Economic and Social Council

Economic Commission for Europe

Conference of European Statisticians

Group of Experts on Population and Housing Censuses

Twenty-first Meeting
Geneva, 18–20 September 2019
Item 4 of the provisional agenda
Research on the use of administrative data for censuses

Problematic issues and possible solutions in the preparation of a register-based population census

Note by Statistics Estonia∗

Summary

The development of society brings about changes in many phenomena and concepts related to them that have been stable for a long time. This makes producing and using statistics about such phenomena complicated and often ambiguous, as the traditional system of statistics often does not have a place for new phenomena and concepts.

Statistics Estonia has solved such problematic issues using an innovative method that relies on the information in registers – to be more precise, on indices created by cross-usage of a great number of registers [1–4].

∗ Prepared by Ene-Margit Tiit.

I. Introduction

1. The number of people in a population, i.e. the number of permanent residents of a country, is generally found by means of a census. In recent times, however, under-coverage of the census has become a problem in many countries – during the census, it has proved impossible to contact all the people who should be counted. This has several causes:

(a) First, people are much more mobile than earlier;

(b) A second reason, however, is psychological – people do not want to share their personal data and they try to protect their privacy in all ways.

2. This is also a global problem which, so far, manifests itself mostly in developed countries, but it has become more pronounced as data protection issues and the cases of leaking and hacking of data are increasingly covered by the media. This presents a serious challenge for providers of official statistics.

3. Since the mid-20th century, statisticians have invented various methods to check the coverage of the census, i.e. to assess how well the number of people in a population as determined by the result of the census matches reality, e.g. by means of a follow-up census [5–6].

4. Usually, such methods permit assessment of the correct population size, but are not so helpful if person-based population statistics are needed, i.e. those which provide precise information about the distribution of inhabitants according to different markers.

5. To acquire such information, it is necessary to use additional data sources to adjust the data produced by an under- or over-covered census.

II. Registers and the system of registers

6. Relatively efficient methods for assessment of the size of the population can be used in countries with a functioning system of high-quality registers which meet the following main requirements:

(a) All objects (persons and dwellings) can be uniquely identified and the identifiers are universal for all the registers that belong to the system;

(b) Objects of different types (persons and dwellings) are mutually connected by identifiers;

(c) The registers are complete, i.e. they contain data of the population of this register;

(d) The information in registers is checked and updated regularly (at least once a year).

7. Estonia has a system of registers that meets the above requirements, and, in addition to persons and dwellings, it also contains businesses as objects with their identifiers.

8. Although the information collected in registers is generally based on documents, the possibility of errors is not excluded:

(a) The data can be outdated;

(b) They can contain accidental errors (which can be caused by carelessness or human error);

(c) There may be systematically erroneous data in someone's interests (most often of the person who submitted the data, sometimes also the register keeper).

9. A precondition for register-based statistics is continuous checking of the quality of the data contained in the registers, particularly their correctness (correspondence between the data and the actual situation).

10. A relatively efficient way for this is the cross-usage of registers, i.e. the data of one register are checked by means of other registers that contain independent data.

11. In most countries, however, the possibilities of such cross-usage are limited by data protection laws, and cross-usage is allowed only for organizations producing statistics, including Statistics Estonia. Statistics Estonia is not allowed to reveal the results of the cross-usage (including the discovered errors) to other organisations [7].

12. Since the law and good practice do not allow the same data to be requested from an individual several times (by different registers), the erroneous data of one register are carried to other registers – for example, this happens to the addresses registered in the population register.

13. From the viewpoint of data users, this is a serious concern, as the precision of addresses is one of the greatest problems in the Estonian system of registers – the addresses in the population register often do not correspond to reality or are not precise [7].

14. Nonetheless, Statistics Estonia has managed to solve several tasks of population statistics by using registers.

III. Correction of census under-coverage in Estonia

15. The most recent Estonian census, that of 2011, was under-covered. Signals about this came from census participants themselves. It was also revealed that the number of people counted was almost 5 per cent smaller than the number of people residing in Estonia according to the population register.

16. To get the actual number of people (and a list of people actually residing in Estonia), a decision was made to use several registers containing personal data. This was the first attempt to correct population data by cross-usage of register data [10, 11].

17. For this purpose, data on the inhabitants of Estonia were collected from more than ten registers and sub-registers. Considering that people's activity in registers depended greatly on their age, and somewhat also on their gender, twelve separate gender-age groups were observed.

18. In each age group, several different models were used – linear and logistic analysis as well as expert assessments for comparison. The results of the different models coincided to a considerable degree. Based on test data obtained from research, the errors in the models were assessed. These were below 5 per cent in all gender-age groups.

19. In conclusion, it turned out that the 2011 census was under-covered by 2.3 per cent [11]. From the year after the census, population statistics used census data that were corrected for under-coverage.

