

ENUMERATORS' ACTIVITY AFTER THE CENSUS

Ene-Margit Tiit
Statistics Estonia

Mare Vähi
University of Tartu

The Population and Housing Census ended on 31 March. It was followed by several months of intensive data revision and on 31 May we found out that a total of 1,294,236 permanent residents were enumerated in Estonia. How was this result achieved? What did the enumerators do and what are they doing after the Census?

Data revision after the Census

Data revision after the Census has taken months or even a year after previous censuses as well. The more questions and the stricter quality requirements, the more time it generally takes. Despite the fact that these days the revision of data is computerised, it still has its limits – programmers cannot predict all the mistakes people make.

In case of historical censuses the questionnaires were filled out and verified according to a certain procedure – the enumerators filled the questionnaires out in pencil, and any necessary corrections were made by striking off, not erased; the head enumerator made changes in black ink and the last-level check was made in red ink. In most cases the corrections concerned logical errors that stood out when reviewing the questionnaires – such as striking differences in the age of parents and children, compliance of occupation with education and status, etc.

Today, when the data is collected by using computers and internet, a programme is used for such verification already during the collection of data by pointing out unlikely situations (does the 22-year-old respondent really have a Doctoral level degree? Is the age gap between the cohabiters really 35 years?). In cases where the response is clearly inaccurate (father is younger than his son), then the programme does not let the respondent to continue completing the questionnaire until the mistake is corrected. In these days the revision of primary data consists of three parts – person identification, address identification and coding. This is followed by identification of dwellings and households and attributing them to relevant persons.

Identification of Personal Questionnaires

Identification of Personal Questionnaires (persons) sounds suspicious – does this really give an opportunity to examine individual personal data? However, this is not the case. If all the respondents and interviewers worked perfectly, this stage could be left out entirely. Unfortunately, many of the personal questionnaires received did not have the personal identification code – approximately one tenth during the e-Census and even more than one third during the interview census. The organisers could not expect that many respondents would not tell their personal identification code during the interview census. There were even reports of persons, who did not tell their personal identification code, because they were afraid that someone could use it to take SMS loan!

Personal identification codes are necessary for two purposes. Firstly, the personal identification code is used to associate the data of the Census Questionnaire with the information that was not asked during the Census, but was retrieved from a register (such as studying); the code is also used for filling in the gaps in collected information with register data. Secondly, personal identification codes are required for removing duplicate questionnaires from aggregated information (i.e. when multiple questionnaires have been received with regard to one person).

One of the time-consuming activities during initial data revision process is appending the personal identification codes to the questionnaires. This is mostly performed automatically, by using the Population Register, and based on person's name and date of birth. If these data are presented correctly, the task is an easy one: on average there are ca 50 people born in Estonia each day, cases of people with the same given name and surname are rather rare (especially in case of Estonians, because we have quite a long list of usable given names and surnames). Should there be two or more people with exactly the same given name and surname, additional information is used, e.g. information about the place of residence.

Things are more complicated in case of incorrect source data. In case of the date of birth, there may be a mistake made in the day or the year. In case of names there are many possible mistakes. When writing the name down by ear, Mati may be mistaken for Matti, Avo for Aavo and Kristiina for Kristina, Anneli for Annely, furthermore – the member of the household responding to the enumerator may be unaware of the name form if they generally use a nickname instead. Further problems arise when transcribing the name that is originally in Slavic or other non-Latin alphabet into the format used by the Estonian Population Register, which is required for determining the personal identification code, e.g. surname versions such as Kholodny, Kholodniy, Holodnõi. This causes problems with identification that cannot always be automatically solved by the programme and that require help from operators, who can see certain additional information about the person if so required, in order to match it with relevant data from the Population Register. This work was mostly completed within a month after the Census.

However, nobody should worry about the operators viewing their personal data in the course of identification process. The operator can only see the part of the questionnaire required for his work, later on the questionnaires are analysed without name, and even the personal identification codes are converted (encrypted) so that they cannot be associated with particular person. However, encrypted codes make it possible to associate the person's data with the data from registers, which is also used in anonymous format.

