

# Estonian Health Interview Survey 2014

## Methodological Report

National Institute for Health Development

# **Estonian Health Interview Survey 2014. Methodological Report**

Mare Ruuge, Ardo Matsi, Leila Oja

Tallinn 2019

The **mission** of the National Institute for Health Development is to create and share knowledge for influencing the attitudes, behaviour, policies and the environment with evidence-based information with an aim of improving the well-being of the people in Estonia.

We would like to thank our colleagues from Statistics Estonia for their cooperation and contribution.

When using the information presented in this report, refer to the publication. Recommended reference:

Ruuge M, Matsi A, Oja L. Estonian Health Interview Survey 2014. Methodological Report. Tallinn: National Institute for Health Development; 2019.

# Contents

|  |    |
|--|----|
| Summary.....   | 4  |
| Definitions.....   | 6  |
| Abbreviations.....   | 9  |
| 1 Survey implementation.....   | 10 |
| 2 Methodology.....   | 11 |
| 2.1 Target population and sample.....                                      | 11 |
| 2.2 Questionnaire.....   | 13 |
| 3 Data collection.....   | 16 |
| 3.1 Fieldwork and data collection.....                                     | 16 |
| 3.2 Work of interviewers.....  | 18 |
| 4 Data processing and data set assembly.....                               | 19 |
| 4.1 Weight calculation.....  | 20 |
| 4.2 Results.....   | 22 |
| 4.3 Complementation of data set with data from administrative sources..... | 23 |
| 5 Data dissemination .....   | 26 |
| 6 Comparability of data from 2006 and 2014 health surveys.....             | 28 |
| References.....  | 30 |
| ANNEXES.....   | 31 |

# Summary

Health is mostly defined as the state of complete physical, mental and social well-being. Health status is influenced by age, chronic diseases and lasting health disorders, the causes of which accumulate over the life-span of a person and are related to the person's past course of life, including their health behaviour and living environment. In terms of health indicators, Estonia generally has some of the poorer ones compared to other European countries. According to statistics, Estonia has one of the lowest average life expectancies for men in Europe, greater part of the causes of death are due to chronic diseases as well as accidents and injuries.

Estonian Health Interview Survey 2014 (EHIS2014) was the third national interview survey concerning the health of Estonian people during the last 20 years.

The aim of this survey was to assess the health status of the Estonian population, the factors influencing health status and the use of health care services, and to see what changes have occurred compared to the results of the previous Estonian Health Interview Surveys (1996 and 2006) (1, 2).

The results of EHIS2014 are used to assess the social and health-dependent coping and need for assistance of the Estonian adult population, and to prognose the health status of the population. Another objective is to explain the changes that have occurred in the population's health status over time, taking into account the effect of health-affecting factors on health status, availability of medical care and the need for assistance as the various aspects of health behaviour. Analysis of the collected data set is used for designing health programmes for the population, planning preventive actions and evaluating the actions undertaken so far. A health survey is distinguished from other social surveys by the fact that different health topics are integrated into one survey and addressed more thoroughly.

Participation in the European Health Survey System enables EHIS2014 to get health indicators comparable with other European Union Member States. EHIS2014 was part of the second wave of the European Health Interview Surveys.

The obligation of conducting EHIS2014 resulted from regulation (EC) No 1338/2008 of the European Parliament and of the Council on Community statistics on public health and health and safety at work (3). Pursuant to this regulation, data is collected via health interview surveys regularly once every five years. The obligation and conditions (data required, reference year, methodology, sample size, micro- and metadata submitted to the Commission (Eurostat) and deadlines) for conducting the European Health Interview Survey were set out by the Commission Regulation (EU) No 141/2013 (4) of 19 February 2013 implementing Regulation (EC) No 1338/2008 of the European Parliament and of the Council. In order to achieve a high level of harmonisation of the indicators gathered with the European Health Interview Survey across the Member States, the Commission (Eurostat), in close cooperation with the Member States, published the "European Health Interview Survey Manual" (5). This manual included a model questionnaire, methodological and practical recommendations and guidelines on sampling and implementation of the survey.