20. It was also revealed that the number of Estonia's residents was overestimated in the population register by almost 3 per cent, the reason being under-registration of departures from Estonia.

IV. Problems related to the country of residence

21. Although there are a number of international standards on how to determine a person's country of residence for the census, these standards and the exceptions to them cause several essential problems that complicate efforts to determine a person's country of residence.

22. Generally, a person becomes a permanent resident only after living in a country for 12 months, but an exception is made if a person has not lived in the country so long but intends to do so.

23. Inevitably, such a definition is contradictory, as it is not possible to check a person's intention in all cases (a person can respond on behalf of another person, but they may not know the intentions of the person s/he represents). It can also happen that the person him-/herself expresses an intention that is not serious or does not materialize for some reason.

24. In principle, the treatment of immigrants and emigrants should be symmetrical. Yet it is completely unrealistic to ask people who have left the country less than a year ago whether they are going to stay abroad for at least a year or plan to return in less than a year.

25. In the latter case, these people should be counted among the population of the country, although it is practically impossible to get data about them if they have no close relatives in the country of departure.

26. The relation between temporary and permanent residents is also somewhat confusing. Temporary residents are people who live at a certain place (address, settlement, country) for more than three months but less than 12 months, i.e. they are not permanent residents. Permanent residents, however, are people who reside at a place permanently, but these people can temporarily stay elsewhere (including in another country) if the length of the stay is no longer than 12 months.

27. Thus, if a person stays three to nine months of a year in one country and the rest of the time, which is also three to nine months in another country (which is not a rare case at all), his/her status as a permanent resident has to be determined according his/her earlier place of residence before migration – if s/he has not declared his/her intention to stay in one of these countries for a longer time. Actually, a person can even stay in three countries in the same way, and it is not easy to determine in which s/he is a permanent resident.

28. There seems to be no problem if the immigrant determines his/her country of permanent residence during an interview for a traditional census (either following the census taker's instructions or according to his/her own opinion).

29. Generally, this was the practice in the past. However, it is not clear how precisely the regulations and time limits were observed in such cases. Family members may have counted the departed household member as a permanent resident of the country even if s/he had been abroad for more than 12 months.

30. Fixing the country of residence is more complicated in the case of various internationally-agreed exceptions. For example, the country of residence of students of higher educational establishments is their country of study (despite possible migration back and forth between their home country and country of study). Also, it is not simple to determine "pendulum migration" unambiguously.

V. Finding population size by means of indices

31. The population size can relatively easily be corrected after the census according to registers, since population counted and registered as permanent residents of the home country versus uncounted persons and those registered as foreign residents can be used as "training datasets". Using the same algorithm for determining the precise number of people in a population during the whole period between the censuses is not appropriate, however, as census data become outdated quickly because of people's mobility.

32. The task was formulated – how to keep the list of permanent residents of the state continuously up to date, i.e. to determine at each moment (in practice, each year), on a personal basis, the set of permanent residents of the country and the persons who have left the country during the years (and who could potentially return). Both sets together form the set of persons that is henceforth called the extended population; these are the persons about whom there are or have earlier been data in Estonian registers (including the population register).

33. To define permanent residents, the residency index $R_i(k)$ was created according to the formula:

$$R(k)=d*R(k-1)+g*X(k-1), \qquad (1)$$

where (1) is the value of the residency index attributed to person i in year k (which changes on the scale from zero to one).

34.   In the formula $R(k–1)$ is the value of residency index of the person in the previous year, $X(k–1)$ is the weighted sum of signs of life [20] collected by person i in the previous year:

$$X(k-1)=\sum a_j b_j \qquad (2)$$

35.   Sign of life b_j corresponding to register j is a binary variable, the value of which is 1 if the person i has been active in register j in the previous year, and 0 if the person has not been active in this register in the previous year $(j=1, 2, ..., m)$, where figure m (number of terms in the formula (2)) is the number of registers and sub-registers used. This number is more than twenty in Estonia at present.

36.   So, the person acquires a corresponding sign of life if s/he, for example, visits a family physician, studies at a school in Estonia, receives social security benefits from the local government or buys a prescription drug.

37.   The first member of the formula (1) is the so-called stabilizing member which connects the person's residency status with the past, and its weight is determined by parameter d; the second member is the so-called life signs member, and its weight characterizes the activity of the person in the previous year [2].

38.   The weights a_j of signs of life b_j are determined based on the data of previous years as the ratio of conditional probabilities of occurrence of this sign of life in the case of definite residents and definite non-residents.

39.   The basis for determining parameters d and g $(d+g=1)$ in the formula (1) is population-political expert assessment: how quickly an immigrant might get the status of a permanent resident, and how long a person will be kept in the list of permanent residents if s/he does not show any activity [2]. In the present case the values $d =0.8$, $g = 0.2$ are in use. A value of threshold 0.7 is used, which, according to the data received until now, minimizes and also balances the errors of both inclusion and exclusion.

