

Estonia's first register-based pilot census

The aim of piloting the register-based population and housing census (REGREL) was to test the quality of the registers¹ used to form the census characteristics, the functioning of the methodology, and readiness of the support software for a register-based census. Like a regular census, a pilot census is conducted by establishing a moment of census and forming a census database as of that moment. The created data tables are verified and the quality level of data is determined in relation to the established criteria and the quality requirements for statistics. Any weaknesses detected in the pilot census will be addressed with relevant measures and the results will be tested in a second pilot census in 2019.

The result of the first pilot census confirms that Statistics Estonia is ready for a register-based census. The state of registers was analysed, based on the requirements of a census, and the quality of census characteristics was evaluated (Table 4).

The pilot census covered the census characteristics, which are mandatory for the Member States of the European Union, except for the characteristics of 'occupation' and 'location of place of work'. In addition, the analysis included the characteristics of 'ethnicity' and 'native language'.

Legal basis of the pilot census:

- Official Statistics Act
- Government of the Republic Order No 391 of 17 September 2015, "List of Statistical Actions of Statistics Estonia from 2015 to 2019"
- European Parliament and Council Regulation (EC) No 763/2008 of 9 July 2008 on population and housing censuses and the respective implementing Regulation (EC) No 1201/2009 of 30 November 2009 and Regulation (EC) No 519/2010 of 16 June 2010

Moment of the pilot census: 31.12.2015

Period of actions of the pilot census: 2.1–8.12.2016

Total population of the pilot census:

- Total population of persons in Estonia
- Total population of dwellings located in Estonia, irrespective of occupation status, and occupied non-residential premises

Goal of the pilot census

Testing the register-based organisation of the population and housing census in accordance with the Eurostat rules and the quality requirements for censuses. Acquiring data from 24 registers and performing data processing and analysis.

Methodology

REGREL is a register-based census, which means that the dataset of REGREL is extracted from registers. Before using registers data as census data, the quality of data is verified in reference to basic statistical criteria. When data are included in different registers, they can be used for verifying the quality of data on the one hand and for selecting the most reliable values in accordance with the developed methodological rules on the other hand.

Generally, census characteristics cannot be acquired directly from registers, because registers have been designed for other, non-statistical purposes and most of the definitions used differ from statistical definitions. It means that data from multiple registers have to be used in order to form certain census characteristics (e.g., the characteristic of 'activity status' requires data from more than 10 registers), while some characteristics are covered by duplicate information in several registers.

¹ In this document, the term 'registers' refers to state registers.

The methodologists of REGREL have solved the following main problems connected with the forming of census characteristics:

- Analysing the relationship of census definitions and value scales with the definitions and value scales used in registers;
- Testing the quality of registers and making efforts to urge register holders to eliminate any shortcomings;
- Determining the number of registers required for forming and, if necessary, verifying each census characteristic and, if some characteristics were not covered by any registers, taking steps to ensure creation of a respective register or register part;
- Establishing optimal rules for forming each census characteristic based on register data and creating the necessary software, using the quality of output characteristics as the basis of optimisation;
- Designing a strategy for using alternative rules when data gaps prevent the use of the optimal rule, as well as for imputation of values based on statistical or logical rules in some cases.

Total population of persons

The total population of persons means the number of residents, or the population size, of the country. This characteristic cannot be formed on the basis of the population register alone, because the previous population and housing census showed an over-coverage in the population register. The problem was solved by developing a residency index, which measures the activity of persons in registers; this aggregate information is then used to decide if a person is a resident of Estonia or not. Greater visible activity of a person in registers, i.e. a larger number of 'signs of life' in registers, corresponds to a higher probability that the person resides in Estonia. More information on the methodology of the residency index is available in the Quarterly Bulletin (http://www.stat.ee/publication-2016_quarterly-bulletin-of-statistics-estonia-3-16)

The total population of the census included 1,315,944 persons.

Total population of households

In the first pilot census of REGREL 1,254,705 persons were distributed between private households and 12,320 persons between institutional households (Table 1). Considering the average size of private households, the share of persons in institutional households and the number of persons who could not be associated with a household due to insufficient residency information, the under-coverage of the population of private households amounts to about 20,000 households.

