

The Population Census 2021 in Latvia

B. Zukula¹ and J.Jukams²

¹ Central Statistical Bureau of Latvia
e-mail: baiba.zukula@csp.gov.lv

² Central Statistical Bureau of Latvia
e-mail: janis.jukams@csp.gov.lv

Abstract

A totally different approach from the traditional census is the register-based census that was developed by the Nordic countries in the 1970s. Denmark was the world's first country to conduct a fully register-based population and housing census in 1981. Under this approach there is no direct collection of data from the population, and the traditional enumeration is replaced by the use of administrative data held in various registers (population register, building/address register, social security register, etc.) through a matching process, making use of personal identification numbers. Information not available in registers is imputed. This approach permits the production of census data at a greatly reduced cost and with relatively limited manpower, once a good quality system of statistical registers has been established (UNECE, 2018).

In 2021, for the first time, a solely register-based census with a reference date of January 1, 2021 was conducted in Latvia.

First, for population estimation, a method developed in 2012 and based on logistic regression model was used to classify all residents in the Population register into two classes - de facto residents and de facto non-residents of Latvia (Aināre et al., 2021). The dependent variable in the model was derived from Census 2011 data and is a binary variable with values 1 denoting de facto residents and value 0 denoting de facto non-residents respectively. Independent variables or predictors in the model are also binary variables describing individual's features (204 regressors in total) and were derived from several administrative data sources corresponding as close as possible to the year 2011. Population and Housing Census 2011 showed that usually resident population of Latvia accounted for 2 074.6 thousand at the beginning of 2011, which was 7% fewer people than in the Population register supervised by the Office of Citizenship and Migration Affairs (2 228.0 thousand). The model has been used for population estimation at CSB Latvia since 2012 and only in a few cases re-estimation of the model's parameter values has been done (e.g. in cases when regressors should be dropped from the model due to changes in administrative registers which (if changes in register occur frequently) can be a problem for supervised approach in general).

Next step, the availability, quality and reliability of different registers and administrative data sources was evaluated. Latvia has many high-quality, regularly updated registers and databases with a good coverage, such as the Office of Citizenship and Migration Affairs, the State Revenue Service, the State Social Insurance Agency and other state information systems, as well as other registers and databases where information on both individuals and dwellings is available. One major factor that facilitates the statistical use of administrative data records is the use of unified identification systems across different sources (Tønder, 2008). Single personal identifier (personal identification number) is used in all registers and databases. It allows data linkage to provide better data coverage, develop methods that are based on data from

several sources. Since 2012, work was started to improve cooperation with Latvia's largest administrative registers and to develop appropriate methodology for preparing all mandatory census indicators. Before deciding whether to use administrative registers in Censuses, it is necessary to develop methods for assessing the quality of registers, their metadata and data. The starting point for the quality assessment is the common statistical quality framework which has been developed at the level of the European Statistical System and the United Nations Statistical Commission (Dygaszewicz, 2020).

The necessary data for census variables in administrative data sources were identified at the level of persons and dwellings (addressing codes) to ensure the preparation of individual level indicators. This cooperation between institutions contributes to improving the quality of registers and databases, making the use of these data more efficient for the public.

During the preparation process it was acknowledged that the state's administrative registers alone cannot provide all the information required for the census as certain groups were missing or information was incomplete. Therefore, data from non-governmental institutions (e.g., artist unions) and private companies (e.g., water and sewerage service companies) were gathered.

In the end, information from 34 different registers and information from municipalities on demolished and uninhabitable residential buildings, as well as information from municipal companies on the provision of district heating, district water/sewerage (data from more than 70 companies) were used for census 2021. When conducting a Census, the essential features of a population and housing census defined by the International Conference of Statisticians (1853) and redefined and highlighted by the Conference of European Statisticians (2015) have to be taken into account (UNECE, 2018). The mentioned features must be ensured regardless of the census methodology used. It enables countries to provide internationally comparable data.

All five essential features of a Census were ensured in Census 2021 in Latvia, i.e.

