

THE REGISTER-BASED POPULATION AND HOUSING CENSUS: METHODOLOGY AND DEVELOPMENTS THEREOF

Ene-Margit Tiit

Statistics Estonia, University of Tartu

Estonia is making preparations for a register-based population and housing census. The article will give an overview of the background of the register-based census and explains what the benefits of a register-based census are and why conducting such a census requires a different methodology. The article will also explain what the prerequisites for conducting a register-based census in Estonia are.

The background of register-based censuses

Censuses conducted and databases created

Censuses are among the oldest statistical activities in the history of the human kind, but the objectives and methodology of censuses have changed repeatedly throughout history. While a few thousand years ago, piling stones in the city square was enough to determine the number of citizens (Misiakoulis 2010), later rulers have wanted to know increasingly more about their subjects or electorate. This is how the census questionnaires – that enumerators, going from door to door, filled out with the assistance of the persons enumerated – came to be. The first census that took place in the territory of Estonia (in 1881) used questionnaires with 15 questions (Lember, Kollo 2010). Over time, the number of questions has increased – in the last census, the number of questions asked from the people of Estonia reached 40 (Tiit 2014). The methodology of collecting data has also changed. A big leap in census methodology took place in Estonia in 2011 when online questionnaires that enumerated persons could fill out by themselves were used alongside interviews. This turned out to be acceptable for the majority of the population. The interview also differed from previous ones – enumerators used a laptop to fill out census questionnaires.

However, there were some persons who remembered that they had already submitted a large portion of their data – for example upon registering their place of residence. Justifiably these persons asked: why is the state requesting data that it already has?

The question is whether the national statistical system can use data which was previously collected for other purposes to conduct the census. Certainly this would have been rather cumbersome at a time when people were only registered in the church books, where births, confirmation, weddings and deaths were written down. Looking up data in the thick volumes used for storing data on births, deaths and weddings, by family, in the Republic of Estonia in the years 1920–1940 wouldn't have been much easier either. What is important is that in the countries where collection and storing of data on persons started relatively early, the skill to value and use statistical data started to develop and a register culture emerged.

Nowadays, inhabitants are asked a fairly large amount of data in every country. There are lists of the electorate and conscript, lists of persons with health insurance, schools register their students and graduates every year, the court system will register persons who have been given a punishment. In the past 50 years, many countries have knowingly started to design and create databases so that they would be as readily available as possible, meanwhile, the fields of use are broadening and private sector companies are starting to become interested in using the data.

The Nordic countries have been in the lead of the process of purposefully collecting, storing and reusing data. There, existing data are used as much as possible when producing statistics,

thereby reducing the amount of data collected in the interviews (Register-based ... 2007). In those countries, the registration of vital events in the church books started already centuries ago (Meetrikaraamatud 2015). Such church documents (parish registers) were also created in Estonia, when Estonia was part of the Kingdom of Sweden in the 17th century, but unfortunately the majority of them have not survived.

As a result of the development and widespread use of the calculating method, creating, keeping and using different databases became substantially easier. Nowadays it is difficult to imagine how somebody could use a database stored in books or punch cards.

What is a register?

A register is a database that meets certain requirements. The most important characteristic of a register is that in a register, all data (line items) are identified. Regarding persons, throughout history the usual identifier has been their name; however, things have always been problematic with that as well. Let us remember that Estonians only got their last names in the first half of the 19th century (Kurro and Kurro 2015), but a first name alone was not enough for identification purposes. And so, in church documents, a person's name was supplemented with the father's name, a married woman's name with the husband's name and often the inhabitants of a farm were named after the farm. These methods didn't suffice when identifying persons and because of that it is difficult to identify persons, including one's own ancestors, before the time when there were last names. With this it becomes clear that giving last names – which turned the country folk into individual persons – was a very important step in the historical development of the statistical recording of the Estonian population.

In reality, sometimes even the first and last names combined are not enough to identify a person. Here is where additional data, such as date of birth or place of residence, help. A personal identification code is a much more efficient identifier. In the past few decades, a personal identification code system has been implemented in many developed countries, including Estonia. As personal identification codes are different for every person living in a country and don't generally change during a person's life, the use of them will simplify the creation and use of registers containing everybody's personal data.

