



Economic Commission for Europe

Conference of European Statisticians

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

Geneva, 26–28 September 2018

Item 2 of the provisional agenda

Methodology, new data sources including big data**From physical area to virtual lists: Toward an administrative census in Israel****Note by the Central Bureau of Statistics, Israel****Summary*

Characterizing a census usually refers to the sources of information and to the data collection mode; Direct data-collection from the population in a traditional census versus secondary use of data-collected for other purposes, in an administrative census. Growing needs, growing diversity and rapidly changing reality define the administrative census as the more suitable mode of action for generating relevant information, given technological, regulatory and other enablers.

The transition process toward an administrative census involves turning points that present challenges and requires decisions to be made in spite of the embedded uncertainty. Moreover, the actual crossing point to an administrative census is a conceptual and practical leap from physical area to virtual lists.

The 2020 census in Israel will integrate data from different sources, including the population. However, unlike the previous census; no area sampling will be involved and a major step toward an administrative census will be made. An administrative geo-demographic census will serve as a sampling frame for a socio-economic survey that is planned to be used for improving the small-area population estimates.

In this document, I would like to shed light on the risks and the opportunities that this critical decision brought about to the future censuses in Israel.

* Prepared by Olivia Blum.

I. Introduction

1. The 1995 Census of Population and Housing in Israel was planned to be the last traditional, door-to-door census. The aim was to follow the Nordic countries example of gradual progress toward an administrative census. It was implemented in 2008 Census, where a geo-demographic file was generated by improved Central Population Register (CPR) data, followed by evaluation and correction surveys that produced over and under-coverage estimates, along with socio-economic information.
2. The growing and pressing needs for timely, frequent and relevant detailed information brought the Israeli Central Bureau of Statistics (ICBS) to investigate the possibility to conduct a rolling census. Engaging research and actual field data collection were not culminated in satisfactory results.
3. At the same time, Israel, as many other countries, experienced an accelerated process of using administrative data for the production of official statistics at large and improved its abilities to use administrative data for census purposes, in particular. However, a full administrative census has not been found to be a viable option yet. Consequently, ICBS reached a critical crossroad regarding the preferred methodology for the next snapshot census.
4. Two key facts enable a shift in the planned methodology: a) Entries and departures to and from the country are well recorded; b) All people in the country have administrative records; the citizens are registered in the CPR and the foreigners are reported in functional records like work permits and visas. A conceptual and practical leap toward an administrative census can be done.
5. The 2020 census in Israel will integrate direct and indirect data collection, by engaging new data sources in different methodological processes. For the first time, an administrative geo-demographic file will serve as a sampling frame for the field survey, aiming toward improving the small-area estimates, and supplementing socio-economic information, totally or partially missing from the administrative records.

II. The geo-demographic administrative census (GDAC)

6. An administrative census is perceived as a feasible option to be carried on in Israel in the short run. All people in the country have administrative records of one sort or another, and entries and departures to and from the country are well recorded. The two main challenges in the production of the geo-demographic administrative census (GDAC) reside in the population frame and in the population location within Israel (Blum & Feinstein, 2017):

A. The Population Frame

7. In the formation of the population frame Israeli citizens are dealt with differently from foreigners living in the country:

The CPR records are of all Israelis who have been citizens since November 1948. The Population and Immigration Authority (PIA) assign inactive status to those who passed away and to those who declared that they immigrate out of the country. Recently, a joint effort has been made, together with the Social Insurance Institute (SII), to identify additional emigrants who did not declare their out-migration. Israeli citizens who do not show signs of life in the SII records, although expected to do so, like collecting old age pension, are defined as potential emigrants. ICBS is taking one step forward and identifies Israelis whose usual residence is not in the country, by their crossing borders patterns. A database of Israelis

abroad is updated by adding or subtracting records of those who cross a threshold of presence in Israel or abroad. The threshold definitions take into account the ease of movement between countries and define 275 days abroad rather than 365 as a threshold to be counted out. However, since the data are needed for policy making regarding services to be rendered to the population, the definition for in-migration is not a perfect mirror image of out-migration. New immigrants are counted in from day one in the country, and returning Israelis are counted in if they are present on census day. They do not have to accumulate 275 day in Israel in order to be counted in. *Symmetry is not required.*

8. In the UNECE (2015) definition of usual residence, continuous stay is not taken literally anymore since assumed intentions are involved:

"The 'usually resident population' of a country is composed of those persons who have their place of usual residence in the country at the census reference time and have lived, or intend to live, there for a continuous period of time of at least 12 months.