The European Health Interview Survey is used to collect data, in order to obtain European Core Health Indicators (ECHI indicators) (6), which can only be gathered by conducting an interview survey among the population. Questions in the survey focus on collecting data on health status, health determinants and use of medical care services. This information represents a collection of minimal statistical data, which should enable better monitoring of the union's programmes on public health and policy areas on social involvement and social protection, health-related inequality and healthy aging. As the health and well-being of people has taken an important position in the comparison across countries, the health indicators collected with the health survey have filled a large data gap within the comparable health indicators of the European Community.

The first wave of European Health Interview Survey was conducted between 2006 and 2009 based on a voluntary agreement (gentlemen's agreement). This meant greater differences in the implementation of the survey, inconsistency in the used methodology and questions, and overall, less internationally-comparable indicators and a great dispersion as regards the year for conducting the survey. These deficiencies of the earlier wave of survey were significantly lessened by the implementing regulation of the Commission that was the basis for the 2014 survey, and the implementation of the methodological guidelines listed in said regulation and presented in the accompanying manual. The consistency of Estonia's participation in the European Health Survey System has been described in the Methodological Report of the Estonian Health Interview Survey 2006 (7).

In conclusion, the third Estonian Health Interview Survey is an abundant and versatile data set for studying the health of the Estonian population. As the used interviewing methods and questionnaire were consistent with other Member States, the data set of EHIS2014 provides indicators comparable with other European Union Member States. When comparing the coinciding data of the previous Estonian Health Interview Surveys in 1996 and 2006 and the current EHIS2014 survey, changes in the Estonian health indicators can be compared for a ten-year and eight-year interval, respectively.

# Definitions

|                              |   |
|------------------------------|---|
| <b>Active movement</b>       | intense physical activity such as jogging, ball games, swimming and skating, but also strenuous gardening   |
| <b>Alcohol unit (dose)</b>   | amount of an alcoholic beverage containing 10 g of absolute alcohol. 1 alcohol dose, for example, equals to 1 bottle of light beer (0.5 l), a glass of wine (100 g) or a shot of strong alcohol (30 g)  |
| <b>Body mass index (BMI)</b> | calculated by dividing body weight (kg) by height squared (meters). Pregnant women were asked to state their pre-pregnancy weight. The ranges of BMI values: <ul style="list-style-type: none"><li>• BMI &lt; 18.5 underweight</li><li>• BMI 18.5–24.9 normal weight</li><li>• BMI 25.0–29.9 overweight</li><li>• BMI ≥ 30 obesity</li></ul>  |
| <b>Depression</b>            | an emotional state mainly characterised by a decline in mood, decreased interest and joy of life, and loss of energy. The prevalence of depression was estimated using the Emotional State Questionnaire EEK-2. Based on the prevalence of depression symptoms (feeling of sadness, loss of interest, feeling of worthlessness, self-accusation, repeated thoughts of death or suicide, feeling lonely, hopelessness about the future, inability to feel happy) in the last four weeks, the respondents have been distributed into groups with significant depression and without significant depression. More information: Emotional State Questionnaire (EEK-2) (8)   |
| <b>Economic activity</b>     | everyone participating in the labour market or serving in the Defence Forces, as well as working old-age pensioners are considered economically <b>active</b> . Unemployed individuals, non-working pensioners, students, homemakers and persons receiving pension for incapacity for work are considered economically <b>non-active</b>  |
| <b>Education</b>             | evaluated based on the highest level of education completed by the time of the interview. Persons with a higher education have graduated from a higher education institution and their average study period is 15–16 years. According to the International Standard Classification of Education ( <i>ISCED</i> ) 2011, vocational education obtained after secondary education is considered equal to the first level of professional higher education and is presented under the section concerning higher education. People with secondary education have graduated high school, upper secondary school or some other educational institution providing secondary education and their average study period is 10–14 years. People who have not completed secondary education form the group with basic education or a lower level education (incl. persons without primary education) |