In addition to the fact that in a register, all line items must be identified, coverage is a logical requirement for registers, i.e. a register must contain line items for all objects who/what belong to the target population. Another aspect that is important about registers is the timeliness of their data: the register's data must be regularly updated to display changes in the situation of the persons (objects) included in the register. Therefore it is important to make a note about a person in the population register if the person leaves the country or dies.

In general, registers have been created in countries in the past 50 years, even more so in the present century, whereas the development of IT possibilities has remarkably sped up the development and use of registers.

Registers as the basis for statistics and censuses

In countries where an efficient register system was created early on, in the second half of the previous century, people started to use registers to produce statistics. For example right now, Statistics Sweden (2015) declares that the majority of the statistics produced in that office is based on registers. Let us not forget that we are talking about a country where theoretical statistics is also at a high level. As censuses are incredibly expensive statistical activities, the idea of moving towards register-based censuses emerged in precisely those countries.

The actual work on conducting a register-based census began in Denmark and Finland in 1970. The first fully register-based censuses took place in Denmark in 1981, in Finland in 1990 and in Sweden in 1995 (Lange 2014; Census ... 2008, 1990 Finland ... 2015, Tønder 2008, Register-based ... 2007). A census which relied heavily on registers took place already in the previous

century also in the Netherlands, where the data of registers were supplemented with data from surveys (The Dutch ... 2006, Nordholt and Linder 2007). The census of 2010/2011 was also register-based in Norway, Slovenia and Austria (Register-based ... 2008; Quality ... 2008; The Slovene ... 2009; Andersen and Utne 2007; The role ... 2006), but there were also non-European countries with a register system developed to such an extent that the census could be conducted register-based. One of those countries is, for example, Singapore (Singapore ... 2003).

Taking into account the resource-intensity of traditional censuses, including their very high cost (caused mostly by the activity of a large number of enumerators), quite a few countries have planned to use registered data as much as possible in the next census. Some countries are planning on switching to a register-based census already in the near future. Our Baltic neighbours, Latvia and Lithuania, are among those countries; Switzerland and even New Zealand and South Africa are considering a wider use of registers, including in censuses (The Swiss ... 2008; Bycroft 2011; Bah 2015). Estonia also plans to conduct the 2020 census based on registers.

The background of register-based censuses in Estonia

The history of register-based censuses in Estonia began with discussions over its feasibility already right after the 2000 population and housing census. The Population Register (PR), managed by AS Andmevara, commenced its official activity in 2000. Preliminarily, the register had been active since the first half of the 1990s, when data were collected and organised. Implementing a personal identification code, marked in identity documents and also used as an identifier in the Population Register, for all permanent residents of Estonia was an important step. Although normally every register has its own internal identifier, it is important that there is also a common identifier, which enables to link data of the same person in different registers, to ensure the cooperation potency of the registers. In Estonia, the personal identification code became such an identifier. Already back then there were optimists who were hoping that the 2000 population census would be the last conventional census in Estonia and in the future people would be enumerated using registers alone. Despite the big hopes of those optimists, the serious preparations for a register-based census did not begin in the first decade of the century, as it did in Austria and some other countries at the time.

The idea of a register-based census came up again in 2006 when talks began for the 2011 census. It appeared that although new registers had been created in Estonia in the meanwhile, time was not ripe yet for a register-based census. And so, in 2008, the Census Committee of the Government of the Republic took the decision that the 2011 census would be conducted using a combined method, in which registers would also be used, but the main work would still be done through interviews which would be conducted both as online questionnaires and face-to-face interviews.

Already in 2010 preparations began for a register-based census. Here, the experience of those countries who had already conducted a register-based census was taken into account. All of the representatives/head enumerators assured that one cycle of enumeration i.e. 10 years is the minimum time period during which it is possible to prepare for a register-based census. So a project was started to develop a methodology for a register-based population and housing census.

Why does a register-based census need its own methodology?

A register-based census must meet general census requirements

A register-based population and housing census must meet all the main requirements set for censuses: the census must be comprehensive, all data must be according to the census moment, the data gathered on all persons and objects must be homogeneous, smaller population groups must also be enumerated.

The output of the census must also meet internationally recognised requirements – that means a list of output characteristics specifically agreed upon, presented as 5–7-dimensional tables (hypercubes) by order of Eurostat.