Individual enumeration	The principle of individual enumeration is a fundamental feature for any census of population. In the case of register-based censuses it is important that each census unit (i.e. individual or dwelling) has a special, uniquely identified record in the registers used. For Census 2021 the information on each person and each dwelling was recorded separately in the database, using individual personal and dwelling codes. It is possible to link individuals with dwellings (address code for both is used).
Simultaneity	The information on persons and dwellings obtained for Census 2021 from administrative data sources corresponds to a certain reference period or a specific point in time (critical moment). 1 January 2021 was used.
Universality (within a precisely defined territory of a country)	Census data must cover the whole country and all population groups. In Census 2021 Population Register (covering all individuals) and the State Land Service Real Estate Cadastre (State Immovable Property Cadastre) Information System (covering all dwellings) were used.
Small area data	Census 2021 provides geo-referenced data on the number of inhabitants and dwellings and different variables, as well as small subgroups of the population. Address coordinates are added to each person and dwelling. It allows to provide very detail data.
Defined periodicity	An advantage of a register-based census is the opportunity to conduct the Census more often than every ten year. Register data are permanently available and more regularly updated. It allowed to prepare test data already annually on 01.01. before Census reference date.

Advantages of register-based census:

- Reduced face-to-face interviewing - during the period of Covid-19 restrictions (in 2020 and 2021), the face-to-face interviewing in data collection was stopped. In Latvia, Covid-19 restrictions didn't prevent the timely preparation of Census data in accordance with national and international legislation, that could be a challenge in case of traditional census.
- Reduced respondents' burden - the data and methodology developed will be used not only for the purposes of the Population and Housing census, but also in regular statistical sample surveys, thus reducing further the respondents' burden.
- Data are available faster - population and its indicators have been published already in the 2nd quarter of 2021, family and household indicators - in the 3rd quarter of 2021, housing indicators - in the 4th quarter of 2021, but variables on level of education and economic activity in the second quarter of 2022.
- Three times cheaper - in 2021, the traditional census would cost 10 mln euro, while using only administrative data sources, it costed around 2.9 mln euro.
- More accurate, more relevant data according to the classification (for example, occupation, industry) - in administrative data the employers indicates more precisely the employee's occupation. At the same time in the surveys the respondents use every day language to describe activity for which sometimes challenging to identify classification codes.

Disadvantages of register based census:

- Variables that are not available in administrative data sources - impossible to provide variables that are not in administrative data sources, for example, native language, language spoken at home, religious. However, this information can be acquired via surveys (for example, in Latvia questions on native language and language spoken at home were included in the international migration survey and will be collected in Adult education survey in 2022).
- Not fully covered information - even beside administrative data imputation methods are used, it is impossible to fully cover illegal employment and de facto place of residence of individuals within the country.

As mentioned before Nordic countries already at the end of 20th century proved that register based census is viable alternative for traditional census if country can develop and maintain good quality registers, further improve quality and expand the registration system. CSB Latvia will not return to traditional census. Our next task is to further cooperation with registers, identify new variables in surveys that can be replaced by administrative data sources. Another important aspect regarding population estimation using register/model based methodology is quality estimation - how do we ensure that our estimates are precise? Regular census coverage surveys are unavoidable in the future to ensure the quality of register/model based estimates beyond Census 2021.

Within the European Union many countries have taken the decision that 2021 was the last traditional census. Census information will be used as basis for register based systems that will allow to carry out register based censuses. Countries request that this approach to be taken into account when developing new regulation on population statistics and census. The variables and obligations mentioned have to be based on systems that are fully or will be fully register based in future. It means that only variables available in administrative data sources or created from information in administrative data sources will be available within the census.

Keywords: Register based census, census features, logistic regression model, supervised learning.

References

I.Aināre, M.Liberts, B.Zukula, S.Šulca, J.Valkovska, B.Opermanis, A.Jurševskis, K.Lece, A.Ceriņa, J.Breidaks, J.Jukāms, R.Beināre (2021) Method used to produce population statistics. Methodological report, Central Statistical Bureau of Latvia, Riga, Latvia. Available at: https://stat.gov.lv/sites/default/files/Metadati/iedz_Metodologija_ENG.pdf

United Nations Economic Commission for Europe (UNECE) (2018) Guidelines on the use of registers and administrative data for population and housing censuses. New York and Geneva, United Nations. Available at: <https://unece.org/fileadmin/DAM/stats/publications/2018/ECESTAT20184.pdf>

Dygazzewicz J. (2020) Transition from traditional census to combined and registers based census. Statistical Journal of the IAOS, vol. 36, no. 1, pp. 165-175, 2020. DOI: 10.3233/SJI-190566

Tønder J.K. (2008) The Register-based Statistical System. Preconditions and Processes. International Association for Official Statistics Conference Shanghai October 14 – 18, 2008. Available at: <https://www.fao.org/3/I9360EN/i9360en.pdf>