|  |   |
|--|---|
| <b>Emotional distress</b>              | condition characterised by several concurrent negative emotions affecting a person, especially stress, anxiety, decline in mood, often accompanied by asthenia (weakness) and sleeping disorders. An individual's emotional distress is assessed based on the Emotional State Questionnaire EEK-2, which was compiled according to the diagnostic criteria for depression and anxiety disorders presented in international classifications of mental disorders  |
| <b>Functional limitations</b>          | limitations on physical functions (such as hearing, sight, mobility) in everyday activities   |
| <b>Health behaviour</b>                | lifestyle of people, which is either <i>health-maintaining</i> and <i>enhancing</i> or <i>health-damaging</i> . Includes many areas, such as eating habits, smoking, alcohol consumption, physical activity, drug use, use of salt, eating fruit and vegetables, preventive visits to doctor, etc.  |
| <b>Household</b>                       | a household is composed of all the individuals living together and sharing, to some extent, a joint budget, which is usually the food budget. A household can also have only one member. Persons who are temporarily away (e.g. due to work, studies or conscription) have been considered as part of the household if they have retained economic bonds with their household. If a household member is temporarily (less than 4 months) in a nursing home, they belong to the household. <b>A private household</b> does not include persons who lived permanently in an institutional household at the time of the census. <b>An institutional household</b> consists of persons living in an institution operating throughout the year, twenty-four hours a day, and offers care for people living there – accommodation, food, and if necessary, also maintenance and care (such as a substitute home, nursing home). Ownership of the institution and sources of funding are not important |
| <b>Nationality</b>                     | depends on the person's self-determination and is not univocally related to their native language, but should comply with the nationality of one of their parents   |
| <b>Person in need of constant care</b> | a person who cannot cope independently with everyday activities, such as shopping, cooking or personal care, due to their health status, and therefore needs a caregiver. This does not include caring for small children   |
| <b>Place of residence</b>              | divided according to the statistical classification of Estonian regional units, NUTS 3 level  |
| <b>Regular smoking</b>                 | daily (practically daily) smoking that has lasted for at least one consecutive year   |
| <b>Self-perceived health</b>           | a general question for assessing one's own health status. The first question of the Minimum European Health Module: "How do you evaluate your health in general?"   |
| <b>Settlement type</b>                 | in the published 2006 survey, an urban settlement is an administrative unit with a population density of more than 100  |

people per km<sup>2</sup>, settlements with a lower population density have been considered as rural settlements

## Trauma

injury or poisoning caused by an accident or deliberate activity

## Abbreviations

|                 |   |
|-----------------|---|
| <b>CAPI</b>     | Computer Assisted Personal Interview                                |
| <b>CAWI</b>     | Computer Assisted Web Interview                                     |
| <b>EHIS2006</b> | Estonian Health Interview Survey 2006                               |
| <b>EHIS2014</b> | Estonian Health Interview Survey 2014                               |
| <b>EU</b>       | European Union  |
| <b>EU-SILC</b>  | Estonian Social Survey (Statistics on Income and Living Conditions) |
| <b>NIHD</b>     | National Institute for Health Development                           |
| <b>PHC</b>      | Population and Housing Census                                       |
| <b>PR</b>       | Population Register   |
| <b>SE</b>       | Statistics Estonia  |

# 1 Survey implementation

Consultations for the implementation of the survey between Ministry of Social Affairs, National Institute for Health Development and Statistics Estonia began at the end of 2011, the first official minuted meeting was held in January 2012. Conducting the survey as a cooperation project was designed to fulfil the obligation of Member States to keep consistency with prior surveys and use the existing resources sustainably.