Address identification

The situation is similar in case of addresses. These days a standard address format is used in Estonia, which should allow identification of every dwelling. This format was also used in the questionnaires. But the persons enumerated were also given an option to use so-called free-hand address. That is the reason behind parallel use of standard address "Harju county, Tallinn city, Kesklinn district, Endla Street, house no 15" and address "Tallinn, Endla 15" and various other formats. In order to ensure accurate representation of each dwelling in the database and to avoid the situation where corner house is shown as two different buildings, it is also necessary to revise the address data. Significant part of this work can be performed by using relevant programme, but more complex problems must be solved by the operators. As for addresses, this Census encountered more problems and greater workload than the previous censuses, because during e-Census (via Internet) people wrote down their address by themselves, resulting in much greater variety than before when the address was written down by the appropriately instructed enumerators. Another problem is the loss of farm names (building names), which are now indicated in the standard address, if the location of the dwelling has no street name. The task becomes even more complicated (but hopefully improves the accuracy of the result) by indicating the place of residence on the map, a solution that is used in Estonia for the first time, and – in terms of self-enumeration – is rather innovative on a world scale as well. Another innovation was that both each household and each person could indicate two places of residence – primary and secondary. This also increases the number of enumerated dwellings and corresponding workload in specifying the addresses.

Coding

The third task the operators have been focusing on during recent month, is coding. In simple words this means replacing textual responses in the questionnaires with numbers, so-called

code. This concerns first and foremost professions and fields of activity, as well as dialects, religions, etc. There is an international set of classifiers for professions and fields of activity, whereas these classifiers, containing thousands of detailed descriptions of sub-divisions, are universal for the entire world, comprising both potato and banana growers, school and nursery school teachers, reverends and prison guards, ministers and janitors.

Over the time these classifiers have undergone remarkable changes due to the emergence of new professions and areas of activity. There are also professions that are no longer known today, or that have assumed a new form – such as carter, which during the first census was a popular job among Estonian male residents of Tallinn. During all censuses there has always been a standard classification of professions and occupations and they have always been distributed among relevant classes. One of the most interesting conclusions that can be made on the basis of the data from earlier censuses is the frequency of and changes in various occupations in the society. In the times of the first censuses conducted in Estonia the so-called “white-collar” jobs were held predominantly by German residents (doctors, teachers, clerks, as well as e.g. clockmakers), military jobs were usually held by Russians and Estonians were engaged in field work, factory work, the majority of the jobs related to handicraft (shoemakers, bakers, tailors), but they also worked as servants. According to the data gained from the first five censuses the Estonians worked mostly in agriculture and fishery.

Processing of duplicates

The next step in processing the Census data is processing of duplicates. Some people have been enumerated repeatedly – for example, a Personal Questionnaire has been completed by a person's spouse, child and the person himself. It is impossible for one person to complete the questionnaire for several times, but there may have been several people, who completed the questionnaire on behalf of their family member. It was also possible that for some reason one questionnaire was completed for person during e-Census, and the other during interview census. This led to duplicate questionnaires, from which it was necessary to select the most accurate, so-called original questionnaire, and leave out the rest. Although our good colleague from the Computer Science Institute thought that using personal identification codes ensured that the number of unique forms was “just one click away”, this actually proved to be much more complicated.

During this Census the personal data were collected by using various questionnaires and documents. In addition to permanent resident questionnaires (the number of permanent residents constitutes the size of Estonian population) there were also temporary resident questionnaires, which were completed both for persons with permanent place of residence in Estonia, and for persons, who permanently live abroad, whereas the questionnaires were mostly filled out by the members of their household; it was also possible to fill them out by the person himself. Close relatives were asked to fill out the questionnaires of the persons, who had left Estonia. Some institutions (e.g. hospital for persons with profound disability or care home) submitted the lists of persons to be enumerated. There is also the list of personal identification codes of persons, who attempted to enumerate themselves abroad, but their enumeration was discontinued, because they indicated foreign country as their country of permanent residence. Some people, who were not enumerated, but informed about that immediately after the Census ended, were interviewed over the phone. For each such category it is clear, whether they are considered as permanent residents of Estonia or not. Thus the following are not considered permanent residents of Estonia: persons, who have already left Estonia (according to the statement of their relatives), persons permanently residing abroad (despite their wish to enumerate themselves) and persons, who reside in Estonia temporarily (3–12 months), but whose permanent residence is abroad.

