

Creation of the income variable from administrative data sources

Urve Kask
(Statistics Estonia)

The transition to collecting data from registers entails both positive and negative side effects. The positive ones are considered to be a smaller response burden, a shorter interview time, more accurate data, less missing answers and lower costs. On the other hand, however, (at least in the first few years) the time spent on and problems with processing data and making them compatible with registers will increase.

Statistics Estonia has transitioned to using administrative data to create the income variable in three stages:

- 1) The first analysis was done in 2005 and it mostly focused on mapping the data of the Estonian Tax and Customs Board and on the option of using this dataset for creating the components of the income variable;
- 2) In the second stage in 2012, all registers were analysed regarding the possibilities for obtaining the necessary components to create the income variable. <http://www.stat.ee/metoodika>
- 3) In the 2013 survey on the income of 2012, no such questions were asked which regarded variables that could be created from registers and which proved to be very compatible in the analysis of the survey and register data: old-age pension, pension for incapacity for work, survivor's pension, unemployment insurance benefit, benefits related to family and children.

1. The main results which is based on the 2005 analysis.

A comparison of the 2003–2005 data of the Estonian Tax and Customs Board with the EU-SILC data: main conclusions. The analysis included the following income components: monetary income from waged labor; contributions to private insurance plans; old-age pension; income from rent; regular payments to other households.

- 1) The database of the Estonian Tax and Customs Board **also provides information on those whose income data was not included in the survey** (helps to improve data quality in the data organisation process)
- 2) If we consider the income of only such persons/households the data of which is in both the survey and the register, **then there are generally no statistical differences in the data**. In several cases, the differences in the indicators of the entire sample are also not statistically significant.
- 3) **In the survey, the number of salaried employees is smaller and the average wage level was higher**. The reason for this is the higher share of non-response among persons with lower wages.
- 4) Although the average magnitude of the income from pension recorded in the survey and the register do not differ significantly, **the discrepancy in the number of respondents is extremely large** and can have an impact on the total income of the survey – the Estonian Tax and Customs Board data are not a suitable source for data on old-age pension. The reason that the vast majority of pensioners do not have to pay taxes.

2. The options for gathering income data from various registers and the main results of the analysis in 2012

The first aim was to map all various registers from where it was possible to gain information on income (some component of the income variable). This compatibility analysis was based on three different register-based databases. With all three of them, Statistics Estonia has binding contracts, which guarantee the timely submission of data and the adequacy of individualised data. Each household and person participating in the survey has been assigned a separate code, which remains the same for all four survey years when the specific person or household is in the sample. This figure is also added to register-based data, which enables both data analysis and substitution.

The coverage of income components by data source:

1	Estonian Tax and Customs Board (ETCB):
1.1.	Tax from income
1.2.	Income from the employer
1.2.1.	Income tax refunds and additional income
1.3.	Unemployment insurance payment
1.4.	Redundancy payment
1.5.	Old-age pension
1.6.	Parental benefit
2	Social Insurance Board (SIB):
2.1.	Pension for incapacity for work
2.2.	Survivor's pension
2.3.	Child benefit
2.4.	Old-age pension
2.5.	Parental benefit
3	Estonian Health Insurance Fund (EHIF)
3.1.	Sickness benefit
3.2.	Maternity benefit

Old-age pension and maternity leave – there are two different data sources. Based on the analysis, it was concluded, what kind the data source will be used.

The more important results of the analysis and decisions based on: In 2009, 13,479 persons were interviewed for the Estonian Social Survey, and in 2010 – 13,426 persons.

Main conclusions:

- 1) **The register and survey data regarding old-age pension are very compatible** and the questions regarding **old-age pension can be replaced with the data of the Estonian National Social Insurance Board (SIB)**. To adopt the old-age pension data from the Estonian National Social Insurance Board to **expect the total number of pension recipients and a slight increase in the total sum and** to expect a slight decrease in average values;
- 2) The data from the Estonian National Social Insurance Board can also be used for both the pension for incapacity for work and the survivor's pension, but **there are certain differences in the number of people receiving pensions and in average values**. To use register data from the Estonian National Social Insurance Board **in determining pensions for incapacity for work to expect a slight decrease in the average values, higher total numbers of pension recipients and an increase in the total sums paid**. To replace the question about **the survivor's pension** in the personal questionnaire of the Social Survey with the register-based data from the Estonian National Social

Insurance Board; **to expect a slight increase in the number of pension recipients and a slight decrease in the average values.**

- 3) The question regarding unemployment insurance benefit can be replaced with data from the Estonian Tax and Customs Board (ETCB). Information regarding additional income tax and income tax refunds can also be replaced. Due to the tendency that people do not remember or indicate the relatively low sums of the received unemployment benefit, the register-based data from the Estonian Tax and Customs Board should be adopted in order to collect more accurate and reliable data. In that case, **the data will indicate a slight decrease in the average sums of unemployment benefit and a greater number of persons receiving the benefit.**
- 4) Among benefits related to family and children, the questions on most variables can be replaced with register data. To adopt the data of aggregated child benefits from the Estonian National Social Insurance Board in order to substitute it for the household question D11 sub-sections F, H, I, J, K, L, N, E, O, P. **A slight increase in average values and total sums is to be expected.**

EU-SILC	Type of benefit	Register
D11F	child benefit	SIB
D11D	pregnancy and maternity leave benefits	EHIF
D11H	childbirth allowance	SIB
D11B	parental benefit	SIB, ETCB
D11J	childcare allowance	SIB
D11K	conscript's child allowance	SIB
D11L	allowance for families with seven or more children	SIB
D11M	child benefit from local government	MISSING
D11N	adoption allowance	SIB
D11E	adoption benefit	SIB
D11O	foster care allowance	SIB
D11P	start in independent life allowance	SIB
DA3	state allowance	SIB

- 5) Information regarding maternity benefits is compatible with the register of the Health Insurance Fund. To collect data from the register of the Health Insurance Fund, **expecting a slight increase in the average values and total sums of maternity benefits.**
- 6) Parental benefit is compatible with the register data of the Estonian Tax and Customs Board and child benefits with the data of the Estonian National Social Insurance Board.
- 7) **There are no substitute sources for child benefits paid by local governments because these data are missing from the registers.**

It was decided that in the 2014 survey (regarding income data from 2013):

- 1) Questions regarding health-related benefits will not be replaced with register-based data from the Health Insurance Fund; in order to maintain the reliability and accuracy of the sickness benefit variable, the compatibility or combining of other databases with the dataset of the Health Insurance Fund should be analysed.
- 2) In order to guarantee the validity of the analysis, for the moment, the question regarding redundancy payment should be retained in the questionnaire of the Estonian Social Survey (ESS). Next year, the ESS questionnaire should be re-examined for compatibility with either the data of the Estonian Tax and Customs Board or the Estonian Unemployment Insurance Fund.

In the 2015 survey on the income of 2014, there will still be some income components. For example: in terms of monetary income, travel expenses are asked about; the question about net earnings has also been retained; additional questions are asked about non-monetary income from waged labour; questions regarding the following components were also retained: business income, property income, insurance, transfers from local governments, and also earnings, transfers and benefits received from a foreign country.