Negotiations and coordination of positions for entering into a cooperation agreement lasted for half a year. In November 2013, a tripartite cooperation agreement was signed between the Ministry of Social Affairs, National Institute for Health Development (NIHD) and Statistics Estonia (SE) for conducting the Estonian Health Interview Survey 2014 in accordance with the instructional materials referred to in the regulation of the European Commission, and establishing the division of tasks and responsibilities. The cooperation agreement detailed the obligations of the parties and the more important deadlines of activities to ensure that the joint conduct of the survey is of high quality and on time. The obligations were generally divided as follows: NIHD was in charge of compiling the questionnaire and guidelines and carrying out data analysis, SE was in charge of data collection and sampling as well as submitting the micro- and metadata to Eurostat.

- NIHD prepared the questionnaires and guidelines for the survey in April 2013.
- SE conducted a pilot study in October 2013.
- Before fieldwork began, NIHD and SE organised a survey-centred training for the interviewers and interview managers.
- Fieldwork was conducted by SE from April to December 2014.
- Data clean-up, coding, weighing and other preparation for the further use of the data in SE lasted until April 2015. Work with the data file in SE continued until October 2015 when a microdata file was submitted to Eurostat.
- Three months after sending the initial microdata, SE forwarded the metadata (information on the data) as required to Eurostat for compiling a quality report. Work on the quality report lasted until December 2016.
- Validation of the data set with Eurostat lasted from February to August of 2016, when the first health indicators of the Member States collected in the second wave of the European Health Interview Survey were published on the databases of Eurostat and OECD.
- Despite the contractual obligation to forward the data set of the survey to NIHD so they could complete their tasks, SE gave NIHD only remote access to the data set of the EHIS2014 questionnaire at the end of March 2016. NIHD published the first selection of data from the survey in December the same year.

Cooperation went smoothly until the end of fieldwork for the survey. NIHD was notably hindered in fulfilling their obligation to publish data by two circumstances, which the parties, despite negotiations, were not able to resolve. Firstly, SE's work on the data set lasted longer than intended (incl. validating and specifying the data set with Eurostat, addressing the need for imputing questions on income, checking and amending linked data sets). The lack of a specific person whose responsibility would be the smooth conduct of the survey in all SE's contractual obligations had an effect on the cooperation. Secondly, SE withdrew from their contractual obligation to forward the survey's microdata set to NIHD for data processing. NIHD only received remote access for data processing as a research institution and not in the scope needed and requested for data analysis. For NIHD to be able to fulfil their contractual obligations, their only choice was to use remote access for data processing – a limitation NIHD had not taken into consideration when planning their activities. Thus, the publication of data to users was postponed for later than planned and happened on a significantly smaller scale.

## 2 Methodology

In addition to the requirements for conducting the European Health Interview Survey described in regulation No 141/2013 of the Commission, the EHIS2014 was conducted on the basis of the methodological guidelines developed in cooperation with the statistical office of the European Union, Eurostat, and the national representatives in the health interview survey working group:

- "European Health Interview Survey (EHIS wave 2) Methodological manual". Contains methodological and practical recommendations and guidelines for sampling and the implementation of the survey, but also a standard questionnaire and conceptual guidelines for the questions (5);
- rules for validating microdata and guidelines for submitting it to Eurostat;
- requirements for compiling the quality report to be submitted to Eurostat.

The quality report for the second wave of the European Health Interview Survey (9) was published on the website of Eurostat in 2018.

### 2.1 Target population and sample

EHIS2014 is a sample survey and the target population were assessed using the data collected based on the sample. The target population of the health survey included permanent residents of Estonia aged 15 and older, living in private households.

The sampling frame was formed on the basis of the individuals in the Population and Housing Census (PHC) 2011 database with the address of their residence at the time of PHC (30.12.2011); this data was amended according to the population register (PR) as of the middle of March 2014 as follows:

1. persons, whose year of arrival to Estonia was marked as 2012 or 2013 in the PR database, were added to the sample;
2. persons with an address in a foreign country in the PR base were excluded;
3. persons with a prison address in the PHC base were assigned the PR address;
4. deceased persons were removed.