The methodology of conventional censuses focuses firstly on the census questionnaire and secondly on the interview. The census questionnaire must be created in a way that the answers to it would provide all the required characteristics. The wording of the questions is important, it must be understandable for both the enumerators and enumerated persons and preferably also motivate people to answer. Organising a census requires substantial preparation, including designing the lists of enumerated persons and the census sheets, training enumerators, coordinating and supervising field work.

Main issues of the methodology of a register-based census

The main issues of the methodology of a register-based census differ greatly from those of a conventional census.

The first issue – and it defines whether a register-based census is even possible in a country – is **the existence and linkability of general registers**. Is there a population register that covers all of the population? Is there a housing register where all usable dwellings are listed? Both registers are only usable if the objects (persons) they contain are identifiable and inter-linkable. Although nowadays in many countries persons are identified with a personal identification code, things are not as clear with dwellings. In several countries (e.g. Norway), an important part of the preparation for a register-based census was identifying the dwellings located in apartment buildings: in cities, apartments were not numbered. Let us remember that in a population and housing census there are several questions regarding the quality of dwellings; however, there also questions regarding persons' living conditions, which require that the data of persons and dwellings are linked. In addition to that there is a variety of questions regarding people's employment – this information is collected by totally different registers, but the link of this information with persons' data is necessary to create census characteristics. Therefore, the first main question of a register-based census is whether in the country there exists a system of cooperating registers.

The next issue is whether it is possible to get the necessary information for **all necessary or obligatory census characteristics** from the registers that exist in the country. Here, the completeness of data is very important, including that there aren't any groups of enumerable objects or persons that might not be covered. As a back-up plan, some census characteristics could be created from research data. This is what some other countries have done (The Dutch ... 2006), but it is more of an emergency solution, as in the case of data derived from research, one respondent often represents several hundred or even thousand persons, so it is impossible to get information on smaller groups of persons, which is what census requirements presume. Even upon creating the output of the census – high-dimensional hypercubes – based on research data, there could be problems with the coherence between marginal distributions (Nordholt and Linder 2007).

The third issue is **the balance between the census and register characteristics**. Registers are not created for censuses, they have their own statute-based objectives for which data is

collected. Even if in the register there is data for essentially the same indicator that is necessary for a census, the definition of a register characteristic differs from that of a census characteristic and in order to make a census characteristic out of the register characteristic, the register characteristic must be converted. Often, the characteristics of one register are not enough to create a census characteristic; it requires the combination of the characteristics of several registers. The most typical example of this is the census characteristic 'social status', which is put together from the data of about 10 different registers.

The fourth issue is connected to **the quality of the registers**. On the one hand, this means data quality in the most general sense, the data meeting the standards that the register itself has set, the lack of systematic and random errors and the completeness of data, i.e. the lack of any gaps. A very important aspect of data quality is the regular updating/completing of data. This stems from one of the main requirements of censuses – acquiring the data at the moment of census. If registers update their data at the turn of the year, then, in case of a register-based census, it would be appropriate to consider the turn of the year to be the census moment.

In the case of a register-based census, another issue arises that does not come up in a conventional census – **defining the target population of enumerated persons**. In a conventional census, the number of enumerated persons or objects will be the census population and for a long time this census population has been considered to be the target population, i.e. it has been presumed that all of the persons or objects who were subject to enumeration, were enumerated. Only in the past few decades, people have been directing their attention to the fact that the census population might not coincide with the target population, it could be under- or over-covered. With a register-based census, there is no automatically generated census population. When several registers are used, often their target populations do not fully match (although theoretically they should). Therefore we must study the reasons behind the discrepancies between the registers and as a result of this analysis we must form a census population. Of course, the aim is to create a census population which is as close as possible to the target population. This means that the number and list of the permanent inhabitants (residents) must be defined as precisely as possible, i.e. an analysis called a **residency analysis** must be carried out.

Is the methodology of a register-based census universal for countries?

Since right now many countries are making an effort to switch to a register-based population and housing census, it is quite natural to ask whether it is useful for every country to work at it alone or whether for example Eurostat, the statistical office of the EU, could design a common methodology that at least all EU Member States could use. An even more ambitious recommendation would be for the UN to design such a common methodology for all future censuses.

