

## Methodology of under-coverage estimation used in Estonian PHC2011

Ene-Margit Tiit, Mare Vähi,

University of Tartu

In order to estimate census under-coverage, it is necessary to know the number of non-enumerated residents (subjects),  $V$ , and the number of census population (residents who were enumerated)  $K$ . The target population (real population size) is  $N$ . We can calculate the estimated target population 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. Knowing the percentage of under-coverage the real population size can be calculated using weighting. 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 and the quality of registers has been more or less thorough assessed.

The initial analysis indicated that the quality of Estonian registers was not sufficient to replace the traditional census method by register-based census. 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.

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. 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 – all who either are Estonian resident and get legal income, and the **Estonian Education Information System (EHIS)** containing data on practically all students studying in all stages of Estonian education system (including also kindergartens). 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 **motor third-party liability insurance 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.

A special question was in Estonian census 2011 to identify the persons emigrated during the last 11 years: Has anybody from your close relations left Estonia during the years 2000—2011 and has not arrived? All the emigrants were identified by their ID-code.

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 were regarded as potential residents of Estonia. Also, they had Estonian identification code. Such 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 would have increased the probability of inclusion errors. Several rules (models) were used for analysis. The rules were as follows:

- Linear discriminant analysis assuming that theoretical errors are equal;
- Logistic discriminant analysis assuming that theoretical errors are equal;
- 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.

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.

In order to obtain an integrated criterion, the five rules (where in discriminant analysis the sizes of allowed errors were somewhat varies) 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.

The people belonging to over-coverage set were not included into census population and all census tables (Hypercubes) were calculated without them. But the corrected population size will be used in population statistics.

### **Assessment results**

The estimated number of non-enumerated residents obtained 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 1,323,000 persons. This result is consistent with current population statistics (which include migration).

The rate of under-coverage is  $28,500 / 1,322,000 = 2.2\%$ , which means that the corresponding coverage rate is 97.8%.