40.   Based on international rules, the period of stability is two years in both directions (which covers the limit of 12 months even if the signs of life appear at the end of the year), but particularly active persons can become residents within one year, although an additional condition is that they must have a (register-based) place of residence in Estonia.

41.   The residency index is calculated each year for all the persons belonging to the extended population, and, in year k, persons whose residency index is higher than threshold c, which is determined empirically, are considered residents.

VI.   Advantages and disadvantages of index-based assessment of the number of people in a population as compared to the traditional census

42.   The number of people in a population has been calculated in Estonia by means of an index since 2012, and since 2016 all population statistics have relied on the index-based concept of residency. The method has been checked by sampling studies where people's country of residence is determined by questioning.

43.   It has been revealed that, when comparing the index-based method with questionnaire data, the estimated inclusion and exclusion errors do not exceed 1 per cent [2].

44.   The advantages of determining residency by means of the index over the traditional question-based census are:

        (a)   Coverage: in principle, information is obtained about persons to be counted, regardless of where they are staying at the moment;

        (b)   Uniformity: the source for all persons' data lies in the same registers and the same calculation algorithms. Thus, the information does not depend on whether the person answers him-/herself or whether the data are obtained from his/her household members;

(c)   Objectivity: the result does not depend on the stated intentions of the person and the willingness to reveal his/her intentions, nor on the person's health or mental status or memory;

(d)   The probability of accidental errors, including human errors, decreases. In the case of question-based censuses, such errors are always inevitable to a greater or smaller extent.

45.   The greatest disadvantage of the index-based method is dependence on the quality of registers, including on how carefully residency definitions are followed in registers (e.g. whether any local government allocates benefits to persons who have actually left for abroad).

46.   Because of the great number of signs of life, errors in a single register are not very significant in deciding on residency.

47.   However, in the case of the index-based method, an absolutely passive permanent resident can also be left out of the population if s/he is not active in any of the registers for several years: that is, if the person does not work, does not receive any benefits, does not visit health care institutions, does not breach any laws.

48.   Such persons are protected from premature exclusion from the permanent population by the stabilizing member of the residency index (see formula (1)).

49.   It is also possible that a person residing abroad may be counted among permanent residents if s/he, when in Estonia, actively uses services there: visits the doctor, buys medicines, performs business transactions, etc. To prevent too early inclusion of them among residents, so-called delay time is used.

VII. Conclusion

50.   In conclusion, in the case of index-based assessment of the number of people in a population, errors are possible too, but their probability decreases if the quality of the registers improves and new data sources are applied. The accuracy of the index-based estimates is assessed through use of additional surveys, and the results are provided with potential estimation error values. Addition of new information (further signs) will result in consistent improvement of the accuracy of index-based estimates.

Bibliography

Tiit, Ene-Margit (2017). Residency Testing. Estimating the true population size of Estonia. Statistics in Transition, Warsaw, 2, 211−226.

Maasing, Ethel; Tiit, Ene-Margit; Vähi, Mare (2017). Residency index – a tool for measuring the population size. Acta et Commentationes Universitatis Tartuensis de Mathematica, 21 (1), 129−139.

Tiit, E.-M.; Vähi, M. (2017). Indexes in demographic statistics: a methodology using nonstandard information for solving critical problems. Papers on Anthropology, 26 (1), 72−87.

Tiit, Ene-Margit; Maasing, Ethel (2016). RESIDENCY INDEX AND ITS APPLICATIONS IN CENSUSES AND POPULATION STATISTICS. Eesti Statistika Kvartalikiri, 3, 41−60.

Tamm, Ebu (2001a) Sample Design for the Evaluation Study of the 2000 Year Census in Estonia.
http://epp.eurostat.ec.europa.eu/portal/page/portal/research_methodology/documents/60.pdf.

Tamm, Ebu. (2001c). Rahvaloenduse andmete kvaliteedi hindamisest. ESS Teabevihik nr 12, lk 8–14.

Registripõhise rahva ja eluruumide loenduse (REGREL) metoodika väljatöötamine. (2013). Koost A. Puur, L. Sakkeus, S. Aben. Tallinn. [www].

Tiit, Ene-Margit. Rahvaloendused Eestis (2011), Akadeemia, nr 12, lk 2276–2312).

Tiit, E.-M. (2013) Eesti rahvaarvu ja Eestist lahkunute arvu hinnangud. – Riigikogu toimetised, nr 27, lk 153–169).

Tiit, E.-M, K. Meres, M. Vähi 2012. Rahvaloenduse üldkogumi hindamine. —*Eesti Statistika Kvartalikiri*, nr 3, lk 79–95.

Tiit, Ene-Margit (2011). 2011.aasta rahava ja eluruumide loenduse alakaetuse hinnang. Eesti Statistika kvartalikiri 4,12.