

ASSESSMENT OF UNDER-COVERAGE IN THE 2011 POPULATION AND HOUSING CENSUS

Ene-Margit Tiit

Statistics Estonia, University of Tartu

Today, census under-coverage is a major problem in all developed countries: some of the people subject to enumeration do not participate in the census. They do not want to be enumerated for various reasons (a desire to protect one's privacy, reluctance to deal with any public authorities, a high degree of mobility), which means that enumerators cannot reach them. In most cases, these people also refrain from any type of self-enumeration. Therefore, the census team must always assess the level of under-coverage and may even have to revise the census results to compensate for under-coverage.

Methodology of under-coverage estimation

In order to estimate census under-coverage, we need to know the number of non-enumerated persons (subjects), V , and the number of persons who were subject to enumeration, which is the target population N . Based on the census results, we know the number of enumerated persons, which is the census population K . We can calculate the estimated target population (Euroopa ... 2010) using the formula $N = K + V$. This means that the rate of under-coverage is $V / (K+V)$. It is usually given as a percentage, indicating the share of non-enumerated persons in the target population. Consequently, the main purpose of under-coverage estimation is to determine the value of V , i.e. the number of persons who were subject to enumeration but were not enumerated.

There are several methods for assessing under-coverage, with a follow-up survey being the most popular option. In this case, a specially designed sample is used for repeat enumeration (using an abridged questionnaire) to determine the share of persons that were included in both the census and the follow-up survey and the share of those who were not enumerated during the census – this information allows a statistical estimation of census coverage. The weakness of this method is that it helps to discover those who were accidentally excluded from the census, but not those who systematically avoid enumeration. The use of administrative registers is currently the most suitable method in Estonia – considering that the next population and housing census is being planned as a register-based census, it is very useful to test the quality of registers and use these for demographic analysis. The methodologists of the population and housing census (PHC) have been analysing the quality of registers since 2008, when the possibility of a register-based census was first discussed. The initial analysis indicated that the quality of Estonian registers was not sufficient to replace the traditional census method (Estonian registers do not yet cover all the parameters required for Eurostat's output). However, registers are useful in several stages of the census, including in the determination of the number of non-enumerated permanent residents, thus providing a reliable estimate of census under-coverage.

A more detailed discussion of registers and the possibilities of using registers to determine the number of non-enumerated permanent residents can be found in the article "Enumerators' activity after the Census" (Tiit, Vähi 2012). A mathematical overview of the rules applied has been given in the articles "Assessment of the target population of the census" (Tiit et al 2012) and "Loendusmeeskond jätkab tööd" (Tiit 2012); these results have also been presented at an international seminar (Joint ... 2012).

The Population Register is the most important register used. It should include all people residing in Estonia, but there is slight over-coverage in terms of Estonian residents, because some of those who have moved abroad have not registered their departure in the Population Register. The estimated number of unregistered illegal immigrants in Estonia is relatively small, because

Estonia has pursued a fairly strict immigration policy. Also, Estonia is not a particularly attractive destination for immigrants. Other important registers include the **Health Insurance Database** covering over 90% of Estonian residents, the **Register of Taxable Persons** covering a large portion of the working-age population, and the **Estonian Education Information System (EHIS)** containing data on almost all students. Additional information can be obtained from the **State Pension Insurance Register** (includes the recipients of pensions, family allowances and benefits), the **Register of Social Services and Benefits (STAR)** and the **Traffic Register**. Over a dozen registers and sub-registers in total were used to determine the number of non-enumerated residents, whereas the decisions were based on the activity of the persons analysed in the respective register databases in 2011.

The following persons were considered and analysed as potential residents: all persons included in the Population Register whose place of usual residence as at 1 January 2012 was in Estonia and who had not been enumerated during PHC 2011 as permanent residents, emigrants or temporary residents. Persons who had an Estonian personal identification code but who (according to the Population Register) were not permanent residents of Estonia at the census moment, were not considered potential residents and were not analysed. The inclusion of those persons in the analysis would have increased the number of potential residents by less than one per cent of the actual population, but it would also have increased the probability of inclusion errors. Five rules (models) were used for analysis; a detailed description of these rules can be found in the articles "Enumerators' activity after the Census" (Tiit, Vähi 2012) and "Assessment of the target population of the census" (Tiit et al 2012). The rules were as follows:

- Linear discriminant analysis assuming that theoretical errors are equal;
- Linear discriminant analysis assuming that the probability of Type 1 errors is less than 0.05;
- Logistic discriminant analysis assuming that theoretical errors are equal;
- Logistic discriminant analysis assuming that the probability of Type 1 errors is less than 0.05;
- Expert assessment.

In the discriminant analysis, the judgement rule was based on the population of 'verified residents' – persons who had been enumerated as permanent residents and were also permanent residents of Estonia according to the Population Register; and on the population of 'verified non-residents' – persons who had been enumerated as emigrants and were also not Estonian residents according to the Population Register.

Following the recommendation of the experts of the PHC Scientific Council, the census team tried to minimise the probability that persons are erroneously counted as residents. An asymmetrical judgement rule was used for that purpose in two cases: the probability of a person's erroneous classification as a resident (Type 1 error) was fixed at a pre-determined level, which was 0.02 for most age-sex groups and less than 0.05 in aggregate. Naturally, this also increased the probability of Type 2 errors, i.e. the erroneous inclusion of a person in the group of non-residents.