However, such a plan is unrealistic. The main reason is that the registers in different countries have developed according to the specific country's needs, possibilities and traditions. The countries that had been among the first at creating and organising registers for the purposes of producing statistics, now have the most advantages. In the 70s-80s of the previous century, data protection limitations were relatively weak or even inexistent and in order to create registers, countries could use census data, which ensured a relatively good quality of register data at the time. Nowadays, the use of census data in administrative registers is generally unthinkable, but it would be a possibility to create statistical registers in statistical offices in which case the movement of data is unidirectional, i.e. the data stored in such registers are not given out to any institution or person on principle.

Therefore, every country must develop the methodology for a register-based census by themselves, although exchanging information with other countries who are solving a similar problem or who have already solved it.

System of registers

The system of registers is different in every country that is preparing for a register-based census, but it does certainly exist in some form. Of course, general registers are not enough to conduct a register-based census but they do make it possible to create the backbone of the census. However, the backbone would generally contain less information than required in the outcomes of the census. It is necessary to also include other registers. The amount of them differs by country and depends on the existing registers. For example in Finland, 30 registers are used in the register-based census. In Estonia, the plan is to use 17 registers, but some more might be added to that number. In order to use a great number of registers together, they must meet certain criteria and be compatible. Essentially this means that the registers describing the same objects (e.g. persons or dwellings) must be identifiable with the same characteristic. In addition to that it is necessary that the registers describing different objects were interconnected with identifiers to create a system of registers. Figure 1 (p. 53) shows a simplified model of such a system.

In every general register, the line items are linked by an identifier – the personal identification code, the company code and the address. General registers are connected to a number of additional registers which characterise the objects of the general register. For example, a number of additional registers are linked to the register of persons (in Estonia, this is the Population Register). These additional registers contain information on the person's education, social status and employment. Companies are identified by a code, and if the register describing a person's employment contains the code of the company, the person is linked to the company. The address identifier characterises the locations of all homes and workplaces.

If the place of residence is marked for every person in the persons register, then the address identifier links the person to his or her place of residence. In exactly the same way, the identifier links a company to its location. These links are characterised by the bidirectional arrows shown in the figure (Figure 1, p. 53). These links ensure the cooperation of registers and that the data presented in the registers can be combined and used when producing register-based statistics, including when conducting a register-based census.

The prerequisites for a register-based census in modern Estonia

When we talk about Estonia's preparedness for a register-based population and housing census, we can find both positive and negative aspects.

General registers and their linkability

The requirements for this prerequisite have been relatively well met in Estonia. What is special about Estonia is that most registers have been created (or redesigned) in the past 10 years. In the registers of personal data, the identifier is the personal identification code, which is used for many different purposes in society (digital medical prescriptions, digital signature etc). The standard for addresses was developed in 2008. This includes spatial coordinates and therefore it will link the buildings to their physical location in the nature. Creating an address standard meant a lot of extra work. Addresses of buildings and dwellings valid until then had to be exchanged for new ones in essentially all of the registers. This has taken several years, it was not finished for the 2011 census and it still hasn't been fully finished. Despite everything, implementing an address standard was a very important step that made it possible to map all objects precisely. The register of offices and companies is also active and its line items are supplied with codes. Although some things are not yet perfect in this system (for example some people's residence data are erroneous for various reasons, and for some companies, only the address of the headquarters is marked and not the addresses of branch offices), the situation regarding this prerequisite is good in Estonia, by estimation even better than it was for some other countries when they started conducting a register-based census for the first time.

The availability of data necessary for obligatory census characteristics

*This is one of the most serious issues that the census team (Statistics Estonia) cannot solve on its own, but needs efficient assistance from registers and the ministries that manage them. Even so, at the moment there is not a census characteristic in Estonia for which there are no data whatsoever in the registers. However, there are characteristics for which data are incomplete or quite lacking. One such characteristic is **the profession**, for which there is a register, but so far, the register only has information on few persons. In some cases, especially in the case of earlier events, we can get help from the database of the 2011 census, i.e. the last one. For example, the education information system does not reflect the education level of older persons. Moreover, no register can show if a person lived abroad in the previous century but was not born there.*

For the purposes of a better coverage of census characteristics in the registers, the Estonian Population and Housing Census (PHC) team has requested amendments in the law (e.g. renewing statistical information in the Population Register in case of an obligatory change of documents), and has held talks with ministries (creating a register of care homes and their inhabitants) and representatives of databases. Regarding some characteristics, the talks have been fruitful, but some problems still remain unsolved, as can be witnessed in the previous paragraph.