A 'continuous period of time' means that absences (from the country of usual residence) whose durations are shorter than 12 months do not affect the country of usual residence".

9. In an administrative census context, *an operational definition is required for "intentions"* and the scope of coverage error in the census, depends on it. Therefore, although all people in Israel have administrative records, and entries and departures are well recorded, the population frame may carry under and over-coverage of Israelis who travel frequently.

10. As for the foreigners in the population frame, work-permits database includes records of all legal foreign workers whose employer applied for their hiring. Most of them leave the country within five years, but new permits are given on an ongoing basis. Meaning that at any point of time, a quota of about four years permits are of legal foreigners who belong to the census population. The individual foreigner is included one year after arrival, and excluded as he leaves. Other temporary residents are also reported in designated files; this is the case with infiltrators and with Jews who chose not to become citizens.

11. The illegal foreigners in the country are mainly those who overstayed their visa validity. In order to improve their identification and to reduce their over-estimation, the PIA has started a special record linkage project, where entries without departures and departures without entries are checked and linked carefully. Moreover, records of Israelis who arrived with a worker or a tourist visa and became Israeli citizens later on, are also crossed with records of entries without exits, since this group is more prone than others not to be identified in the CPR records (Different passport numbers, different last names, different spelling of foreign names in Hebrew etc.).

12. At the end of 2017, the estimated population was 8.8 million Israeli residents, excluding foreigners. The CPR comprised of 11.66 million records, 2.2 million inactive, and additional 0.6 million were defined as Israelis abroad.

13. At the end of 2017, the estimated number of legal foreigners (including infiltrators) was 200 thousand, and the estimated number of illegal foreigner remained in a wide range of 30–70 thousand.

B. Census Address

14. In a geo-demographic administrative census, defining the population location may be perceived in two ways: The ability to define a census address to each identified person in the population frame; or the ability to ascribe census address to each record in the population frame, in a way that ensures reliable estimates for small areas and small groups. This is the difference between the register approach and the statistical approach and it influences the census methodology. The Israeli CPR carries about 20% erroneous addresses on the

individual level but only 8% coverage error on the locality level (on average). Since this is not the required accuracy in a census and since coverage error can reach 40% in small localities, the aim in the geo-demographic administrative file is to assign the best census address to each identified record. It also enables the formation of administrative families.

15. The CPR erroneous addresses are of two types: Addresses of people who moved in or moved out from, without reporting the PIA, and addresses of people who have more than one address and only one should be defined as a census address. The additional addresses, found in other administrative files, point at most of the problematic records in the CPR.

16. ICBS plans to engage machine learning in order to generate a model to assign an address to each record, and the open question now, opened by Zhang (2018) is whether to produce probability scores to each address of a single person (summed up to 1) or to decide which address is the best (0 or 1 scores). One observation has already been made: although the first option can serve for the population estimates, the second should be used for the users.

17. Once we move conceptually to an administrative geo-demographic census, the continuous improvement process becomes a core issue in the planning. Adding qualitative data sources and improving the algorithm of address choosing, are to be done and evaluated on a yearly basis.

C. Demographic Attributes

18. The objective demographic attributes of a census, like age, sex, country of birth, and religion, are part of the CPR and its supporting files. The missing values scope is minute but the accuracy of some of the variables is not even for all population groups. For example, country of birth can be CSSR or Czechoslovakia rather than the specific republic in current borders. Year of birth for people, who were born in countries where birth registration was not mandatory, is not as accurate as for Israeli born population. These data are to be improved by previous censuses, and ongoing surveys.

19. The subjective attributes of a census, like gender and ethnicity (involving self-perception), cannot be covered in an administrative census.

20. Another major issue is the formation of families and households. Israel does not have yet a dwelling unit address, adopted by all authorities and individuals. Therefore, an address is usually a building address. Family relations in an address are constructed directly from the CPR and since people from an extended family can live in the same building, only nuclear families are defined as census entities. It is clear that cohabitation is not easily identified if children are not involved, and so is the case with same sex couples. Plans are made to engage as many as possible relevant stakeholders in the central government and in the local authorities, to enforce reporting on an identified dwelling unit level.

III. The sample survey: population and area

21. The working assumptions that the population frame is known, allows for the disengagement from the area. The GDAC is the sampling frame to be used for a census survey. Stratified sampling can be based on different variables, including geographic affiliation, and data collection can be done from afar, i.e. internet and telephone. It is a major change in the census methodology and in the expected statistical methodology.