Things become complicated when duplicate (and often contradictory) questionnaires or documents were generated for more than 100,000 persons. For example – a person attempted to enumerate himself abroad, the process was interrupted and he came to Estonia and completed the questionnaire of permanent resident here. Another rather typical situation: mother has completed the questionnaire of the person as permanent resident, sister confirms that the person

has gone abroad, and brother thinks that the person lives in Estonia as temporary resident. All these situations needed watertight guidelines to be worked out that comply with international rules and good census practice and help to decide, whether the person is permanent resident of Estonia or not. Things are further complicated due to the fact that the definition of permanent resident used in the Census differs from that used in daily life and in research. For example, many parents of university students find it difficult to accept that – for the Census purposes – their dependent child is considered as a separate household with permanent residence in the university city and the home of the parents is listed only as a second place of residence. Another seemingly unusual thing is the international agreement on granting the status of permanent resident to a person after one year – the person, who left home previous month, is not yet considered a foreigner.

The oldest and most important tradition to be reckoned with while processing the Census data is reliance on Census data. Generally, Census data are not compared to the data gained from other sources (e.g. registers) for the purpose of revision and specification. Other data sources are only used for supplementing the Census data if some of the indicators are missing from Census data or are clearly controversial (so-called grave error).

Understanding the questions and providing accurate responses was definitely the greatest problem in case of self-enumeration, which was used in Estonia as the primary method (e-Census) for the first time. This was probably selectively used during earlier censuses concluded in the Republic of Estonia (1922 and 1934), when the enumerators were allowed to bring questionnaires to the household, give instructions about filling them out and collect the questionnaires at the second visit. Unfortunately there are no reviews and analyses about the volume and results of the self-enumeration carried out at that time.

The publicity campaign preceding this present Census provided in-depth explanations on when to consider a person as permanent resident of Estonia and when as a foreigner, relevant explanation was also indicated in the questionnaires, and in the instructions given to the enumerators. Therefore there is no verification of whether the person, who has indicated Estonia as his or her permanent residence (or who has been indicated as Estonian citizen by the members of his or her household), really lives in Estonia. One reason for lacking verification is that according to the international rules even the persons staying in the country illegally and not entered in any register are also subject to enumeration. However, it is necessary to analyse the situations, where the information gained during the Census seems to be contradicting.

If the duplicates have been filled out by different people, then the questionnaire completed by the person or a parent of a child less than 15 years of age is considered as primary (original) questionnaire. But if some of the data is lacking in the original, it will be supplemented on the basis of duplicates, while verifying the compliance of the added data with the data of the original. After selecting the originals and removing excess duplicates it is possible to state the number of people enumerated and people, who have left the country as revealed from Census data. As a result of this work, the preliminary number of the enumerated persons has been found out. However, in case of people enumerated and people who have left the country, one has to consider that this is information collected during the Census, which is not absolutely comprehensive and accurate.

Over-coverage and under-coverage of the Census data

In case of previous censuses, the number of people enumerated has been used as a reference to the size of population. This is actually not quite true. There have always been cases, where some people have not been enumerated for various reasons, leading to the census result that is smaller than actual population. Such a situation is referred to as under-coverage and it stands for the ratio of persons, who were not enumerated and population (usually indicated as percentage). Enumeration may also encounter another error – too many people are enumerated. This represents over-coverage. Over-coverage stands for the ratio of excess enumeration of people and population. One of the reasons for over-coverage may be enumeration of temporary

residents as permanent residents, but the most frequent is repeated enumeration. Historically speaking, the people have been rather settled, and thus under-coverage and over-coverage have been relatively low and more or less balanced. Census data was the most accurate estimation of the actual size of population available, yet it was possible to compare the Census data with the size of population gained from previous census, which had been supplemented on the basis of current population statistics. Therefore it was justified to use Census results as a reference to the size of population and it is still used for that purpose in many countries. The official population published by Statistics Estonia and the population with migration are based on the data of the last Population Census, which are amended on an annual basis according to the number of births and deaths. In case of population with migration both immigration and emigration are taken into account.