Converting register characteristics into census characteristics. A statistical register

The questions where and when (administrative) register characteristics should be converted into census characteristics, who should be doing it and where the census characteristics would be stored have been solved differently in different countries and for different characteristics. In some countries, a statistical register has been created in the statistical office. The statistical register contains all or part of the census characteristics.

A statistical register differs from an administrative register by the lack of other objectives besides producing statistics. A statistical register uses as an input mainly the data collected by administrative registers, but does not follow the documentation principle of administrative registers to the full extent. Therefore, conversions of data are permitted and possible in a statistical register, including correcting incorrect data and imputing missing data. In the case of all of these conversions, the principles of statistical ethics and logic are followed. The statistical register is located in the statistical office and no information is made available there.

Converting register characteristics into census characteristics is a task which greatly varies in workload and complexity, depending on the characteristic at hand. Some characteristics, for example the legal marital status, can be transmitted from the register (in this case, the PR) almost immediately. At the same time there are also characteristics which require the use of several registers and relatively long and complicated algorithms to form. Particularly big problems which are different than in the case of conventional censuses are those relating to the characteristics of the family and household.

While in a conventional census it is possible to register the situation where there are more than one households living in one dwelling, it would be impossible in a register-based census, as no register records, who of the persons living in a common dwelling form a household and how many households are living in a dwelling. Therefore, in case of a register-based census one must assume that all persons who live in the same dwelling form a single household. With this the number of households in the population is inevitably smaller compared to the real number (based on what people have said). Luckily, multi-household dwellings are not very common in Estonia nowadays. Sometimes several generations live in one dwelling (more often in a family dwelling) and these generations have separate budgets and consider themselves to be in separate households. University students who have sublet an apartment or a house form each their own separate one-member households. However the difference between the sizes of

different households (dwelling-based and joint household based) is not very big. According to the data of the 2011 population census, multi-household dwellings formed approximately 5% of all dwellings.

A much more complicated task is creating a matrix of the relations between the members of a household and creating family nuclei on the basis of register data. Some relations – those which represent marriage and child-parent relations – can be found in the Population Register. Unregistered partnership, however, which could also be the basis of a family nucleus, is in many cases impossible to identify with only register data. Kairiin Kütt, who has written a Master's thesis on the topic (Kütt 2014), created an algorithm which defined persons as partners, based on their sex, age, marital status and familial relation. The algorithm gave a good result in identifying the partners among the persons living in the same dwelling. Identifying the partners was made possible by comparing the result of the algorithm with data from the previous census. When the partners have been identified, it is possible to define the family nuclei, the status of the persons in the family, the size of the family and the same indicators for households as well.

The Estonian PHC team is planning to use the logical solution of calculating census characteristics in Statistics Estonia directly by the census team and storing those characteristics as a statistical register.

The quality of registers in view of censuses

The possibilities for registers to assess the quality of their data are rather limited as errors in the content mainly become apparent when using several registers at the same time. This fact is also reflected by the saying that the older the register (i.e. the longer it has been used), the higher its quality. Of course it is possible for registers to detect and correct technical errors. It is also possible to perform logic checks which highlight register errors to register managers. Despite all of this, solving the quality problems of registers is very important from the point of view of a register-based census. Here, the main problem is not the occurrence of random errors but rather the data gaps, data outdated (i.e. untimely updating) and systematic errors which have occurred for various reasons.

Upon acquiring data from administrative registers, the PHC team of Statistics Estonia will check their quality. Compared to the work of register managers, finding errors is more thorough thanks to the PHC team's experience as well as handling a larger amount of data together. Here, it is important that Statistics Estonia must not report the discrepancies that were detected upon cross-checking of several registers to the registers, however, analysing the registers together might lay ground to some general quality assessments.

Still, improving the data quality of registers should not be handled as solely a problem of statistics and census. Register data must be of high quality first and foremost to fulfil the main tasks of the registers. For this purpose, the state of Estonia set up a project to improve the quality of register data. The Land Board has been particularly active in this activity; the situation has improved regarding many areas whose address data were previously of rather low quality.

Register data quality is checked by the PHC team with logic checks, which are implemented both upon importing data from registers and upon creating census characteristics.

Target population of a register-based census

Defining the target population for the register-based census is an important and novel task for Estonia. It would make sense to use the list of residents of Estonia according to the Population Register or the census population of the previous census, adjusted with the current population statistics, as census target population. However, the problem is that the populations differ by several percentage points and neither of them are 100% trustworthy.

