changes in the model, and therefore the model outcomes are stable and comparable:

Table 1

Emigrants in the National Population Register 2011-2015

End of year	# of Active Records in the NPR	# of Emigrants (Model)	% of Emigrants (Model)
2011	8,461,241	575,623	6.80
2012	8,615,131	601,531	6.98
2013	8,770,914	615,545	7.02
2014	8,939,414	625,901	7.00
2015	9,111,987	641,641	7.04

8. Better identification of emigrants improves the quality of the IAF, by reducing its over-coverage. The IAF as a sampling frame carries a smaller over-coverage error and therefore the estimates generated by the designated survey are of better quality.

9. Another shortcoming of the NPR is that it carries records of Israelis only. Legal and illegal foreigners who belong to the census population (over a year in the country), are not included. Relevant information about legal foreigners (as recorded by the Population and Immigration Authority), mainly foreign workers, was not integrated into the IAF so far, and their estimation was generated in a separate process (see Table 2). Foreigners were under-covered in the last census.

10. The construction of the IAF, based on the NPR minus emigrants plus foreigners, together with demographic analysis, is addressing the global coverage challenge. However, correcting the local coverage in the NPR is a huge challenge in itself since it undergoes quality deterioration over time (see also Blum, 1999; Blum, 2003; Glickman et al, 2004; Nirel et al, 2003). Addresses are not fully updated because of regulations not enforced and because of differential benefits, monetary and others, allocated geographically (see also Zadka and Feinstein, 2012). Using administrative sources to correct addresses in the IAF has been found to be not justifiable since public institutions tend to use the NPR addresses as the formal address, and although some of addresses are corrected, others are hampered. The implication is that there are errors both ways, i.e. local under-coverage and local over-coverage of the IAF. The coverage surveys in the census aim to produce estimates of these local faults.

Table 2
Foreign workers who entered Israel with working visa in 1995-2015 and whose exit was not recorded by the end of 2015, by year of entrance

Year of Entrance	# of entrances a year (in thousands)	Exit was not recorded by the end of 2015	
		# (in thousands)	%
1995	78.3	1.8	2.3
1996	90.8	2.5	2.8
1997	64.0	1.4	2.2
1998	64.2	2.3	3.6
1999	53.3	1.8	3.4
2000	49.0	1.9	3.9
2001	75.2	2.9	3.9
2002	32.4	1.2	3.7
2003	25.1	1.1	4.4
2004	32.0	1.6	5.0
2005	29.4	1.3	4.4
2006	32.7	1.4	4.3
2007	36.5	1.6	4.4
2008	30.3	1.6	5.3
2009	26.6	1.5	5.6
2010	32.0	2.4	7.5
2011	32.7	5.1	15.6
2012	29.3	4.5	15.4
2013	35.6	11.0	30.9
2014	38.1	19.3	50.7
2015	42.6	35.0	82.2
Total	930.1	103.6	

III. Coverage of census area and of area cells

11. The other two frames of the integrated census are the census area and the residential area cells. The process of forming the sampling frame for the census surveys, involves an elimination process of areas pre-classified as not relevant. It is a tricky and an error-prone process; The census area is usually well recognized. However, the classification of area cells (basic data-collection working units) within a statistical area (the estimation unit) as residential, may introduce a challenge. Moreover, except for the Bedouins outside localities in the South, residential area cells are identified as such only within the boundaries of localities. As for their classification, the census area is not canvassed and area cells are classified in the office according to a set of indicators, spatial and others. Area cells not classified as residential (open areas, industrial areas, commercial areas etc.) have a smaller probability to be sampled, if at all, and therefore misclassification is a potential source of under-coverage.

Table 3
Samples of area-cells by cell type, 2008 Census

Cell type	# of Cells	# of Sampled Cells	Sampling fraction
Residential	50,513	9,196	18%
Bedouins outside localities in the South	1,053	162	15%
Residential area in construction process	2,014	111	6%
Residential in Industrial area	58	20	34%
Industrial	1,566	0	0%
Green area	6,181	0	0%
Public Domain	6,119	0	0%
institutional	287	0	0%
Others outside localities	333	0	0%
Total	68,124	9,489	14%
Total Residential	53,638	9,489	18%
Total Not-residential	14,486	0	0%

12. Area cells not classified as residential should be sampled in a very small sampling fraction, in order to estimate the quality of the classification and to avoid under-coverage. Moreover, some of the area cells outside localities (other not Bedouins) should be part of the classification process, especially areas within a regional municipality, and should be sampled accordingly.