Today there have been changes in organisation of a census in the developed countries and quality requirements have become stricter. Over-coverage can be almost completely avoided by using personal identification codes, but under-coverage poses a serious problem worldwide. People are mobile, they have several places of work and residence and the enumerators have a hard time getting hold of them. Besides that certain people are keen on withholding their personal data and consider any publication thereof a violation of privacy. There are various other reasons why certain amount of people is left out of the enumeration process. Hence the understanding that today the number of people enumerated does not match exactly to the actual population and main census-related issue is under-coverage.

After the Census of 2000 a follow-up census was performed in order to assess census quality, mainly in terms of coverage. According to the calculations made by Ebu Tamm the Census had a under-coverage of 1.2%, but the media implied even greater under-coverage. However, the Estonian population was not adjusted according to census under-coverage, as there was no relevant methodology and necessary additional information concerning the distribution of error by sex/age and geographical distribution.

Under-coverage of the 2011 Population and Housing Census

After the Census conducted in Estonia there have been reports of people, who were not enumerated. Here are two typical examples that even reached the media.

“The enumerator did not want to enumerate me and said that all the people in my apartment have already been enumerated. My daughter is registered in my apartment, because I want her to inherit my apartment. She actually lives elsewhere, but when completing the questionnaire online, she stated that she lived here.” In case of Estonia, this is a typical situation that was anticipated by the organisers of the Census, but it has virtually no cure. If people enumerated themselves on the Internet and indicated incorrect address (most frequently their registered address) and that all the persons and households in that dwelling were enumerated, then the enumerators did not visit that dwelling according to the previous agreement and public information. The people actually living in that dwelling, who did not enumerate themselves during e-Census, might have been left out of the enumeration. In case of this example, confusion was caused by the daughter of the gentleman, firstly by wrongful indication of her own place of residence (ignoring the requirement to indicate actual dwelling), and secondly by completing father’s questionnaire about his dwelling and not indicating that her father also lived in that apartment.

The second example concerns Supilinna district in Tartu. A reporter of Eesti Raadio, who was asking about the reliability of the Census, randomly interviewed some men at the shop and found several people, who claimed that they had not been enumerated. For instance, one interviewee told the reporter the following: *“I have not seen the enumerator. The enumerator has come to my house, but I live on the second floor and keep my doors locked. There was a note left in the mailbox, but I found it only later.”*

In either of the cases there is no reason to blame the organisation of the Census or the activity or inactivity of the enumerators, this is more likely a case of unexpected behaviour of the people enumerated.

There is inevitably a considerable number of people in Estonia, who do not check their mail, keep their door locked, do not use a phone and therefore have been left out of enumeration process, despite being permanent residents of Estonia. Enumeration team proposed several other options to “catch” them – inviting enumerator over by phone, or registering themselves on the homepage after the Census was over. That is how the team found out about another few thousands of people, who were not enumerated (Map 1, p. 106).

Map 1 shows that the distribution of people, who were not enumerated at first, but who informed about themselves, corresponds roughly to the population density, being greater in major cities and smaller in under-populated areas. This indicates that it was a random process. However, the enumerators did not receive the information about all people, who were not enumerated – this group includes people, who were left out because they avoided the Census intentionally, or were just passive or disloyal. The Census data are presumably subject to at least 1% under-coverage, i.e. the amount of people, who were left out of enumeration despite all efforts. In international comparison this is not a bad result, but it definitely does not satisfy all those people, who want to know the exact size of Estonian population and not just as a single number, but also its distribution by sex/age/ethnic nationality and the exact geographical distribution.