The Estonian population is over-covered in the Population Register. This is caused by unregistered emigration. However, there are also some permanent residents in Estonia who are

not included in the PR; therefore, a slight under-coverage is added to the over-coverage. On the other hand, census population was under-covered and 2.3% of persons were added to it based on a statistical assessment. It remains unclear whether the persons who were not enumerated back then still exist in the population.

Why is it necessary to define the PHC population as precisely as possible?

Firstly, one of the aims of any census, including a register-based census, is to find the most precise population, which is also a prerequisite to knowing all the other indicators precisely.

In the case of a register-based census there is another reason – if the presumed target population (which is also the census population) is over-covered, it will possibly create a lot of pointless additional work in the form of finding, checking and imputing census characteristics for the persons who do not actually live in Estonia (as it can be assumed that in that case, registers often lack the information necessary to create census characteristics).

The problem of residency was solved in Estonia for the first time when assessing the 2011 population census under-coverage (Tiit 2012; Tiit et al. 2012). More than ten registers were used for this and all persons, who had not been enumerated, but who lived in Estonia according to the PR (approximately 60,000 persons) were analysed. Persons' activity in the registers in 2011 was registered and a statistical discriminant analysis problem was solved based on this data. It became evident that roughly half of the persons analysed were active in several registers and those persons were included in the permanent population. For the rest it was assumed that they had left Estonia. The census population that was supplemented person by person was taken as the basis for the population statistics calculations done and published in Statistics Estonia.

Solving the residency problem to define the target population of PHC

Since there is no reason to assume that unregistered border-crossing would have stopped, whereas those who have once left without registering don't even have the possibility to register arriving from abroad, it does not make sense to automatically take the census population, which is only supplemented based on the vital events occurred in between, as the PHC target population.

Taking this into account, the residency problem was yet again solved as part of the PHC preparations, however, this time a higher number of persons was analysed. The solution to this problem formed the content of Ethel Maasing's Master's thesis (2015), whereas a large part of the work consisted in the preliminary organising of data and preparing it for analysis. Parallel to Ethel Maasing, some Master's degree students (mostly from the mathematical statistics programme) participating in the statistical data processing seminar solved the same problem so as to compare the results and find the most suitable methodology.

As at 1.01.2015 there are 1,462,859 persons in the Estonian Population Register, but 1,363,615 are listed as permanent residents of Estonia. According to Statistics Estonia, 1,313,271 permanent residents were living in Estonia at that time. When solving the residency problem, we must find the reason behind the difference between two figures, which reflect fundamentally the same data but which differ by approximately 50,000 (nearly 4%) persons.

In order to specify the actual size of the population of Estonia in the beginning of 2015, all registers that agreed to cooperate and order data were used. Along with the sub-registers, there were 21 registers in total in addition to the PR. A person was considered to be included in the register if he or she had been active in the register in 2014 at least once. As activity in the registers heavily depended on the person's sex and age, the data was divided into subgroups. The problem was solved by using logistic regression analysis, linear regression analysis and discriminant analysis, but the attempts at applying cluster analysis did not achieve a positive result. The statistical methods mentioned require a so-called study group (objects with a known belonging). The selection included, on the one hand, persons who were residents of

Estonia according to the PR and who participated in the 2011 census, and on the other hand persons who were not residents according to the PR and who had not been enumerated either.

The Master's thesis of Ethel Maasing gave as a result a population which was a few percentage points lower than expected, but it showed that the methodology which had initially been developed to distinguish 60,000 objects (Tiit et al. 2012) can also be applied in the case of a dataset consisting of 1.5 million objects. The need to include more registers in the analysis also became apparent, as the existing ones do not cover all the sex and age groups evenly. In the course of work it also became evident that alongside the register managers who are cooperative and helpful and who supplied the necessary data in due time, there are also those who could not hold up their agreement with Statistics Estonia and respond to the requests of the author of the paper in due time, and because of that a part of potentially necessary data was left out of the study.

Residency index and current calculation of residency

In principle, the residency problem should be solved every year, but this is a relatively work-intensive task that involves additional tasks, e.g. defining study groups. Here it is characteristic that a large part of the persons analysed will maintain their status and the decisions made in the course of the residency analysis should ensure this reasoned stability.