IV. Multi-frame potential under-coverage

13. The integrated census methodology aims to have estimates for Statistical Areas (SA) based on the outcomes of the coverage surveys, using the dual system estimate model (DSE). Estimates of over, under and net-coverage are provided for each SA, within localities, vis-à-vis the population of this area in the IAF. The derived enumeration process risks potential holes in the crisscross of population and area. People who tend to be under or over-covered fall in the following categories:

Table 4
Potential Under-coverage

In IAF/ Outside census area	In census area/ Outside localities	In census area/ Outside IAF addresses	In localities/ Outside residential area-cells	In census area/ Outside IAF
Diplomats, Living in unrecognized outposts, <i>Living in marine vessels</i>	Bedouins, Living in or out of regional municipality	Living in new residential area, <i>Arab localities with no addresses</i>	Living in industrial/ commercial areas	Foreigners

14. Individual record in the IAF, living outside the census area:

(a) Israeli Diplomats abroad are a potential under-coverage. If they are not identified as diplomats, those who are over a year abroad may be defined as emigrants, and those who are under a year abroad are a potential over-coverage in their IAF addresses;

(b) People who belong to the population and live outside the census area are a potential under-coverage. They are defined as over-coverage in their IAF addresses but are not defined as under-coverage in their place of usual residence (marine vessels, unrecognized outposts). Unlike the diplomats, they are counted in the global estimates, but not in their census addresses.

15. Individual record in the IAF, living in “addresses” outside localities:

(a) Bedouins in the South, who live outside localities and are identified as such in the IAF before the census, are not ascribed to the area they live in but rather to one aggregate of people living outside localities. Local coverage should be improved and defined according to geographic delineation in fact.

(b) People, who live outside the formal boundaries of a locality, may be missed in the local estimates, especially if their number does not cross a minimum threshold in a single location. In regional municipalities, under-coverage has local consequences since the characteristics of those living outside the localities tend to be different from those who are living in localities within the regional municipality.

16. Individual records in the IAF, living within the localities, in statistical area without addresses in the IAF:

(a) People living within the locality’s boundaries but in a statistical area with no addresses in the IAF, are not ascribed to the area they live in. This is usually the case in new residential areas. They are defined as over-coverage in their IAF addresses, but are not defined as under-coverage in the area they live in. This area remains empty since no records are found in the IAF to be inflated.

(b) In cases where the statistical area carries addresses having relatively small representation in the IAF, those who are registered in these addresses are assigned with (too) large weights. As a result, statistics for this area may be biased.

(c) In Arab localities with no addresses, the under-coverage may occur on a statistical area level only, since the records are assigned to the locality.

17. Individual records in the IAF, living in statistical areas classified as not residential – People who belong to the census population, who live in areas classified as not for habitation, may not be ascribed to the right statistical area since these areas are under-sampled if at all.

18. Individuals living in the census area, with no IAF record – Foreigners are a potential under-coverage of this sort

(a) The global estimation of legal foreigners in Israel is supported by the Population and Immigration Authority records and can be considered good enough. However, their regional distribution is not updated and if foreign workers live in areas not classified as residential, they may not be found in the census at all. Those who live in residential areas are in Israeli households, and those who live in foreigners households in Israelis neighborhoods, are included in the regular main-stream process. Those who live in foreigners areas were not addressed in previous census because of negligible size, but should be addressed in the next census;

(b) Illegal foreigners are estimated globally (entrances to Israel minus exits, by country). They are not defined as a target population in the census and some of them may

be randomly included in the next census. They are under-covered locally, and it is not clear if they are under or over-covered globally (See also Sheps, 2016).