Table 1. Persons and households

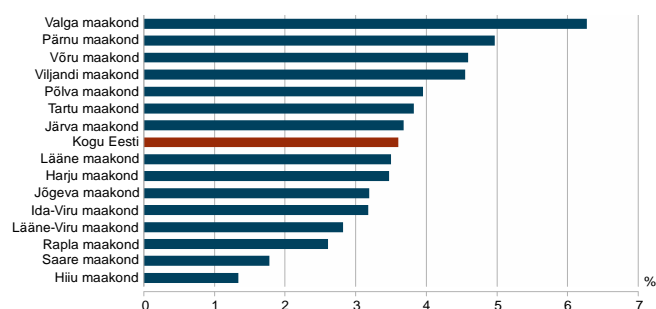
	Number of persons	Number of households
Private households	1,254,705	527,443
Institutional households ^a	12,320	–
Place of residence determined at municipality level	47,358	19,500 ^b
Place of residence unknown	1,561	500 ^b
Total	1,315,944	

^a Members of institutional households are not distributed between individual households.

^b Estimated number.

A person's place of usual residence is the place of residence, which is entered in the population register. However, in case of some persons, the population register entry is not specific enough to include dwelling data. The share of persons, whose place of residence has been entered only at municipality level, varies between different counties (Figure 1). The share of such persons is highest in Valga county and lowest in Hiiu county. A larger share of persons, whose place of residence has been entered at dwelling level, results in a more accurate and higher quality of household and family characteristics.

Figure 1. Share of persons with place of residence entered with the specificity of municipality in different counties



Total population of dwellings

Table 2 shows the breakdown of the total population of dwellings by type.

Table 2. Total population of dwellings by type and occupation status

Type of dwelling	Number of dwellings	Number of unoccupied dwellings	Number of occupied dwellings	Occupied dwellings, %
Conventional dwellings	725,539	199,074	526,465	72.6
Other housing units	978	0	978	100.0
Collective living quarters	4,359	0	4,359	100.0
Total	730,876	199,074	531,802	72.8

According to the definition, collective living quarters and other housing units have to be occupied. Conventional dwellings can be occupied or unoccupied. As 96.3% of the population is distributed between dwellings, the total population of occupied dwellings is undercovered and the total population of unoccupied dwellings is overcovered. According to estimates, if all persons were distributed between dwellings, the number of occupied conventional dwellings would be 546,800 or 75.4% of the total number of conventional dwellings.

Table 3. Sources of information on dwellings

Source of information	Unoccupied conventional dwellings	Occupied conventional dwellings	Other housing units	Collective living quarters	Total	Relative importance of the source, %
ADS	8,625	19,082	479	1,726	29,912	4.1
ADS and EHR	190,449	507,383	499	2,633	700,964	95.9
Total	199,074	526,465	978	4,359	730,876	100.0

Data on 4.1% of dwellings is available only in ADS; the technical characteristics of these dwellings can only be found in the data of PHC2011.

Quality of census characteristics

When estimating the quality of census characteristics, we should observe whether the definition of a characteristics corresponds to the international definition of the corresponding census output characteristic.

Coverage is one of the main quality criteria for censuses. This estimate is based on the share of missing values of individual characteristics. If the share of missing values exceeds a target level, the values have to be replaced – imputed – using mathematical or statistical rules. The most thorough verification of data quality is possible when we compare the breakdown of census characteristics with the corresponding breakdown of characteristics from a different source. When using this method, it is important to keep in mind that the breakdowns in two sources can be different: the difference can be caused by a sampling error when comparison is made to a sample survey, by

changes taking place over time in case of comparisons to the previous census, or by a different definition in case of another (comprehensive) source.

The share of missing values was estimated in the current quality analysis of the REGREL pilot census. Furthermore, definitions were assessed in terms of compliance with the definitions of the Regulation; they were fully compliant in most cases. Consistency with references sources was assessed based on suitably derived indicators.

The average quality rating of individual characteristics on a five-point scale was 3.4, i.e., 'satisfactory'; the population data had the average quality rating of 3.9 and the dwelling and household data had the rating of 3.1. Seven characteristics had a poor quality rating – all of them were characteristics related with place of usual residence (Table 4).

The quality of output cubes² was estimated in addition to the quality of individual characteristics, which was estimated on the basis of expert assessments relying on objective criteria. The average quality rating of cubes on a scale of 0–1 was 0.73. A rating was not calculated for six cubes, which included two census characteristics that were not formed in this pilot census. 13 of the 41 output cubes received an excellent quality rating and 17 had a satisfactory rating. The quality of five cubes was rated as 'poor' – these are cubes with a small number of characteristics, which include information on the place of usual residence and the type of household/family. Considering all the criteria, the overall quality of REGREL characteristics is satisfactory and the quality of population characteristics is good. As it was only the first pilot census, this is quite an acceptable result. Characteristics related with place of usual residence require the most work during the years before the next census.

Conclusion

The first pilot census of REGREL has been successfully completed. The pilot census showed that a register-based population and housing census is feasible and the preparations for the census have been purposeful.