A list of persons with maximum coverage (a maximum statistical register MR), which contains both all the persons included in the PR and also persons included in the other administrative registers, and also residents of all the previous years (starting from 2011 when such an analysis was first carried out).

It probably makes sense to use **the residency index** which characterises every person included in the MR register, obtaining a value between 0 and 1, where (just like in the process of differentiating) 1 equals residency status and 0 equals non-residency. For persons in whose case residency remains unclear, the value of the index is between 0 and 1, whereas the decision regarding residency (which has to be made every turn of the year) is made according to whether the value of the index is higher than the threshold c (the value of which is, for example, 0.5) or not.

Calculating the index should take place every year automatically according to the data received from registers, pursuant to the following rules:

All persons who belonged to the register MR in the year n (and who therefore have the index value calculated in the previous year), will have the following possibilities in the year $n+1$:

- They belong to the register as possible residents also in the year $n+1$ (group A);
- By the year $n+1$ they will have (officially) left the country (group B);
- They are dead so they are no longer included in the register MR.

The index value of all group A persons for the previous year will be automatically reduced by the factor b ($0 < b < 1$), the index of group B persons will be given the value 0 but they will remain in the register MR and their analysis will continue.

In the course of a year the indices of all the persons who are included in the register MR will be changed pursuant to the following rule:

- A new person who is born or who immigrates will be given the index value 1.
- Every person who takes an active step/causes an event in a register, will have the value a ($0 < a < 1$) added to their index value. If the person's index value is 1, no value will be added.

At the end of the year, all the persons of register MR whose index value is greater than c will be enumerated and considered to be residents.

The most suitable values for the parameters a, b, and c must be assessed experimentally. It is suitable to use values such as $a = 0.2$, $b = 0.8$ and $c = 0.5$ as original values. It is possible to specify the analysis by attaching a particular value a to each register, and this value would confirm that the object is included among the residents.

It would probably be useful to carry out a complete residency analysis among the objects under inspection (or only those, whose index by the end of the year is between 0 and 1) every once in a while, using all existing registers.

Is a register-based census worth all the effort?

This question crops up every now and then, especially when we think about how conventional population censuses were conducted in the past – with a short preparation time and at a low cost. Perhaps we should give up registers, laptops, the Internet and GPS, and once again send volunteers to go door to door with enumeration sheets. We could consider using the computer only when counting the results.

Unfortunately that would get us nowhere: it is not that easy to turn back time. It would certainly be difficult to find close to twenty thousand volunteers who would agree to become enumerators. It is much more difficult to find people and even more difficult to motivate them to cooperate than previously. The quality of such a census would probably be unacceptably poor.

Its low cost has been emphasised as one of the main benefits of a register-based census. This does not apply in the case of the first register-based census – this is, on the contrary, even more expensive than a conventional census, as it includes massive one-time tasks such as developing the methodology and IT solutions, and organising registers.

The situation will get better in the future as a register-based census does not involve the work of enumerators, which forms a majority of the census budget (even if a substantial part of the census is conducted over the Internet, as shown by Estonia's experience), and most of the expenses are related to the work with data. The experience of Finland and other Nordic Countries also demonstrates the remarkable decrease in the cost of a population census upon switching to a register-based methodology. In those countries, the price of the last population census per enumerated person was assessed as the lowest (Ruotsalainen 2011; Main results ... 2009). In the case of Estonia, the improvement of the results will probably not be as radical, as the ratio of the enumerators' work to data-related work is smaller in a country with a smaller population.

The biggest advantage of a register-based census is the possibility of conducting population censuses more often, even every year, in order to monitor and analyse changes in the population regarding employment, education and distribution, foresee potential dangers and take measures to reduce them on time. For a country like Estonia, which is small and has a decreasing population, this is extremely important.

It can be assumed that other statistical activities will follow the census – a well-functioning system of registers should attract other researchers to use it in their analyses. Without undermining the importance and usability of sample surveys, it must be noted that a register-based analysis is exhaustive; therefore, random errors do not occur and the level of precision is much higher compared to sample surveys.

It requires serious work to establish and further develop a methodology of a register-based census that would be suitable for Estonia and which, at every moment, would use the statistical material collected in Estonia and the statistical knowledge accumulated in the whole world to its fullest extent in order to solve statistical problems that are important for Estonia at the moment.