Census data on persons, who have left Estonia

The data on persons, who have left Estonia, collected during the Census, gives further information on the actual size of population. Unfortunately this information is not quite complete, because if an entire household has left, there is no one to fill out the questionnaire about the persons, who have left. Therefore one can assume that the aggregate information on people, who have left Estonia, is subject to even greater under-coverage in comparison to the aggregate information on the Estonian residents, who were enumerated. At the same time this aggregated information needs further adjustment, because those indicated as people, who have left, may include people that – pursuant to international rules – were permanent residents of Estonia at the Census moment as they left only a few months before the Census moment or even after that moment. On the other hand, there are many duplicates in the aggregated information on those, who have left, because several close relatives (parents, siblings, spouse, or children) might have indicated the person as a person, who is no longer resident of Estonia.

The Population Register and actual size of Estonian population

The next work stage for the PHC team is the assessment of actual size of Estonian population. For that purpose the existing registers in Estonia are applied, whereas encrypted codes are used for associating various data in order to avoid the identification of persons.

Although the majority of Estonian registers have been created and launched within recent decade, i.e. they are rather new and their quality does not allow conducting the Census 2011 entirely on the basis of registers, they still contain plenty of information that can be used for assessing the size of Estonian population.

The Population Register. *One of the oldest and most accurate registers in Estonia is the Population Register that should, in principle, reflect all Estonian residents and their basic information, the most significant of which are personal identification code and place of residence, including the country of residence. The Population Register contains more than 83,000 people with Estonian personal identification code, who reside here temporarily or who have left abroad. Unfortunately, the exact address of people living in Estonia represents the weakness of the Population Register – because for many years it was not mandatory to register the place of residence in Estonia, and it is still only formally mandatory (it is generally not checked). There are several reasons for concealment or improper registration of the actual place of residence – benefits provided by local governments, school preferences, etc. Therefore the place of residence indicated in the Population Register differs from the actual place of residence in about one fifth of the cases. Sadly this includes cases where a person has left Estonia, but is still registered as Estonian resident in the register, which causes over-coverage in the Population*

Register, because immigrants are assumedly entered in the register in a more accurate way (this is confirmed by international practice). This is not meant as a reproof to the register, but rather to the legislation. This is where Estonian residents show rather weak registration culture and loyalty, which falls behind that of Nordic Countries that set an example for us.

When comparing (based on encrypted codes) the list of persons residing in Estonia at the Census moment according to the data of the Population Register with the list of persons enumerated, it is possible to determine the number of persons not enumerated. The hypothesis of over-coverage in the Population Register, which was supported by previous surveys, was also proved at the 2011 Census: the number of people indicated in the Population Register as Estonian residents (1,365,552 persons at the Census moment) is higher than the preliminary number of persons enumerated as permanent residents (1,294,236). In case of people entered in the Population Register, but not enumerated, there are two options – either they live in Estonia and were not enumerated, or they have left Estonia without registering their departure. A decision has to be made about each person in that list concerning the category they belong to – whether they are permanent residents of Estonia or they have left Estonia. That is when it is necessary to use other registers.

Other Estonian registers and their use in determining actual size of population

*Sources of further information are required for people, who were not enumerated as permanent residents or people, who had left. For that purpose various other Estonian registers can be used, (Figure 1, p. 108). Some of the major registers include the **Health Insurance Database**, containing more than 90% of Estonian people, the **Register of Taxable Persons**, covering significant part of working-age population, the **Estonian Education Information System (EHIS)**, containing comprehensive information about students, thus providing an excellent source for identification of persons of school age. Quite a lot of additional information is available from the **State Pension Insurance Register** (contains pensions, family benefits and allowances), the **Data Register of Social Services and Support (STAR)**, as well as the **Traffic Register**, containing the information on both vehicle users and owners. Considering the connection of a parent and a child (children) with parental benefit and family benefit, they have both been deemed as registered.*

In view of the persons, who have actively used all of the above-mentioned registers in 2011, it appears that, on average, each Estonian resident is entered in three more registers (in addition to the Population Register) during a year (Figure 2, p. 109). Yet there are also persons, who have been registered as Estonian resident in the Population Register, but have not been indicated in any other registers during the year. It would be only logical to assume that such persons have left Estonia. However, this decision cannot be rushed, it is necessary to analyse all registers and the extent to which the registers actually cover the population. The relevant analysis is provided below.