All total populations of the first pilot census and all mandatory characteristics envisaged for the pilot census were formed. The largest amount of work was associated with the development of a methodology for total populations. A residency index was created to support forming the total population of persons. A combination of ADS and EHR is used to form the total population of dwellings, because EHR alone is undercovered. The number of populated dwellings and private households is undercovered, because for 3.7% of the persons the place of residence is not known with the specificity of a dwelling.

Implementing the ADS standard has been the greatest step in the preparation of REGREL after PHC2011. This has made linking place of usual residence and dwelling data much easier.

The first pilot census of REGREL was based on data from 24 registers; automatic acquisition was used in case of three registers (RR, ADS, KOPIS) and database excerpts were used in case of 21 registers. The largest number of databases was needed to form the total population of persons (Table 5).

The quality of census characteristics was checked for conformity with the quality criteria (according to European Parliament and Council Regulation (EC) No 763/2008). Of the total 38 census characteristics formed, the following had the best quality ratings: sex, age, legal marital status, country of birth, country of citizenship, total population, ethnicity, native language, location of dwelling, living quarters by type of building (Table 4).

The quality of the mandatory output of REGREL (hypercubes) was estimated according to the quality criteria of the pilot census. 13 of the 41 output cubes received an excellent quality rating and 17 had a satisfactory rating.

The largest problem for REGREL is the difference between registered and actual places of residence. This affects the breakdown of the lowest level of the place of usual residence (municipality) and all household and family characteristics (Figure 2). The situation could be improved (to achieve a better match between actual and registered data) by motivating people to register their actual addresses in the population register. This can be facilitated by awareness campaigns by local governments and the media.

² Census output is presented as cubes, i.e., 6–8-dimensional tables. The quality rating of cubes was calculated as the average of ratings of the individual characteristics included in a cube.

Figure 2. Families by type of family nucleus

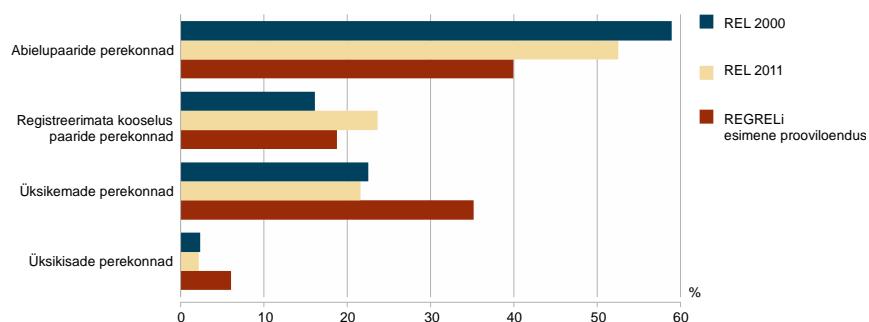


Table 4. Aggregate rating for the quality of characteristics

Rating	Characteristic
5 (excellent)	sex, age, legal marital status, country of birth, country of citizenship, total population, ethnicity, native language, location of dwelling, dwellings by type of building
4	employment status, educational attainment, type of dwelling, number of occupants, useful floor space, number of rooms in dwelling, water supply system, toilet facilities, type of heating, density standard
3 (satisfactory)	place of usual residence one year prior to the census, employment status, industry, year of arrival in the country, household status, size of private household, size of family nucleus, housing arrangements, occupancy status of conventional dwellings, bathroom
2	living quarters by period of construction
1 (poor)	place of usual residence, household occupancy status, locality, family status, type of family nucleus, type of private household, type of ownership

^a Explanations of database abbreviations are provided in Annex 1.

^b Data on the location of place of work and on occupation can be acquired in the future from the register of employment (TÖR).

Annex 1. Abbreviations

ADS	Address Data System
ARIREG	Commercial Register
EHIS	Estonian Education Information System
EHR	State Register of Construction Works
EMPIS	Estonian Unemployment Information System
EMSR	Estonian Medical Birth Registry
ETR	Register of Residence and Work Permits
KIR	Register of Prisoners
KIRST	Health Insurance Information System
KMAIS	Register of Identity Documents
KOPIS	Register of Mandatory Funded Pension
KR	Land Register
KVKR	National Defence Obligation Register
MKR	Register of Taxable Persons
PKR	Estonian National Pension Insurance Register
PHC2000	Population and Housing Census of 2000
PHC2011	Population and Housing Census of 2011
RETS	Estonian Medical Prescription Centre
RKOARR	State Register of State and Local Government Institutions
RR	Population Register
SAP	The State Human Resources Database
SPR	Estonian Causes of Death Registry
STAR	Social Services and Benefits Registry
TÖR	Register of Employment
Liiklusregister	Estonian Traffic Register