22. The idea that the census refers to the population first and only then to the area seems to be more relevant in today's conditions of people moving freely between states and continents. The governments have to take into account wide margins of uncertainty regarding the population to be served.

23. The definition of the boundaries of official statistics has already gone through important changes:

In the SNA 1993 the reference is to the economic territory:

“The economic territory of a country consists of the geographic territory administered by a government within which persons, goods and capital circulate freely”.

In the SNA2008 it is defined more broadly as

“Any geographic area or jurisdiction for which statistics are required,” and “the area under the effective economic control of a single government”.

24. The survey aims to supplement socio-economic information and to correct population estimates in areas identified as carrying address errors in the CPR. For the later, the sample design can benefit from previous knowledge accumulated regarding reporting patterns and migration of different groups. For example, the 21–30 age group tends to update the CPR address only when got married; Arab men tend to migrate within the same locality; Recently inhabited area carries under-coverage error until it is fully populated. As for the socio-economic characteristics, area should be considered more where variance is high in order to enable small area estimation.

25. Administrative files being the backbone of the census is used to reduce the response burden selectively. For example, the education register is good for the people under the age of 65 but less good for the elderly. Moreover, people who migrated to Israel recently are less likely to have updated education record since they acquired their education abroad. Only those people with no information or with not updated information in the admin files will be asked to answer the education questions.

26. The combination of identified population, computerized data collection and advanced technologies of information security, is used to further reduce response burden by embedding a control file with navigation instructions on the individual level. ID number and a yes/no sign will navigate the respondent to a question only if the information is not known. For example, if the GDAF record of a person has USSR as a country of birth, he will be asked for the country of birth in today's borders. Others are not going to be asked this question. The background navigation file will include the whole population and not only the sampled records since data are collected for the household and we assume no knowledge of its structure and components beforehand.

27. Having identified individuals as sample units enables the census questionnaire to be addressed to them specifically, rather to an area with unknown population, and remote answering modes can be suggested. It will be the first time that gradual data collection internet-telephone area will be used and in this order.

28. Although the sample unit is an individual, the research unit is a household, meaning that answers by proxy are going to be provided with no interviewer involved. The quality of the answers is to be thoroughly checked before the census.

29. The statistical methodology, led by the National Statistician, Prof. Danny Pfeffermann, focus on small area estimates to be produced by using the survey information, the GDAC and other sources of information. The rational is that the estimates for “good areas” are not going to change the administrative estimates, while the estimates of “bad areas” in the admin files will do. Over-time, the contribution of the survey to the administrative population estimates will be reduced, until abolished.

IV. Concluding remark

30. The late shift in the census methodology in Israel brings about many risks to be carefully managed. However, the GDAC can be perceived as the safety-net together with quality assessment. A post enumeration survey is still to be planned. A debate regarding its purposes and derived methodology is not concluded yet.

31. Leaning more on administrative sources of information brings about new opportunities to the census processes and outcomes. Referring to identified population in a known population frame implies substantial change in the concept of census. Area boundaries become, in a way, a virtual entity rather than the main physical entity in a census. The theoretical and socio-economic implications, and the influence on policy making, should be further discussed.

32. Direct data collection can be significantly reduced if administrative information will be used to navigate the identified respondent to answer the questions with missing data only. The pre-requisites for this process are: tight information security, unique ID for all (or even partial) population, computerized working environment that enables on-line navigation.

33. Official statistics at large and administrative censuses in particular, will have to deal with different types of information sources. The availability of mass information and private data is to be harnessed to statistical processes and processing, carried on by data scientists.

34. After the 2017 census experiment, ICBS reached a crucial decision to abandon the rolling census and to go back to the administrative census path, paved after the 1995 census. However, the existing conditions (needs, data sources, technology) enable a paradigm shift, as suggested by the UNECE expert group in 2017 and ICBS plans to seize the moment.

References

Blum & Feinstein, 2017. Estimation of the Total Population in the 2020 Integrated Census in Israel. UNECE group of experts on population and housing censuses. Geneva.

UNECE Conference of European Statisticians (2015). Recommendations for the 2020 Censuses of Population and Housing. United Nations, New York & Geneva.

Eurostat, IMF, OECD, UN, World Bank (1993). System of National Accounts. Brussels/Luxembourg, New York, Paris, Washington, D.C.

Eurostat, IMF, OECD, UN, World Bank (2008). System of National Accounts. New York.

Zhang (2018). Towards a system of register-based population counts at ICBS. Consultation report. Israel Central Bureau of Statistics, Jerusalem.