V. Concluding remarks

19. In a traditional census, the main frames are "tangible": The census area, and the sampling frame for the socio-economic questionnaire, created in the field as a result of thorough canvassing of the area and of the dwelling units. These frames usually overlap. In an integrated census, more frames are created, some of them are "virtual, and they do not necessarily overlap. Area-wise, the challenge resides in the decision what area-cells should be included in the frame. Population-wise, the challenges are of two types: building as accurate as possible statistical population register (with foreigners, without Israeli emigrants abroad), and producing fit-for-use local estimates in all census area. It is not necessary to be present everywhere and to enumerate everyone in order to produce unbiased estimates with a reasonable variance. However, the gaps described above inflict the estimates of specific small areas and populations, resulting in insufficient quality and specificity (higher variance).

20. Some of the gaps produced by these frames were taken into account in the estimation processes. For example, although not all population outside localities was ascribed to its specific geographic location, all of it (except foreigners) was included in the census estimates. Foreigners were included in the ongoing demographic estimates, globally. Yet, more attention should be paid to processes and methodologies that can be further developed and supplemented to improve quality and resolution of the estimates. We have to ensure that all Statistical Areas are represented and that enough area cells are sampled from each, to reflect the variance.

21. Planning for the next census, methods, methodologies and processes are developed in order to overcome most difficulties encountered in the previous census. ICBS is planning to refer to the whole area rather than to area of localities, classification of area cells will be better checked, non-residential area cells will be sampled in a small fraction in order to validate their classification and to find out whether population reside there. Functional fieldwork processes will be developed for this end. Moreover, ICBS is changing the sampling design and moving to a multistage sampling in order to improve the sample efficiency.

22. As for the population, ICBS is improving the identification of those who do not belong to the census population, including those who live outside the census area, and those who do belong, including legal and illegal foreigners.

23. The foreigners' population and its geographic distribution can no longer be ignored. The information collected from the administrative files and from the field surveys is lacking. Legal foreign workers have their employer's address in their records but may live in a different residential address, and illegal migrant workers are often entirely unlisted. In field surveys they are hard to count and do not always cooperate with the interviewers. For 2020, the Population and Immigration Authority is making an effort to monitor the whereabouts of the legal foreign workers, ICBS plans to contact employers and find out if they can improve the information regarding local coverage of their workers, and to develop designated fieldwork processes to get cooperation and better coverage of this population. Illegal foreigners are estimated globally but we may be able to improve the local estimates by better covering the legal ones and analyzing similarities in living preferences, between the two populations.

24. ICBS is also improving the model for identifying Israelis who do not belong to the census population, mainly emigrants. One of the problems encountered after 2008 has been

the over-estimate of old people. Even though the dual system estimates model handles the global over-count caused by unidentified emigrants in the improved NPR, differential corrections are needed for different population groups (Jews and Arabs; age-groups) and the global over-count have to be allocated (by calibration) intelligently accordingly.

25. The integrated census is an ongoing challenge in itself and especially when using more and more administrative files in the combination of direct and indirect data collection. It is reasonable to assume that not all problems will be solved, using methodology or work processes; however, it is possible to minimize them by making sure that the frames as an infrastructure will be as complete, identified and integrative, as possible. A thorough evaluation of critical components of the integrated census, and its results, will enable us to continue improve means and ends.

VI. References

Blum O. (1999), Combining Register-Based and Traditional Census Processes as a Pre-Defined Strategy in Census Planning. Statistical Policy, Working Paper 30. FCSM, Washington DC.

Blum O. (2003), Wholistic Approach to Multifaceted Integrated Census. FCSM, Washington DC.

Glickman H., R. Nirel and D. Ben-Hur (2003) False Captures in Capture-Recapture Experiments with Application to Census Adjustment. Bulletin of the International Statistical Institute, 54th Session, Contributed Papers, Vol. LX (2003), pp. 413-414

Nirel R, H. Glickman and D. Ben Hur (2004), A Strategy for a System of Coverage Samples for an Integrated Census. Proceedings of Statistics Canada Symposium 2003: Challenges in Survey Taking for the Next Decade.

Sabah M. (2012), Israeli Emigrants Stock. Technical Paper Series. Israel Central Bureau of Statistics (in Hebrew).

Sheps M. (2016), Challenges in Estimating Irregular Migration in Israel since 1995. UNECE Work session on Migration Statistics. Geneva, Switzerland.

Zadka P, and Y, Feinstein (2012), The 2008 Integrated Census in Israel and Future Censuses. United Nations International Seminar on Population and Housing Censuses: Beyond the 2010 Round. Seoul, Republic of Korea.