Persons whose activities are shown in several other registers besides the Population Register, are most likely to live in Estonia and if they are not enumerated, this may be due to enumerating someone else at their address (if a fifth of all people with different actual and registered place of residence indicated their registered dwelling as their place of residence during e-Census, this may account for not enumerating approximately 1% of people). Another reason for not being enumerated may be a temporary stay abroad or just evasion. The decision about whether the person resides in Estonia but was not enumerated, cannot generally be made on the basis of just one register, because there are people, who use certain social services in Estonia, while living abroad. Figure 2 (p.109) reveals that register coverage depends to significant extent on the age of the person and the group least covered by registers consists of people aged 20 to 60, especially men (Figure 3, p. 109).

Comparison of Census data with the register data

In early March after enumerating more than one million people, including both permanent residents and those, who had left Estonia, a comparison (based on 2011 register data) was performed with regard to the activity of two categories of enumerated persons – permanent residents and those, who had left Estonia (Figure 4, p. 110).

When comparing the data shown in Figures 1 and 4, it appears that there are relatively more active people among enumerated permanent residents than among the Estonian residents listed in the Population Register – both in terms of tax liability, use of the Estonian Education Information System, the Health Insurance Fund, motor third party liability insurance and parental benefit and family benefit. When comparing the data provided by the Population Register about people living abroad and the data gained from PHC about the activity of people, who have left Estonia (register data), the latter was remarkably higher in case of most registers.

That aggregated information was used as a basis for developing initial rules for dividing the persons not enumerated, but entered in the Population Register into two categories – permanent residents of Estonia and people, who have left Estonia unofficially. However, this is not an easy task, because the information available in the registers is contradictory at times, partially due to intentional behaviour of people. However, it revealed a series of reliable combinations of register entries that allow to decide, at great probability (minimum risk of error), whether a person really lives in Estonia or not (Table 1, p. 111).

But it is significantly more difficult to convincingly prove that a person does not live in Estonia. There are more than 2% of people among the Estonian residents enumerated as permanent residents, who are not entered in any of the registers, plus approximately 4% of people, who are only included in one register. It is not possible to follow the rule that classifies all such people as foreigners. It appears that according to the list of Estonian residents at the Population Register there are quite many people who are not entered in any of the other registers, and also people entered in just one register, whereas the majority of them are enumerated as permanent residents and there is no reason to suspect that they do not live in Estonia. Thus it is necessary to find additional sources of information to improve the assessment of non-enumerated Estonian population; this work is continuously in process.

Assessment of Estonian population and net migration

Estonian population is determined by adding the number of enumerated permanent residents to the number of persons, who were not enumerated, but who are very likely to permanently reside in Estonia as indicated by other information (effective registers). Such calculation of population also allows assessing PHC2011 coverage.

It is of great interest to assess the number of people emigrating from Estonia during the period between the censuses. Registered migration balance is derived from the difference in official population without migration and population with migration, which is 21,653 persons. The assessment of unofficial (unregistered) migration consists of two components. One of them is based on the information collected during Census about the people, who have left, while taking into consideration only those currently residing in Estonia according to the data of the Population Register. According to the data of the Population Register, the second component of unofficial migration is composed of non-enumerated persons residing in Estonia, who, according to the analysis of registers, were classified as persons, who have left Estonia. When comparing the data of the Population Register with the enumeration results, we can also assess the unofficial immigration, which means that in addition to the registered migration balance we can also assess unregistered migration balance.

Inevitably, there is still room for error. Perhaps, despite everything, the list of Estonian permanent residents does not include a man from Supilinn, who lives off collecting bottles and does not need social assistance, who has not yet attained pensionable age, has managed to avoid visiting a doctor within the last year, and has thus left no trace in any of the registers during the last year. The enumeration team makes their best efforts to give the most accurate assessment of the number of residents by utilising available information. From the statistics viewpoint the best result is not “the largest possible”, but the “most accurate” one.