The same source data – activity of the analysed persons in administrative registers during 2011 – were used in all cases. An automatically selected optimal set of parameters (with best differentiation and not correlated), selected from the group of source parameters, is used in discriminant analysis models. This set was somewhat different in case of linear discriminant analysis and logistic discriminant analysis. The selection of parameters for expert assessment was based on an analysis of the content of registers. All judgement rules produced relatively similar results, but the number of persons classified as residents was somewhat smaller when asymmetric criteria were applied (and bigger when symmetric criteria were applied).

In order to obtain an integrated criterion, the aforementioned five rules were treated as equivalent, with a person deemed as a resident if four of the five criteria supported this conclusion.

It appeared that 46% of all analysed persons could be counted as residents in such a case. The remaining persons in the survey sample were classified as non-residents. This resulted in 0.044 probability for Type 1 error (inclusion). Consequently, the number of persons erroneously counted as residents would constitute less than 0.1% of the total population, while the number of persons erroneously counted as non-residents would be less than 0.5% of the total population.

In case of all the persons categorised as residents as a result of the analysis, a large portion of the information asked in the census is available in registers. Consequently, it would be possible, in principle, to include them in the census population (a practice used in Latvia, for example). This would also be allowed by the Estonian census methodology, which includes provisions for the use of registers in the census, incl. in imputation. However, no harmonised international methodology for under-coverage correction has been developed or recommended (it probably cannot be developed in the near future). Therefore, most countries submit their census results without adjustment. Statistics Estonia has decided to do the same.

Assessment results

The estimated number of non-enumerated residents was approximately 28,000, based on the methodology used.

The size of the census population is 1,294,455 persons. This means that the estimated target population would be more than 1,322,000 persons. This result is consistent with current population statistics (which include migration).

The rate of under-coverage is $28,000 / 1,322,000 = 2.1\%$, which means that the corresponding coverage rate is 97.9%.

Under-coverage by county

The rate of under-coverage is not the same everywhere – in some counties the rate is almost twice as high as in some other counties. Higher than average levels of under-coverage were detected in Valga and Ida-Viru counties, i.e. in areas next to the border where the share of people of other ethnic nationalities is relatively higher (Figure 1, p. 112).

Under-coverage by sex and age group

The level of under-coverage was the lowest in case of school students and increased rapidly among young men (Figure 2, p. 112). In case of the working-age population, the rate of under-coverage was considerably higher among men than among women. This can be explained by men's generally smaller motivation to take part in surveys, and possibly by some men's disregard for (social) norms. It is also possible that some long-distance commuters were not enumerated – these are people who work abroad but whose families live in Estonia and who spend most of their free time here.

Under-coverage based on citizenship, ethnic nationality and mother tongue

*In order to know who the non-enumerated persons are, the census team analysed their distribution by ethnic nationality, citizenship and mother tongue – official data on all these parameters have now been published and are also available in the Population Register. Distribution by **mother tongue** was the most informative (Figure 3, p. 113), even though the Population Register did not contain any data on this parameter for one fifth of the persons categorised as residents in the analysis. Of all non-enumerated persons with a known mother tongue, nearly a half (46%) had a mother tongue other than the official language – it is likely that the census information campaign did not reach them. Importantly, a significant share of these*

people had a mother tongue other than Russian, which was the second language of the information campaign.

In terms of the **ethnic nationality** of the non-enumerated persons (Figure 4, p. 114), 13% of these persons had not specified their ethnic nationality in the Population Register. But Estonians represented only 55% of those whose ethnic nationality was specified. Consequently, a large share (probably at least half) of the non-enumerated residents are non-Estonians. This also explains why there were more of this type of exclusions in Ida-Viru county where Estonians are in the minority. In Valga county, the coverage rate was lower than in other counties, which was probably caused by the considerable share of Latvians in the population and their low level of inclusion in the census.

For comparison, Figure 4 (p. 114) includes Estonians, who had the lowest under-coverage rate (1.5%). The figure shows that people from countries that are physically and culturally distant from Estonia were more likely to avoid participating in the census.

Citizenship has a somewhat smaller impact on coverage, with a quarter of the non-enumerated persons being without Estonian **citizenship**. Figure 5 (p. 114) outlines the countries whose citizens are represented by at least 1,000 residents in Estonia.

Conclusion

The coverage rate of the 2011 Population and Housing Census was 97.9%, which is very close to the reference figure 98% established by the PHC Council as an indicator of very good performance.

Approximately one half of the non-enumerated residents are non-Estonians whose mother tongue is not Estonian. This group includes representatives of many different countries and languages. Many of them were probably foreign exchange students or migrant workers, who at the census moment were Estonian residents according to the enumeration rules.

It is likely that non-Estonians, in particular those who do not speak Estonian, were not motivated to take part in the census, especially if a census had already been conducted in their homeland.

Furthermore, the non-enumerated residents included a large share of working-age men, many of whom were probably not at home in the census period, are long-distance commuters or simply did not want to talk to an enumerator (there were several reports after the census about this kind of reluctance to be interviewed). Lack of any activity in registers could have been a reason for non-enumeration in case of some elderly persons.

Consequently, the analysis did not confirm the hypothesis that most of the non-enumerated persons were Estonians who have left Estonia but continue to use Estonian public services due to previous connections and who are, as a result, still listed in Estonian registers. Only less than a half of the persons analysed could theoretically belong to this group.