Persons under 15 years of age and about 13,000 persons living in institutions (prisons, nursing homes, hospitals, substitute homes) were excluded. Last amendment was conducted in March 2014.

The population was divided into 4 strata based on their place of residence. For sampling, subjects within a stratum were first sorted by county and then by their personal identification code. Sampling occurred via systematic selection.

Population and sample sizes in the strata:

| STRATUM   | POPULATION SIZE  | SAMPLE SIZE  |
|---|------------------|--------------|
| Tallinn   | 337,091          | 2,500        |
| Harju (excl. Tallinn), Ida-Viru, Lääne-Viru, Pärnu, Tartu Counties        | 496,854          | 3,500        |
| Jõgeva, Järva, Lääne, Põlva, Rapla, Saare, Valga, Viljandi, Võru Counties | 241,224          | 2,100        |
| Hiiu County   | 7,228            | 400          |
| <b>TOTAL</b>  | <b>1,082,397</b> | <b>8,500</b> |

Address information was verified with the address database of SE's geodatabase eGEOstat. (The task of the geodatabase is to collect, manage, process, analyse and visualise SE's spatial data.) Persons in the sample were distributed between 132 interview regions.

Specifications in compiling the frame and sample were the following:

1. persons in the PHC base but not in the PR (excl. the deceased) were kept in the frame;
2. persons who have participated or are participating in the surveys of Statistics Estonia were kept in the frame;
3. only persons with an address of a sufficient level of precision were taken into the sample – in urban settlements, the address was known at least down to the house number.

The sample was distributed between seven periods or survey moments. The number of the survey moment indicates the number of the month when the object was interviewed. The interview rules for each survey moment had the following scheme:

1. web interview (method CAWI – Computer Assisted Web Interview) in the month (April to October) corresponding to the number of the survey moment,
2. personal interview conducted by the interviewer (method CAPI – Computer Assisted Personal Interview) starting from the month following the month corresponding to the number of the survey moment until the end of the survey period on 30 November 2014.

The interview schedule or rules (survey moments, methods and their duration) were the following:

| Survey moment | CAWI      |            | CAPI      |            |
|---------------|-----------|------------|-----------|------------|
|               | Start     | End        | Start     | End        |
| 4             | 1.04.2014 | 30.04.2014 | 1.05.2014 | 30.11.2014 |
| 5             | 1.05.2014 | 31.05.2014 | 1.06.2014 | 30.11.2014 |
| 6             | 1.06.2014 | 30.06.2014 | 1.07.2014 | 30.11.2014 |
| 7             | 1.07.2014 | 31.07.2014 | 1.08.2014 | 30.11.2014 |
| 8             | 1.08.2014 | 31.08.2014 | 1.09.2014 | 30.11.2014 |
| 9             | 1.09.2014 | 30.09.2014 | 1.10.2014 | 30.11.2014 |
| 10            | 1.10.2014 | 31.10.2014 | 1.11.2014 | 30.11.2014 |

Sample size was 8,500 persons. Based on the information collected in the survey, the persons in the sample were divided into three groups – respondents, non-responses and error in the frame:

| Sample distribution: ineligible sample cases, non-responses and respondents |  | Number |
|---|--|--------|
| <b>1</b>  | <b>Total released sample cases</b>   | 8,500  |
| <b>2</b>  | <b>Ineligible sample cases / out of scope units</b>  | 347    |
| <b>2.2</b>  | <b>Status changes</b> between sampling and interviewing (deceased, left Estonia, living in an institution, etc.) | 289    |
| <b>2.3</b>  | <b>Out of target units</b>   | 49     |
| <b>2.4</b>  | <b>Other ineligible (here: members of the same household)</b>  | 9      |
| <b>3</b>  | <b>Eligible sample cases / in-scope units</b><br>► [3] = [4] + [5]   | 8,153  |
| <b>4</b>  | <b>Total non-response</b>  | 2,701  |
| <b>4.1</b>  | No contact made  | 659    |
| <b>4.2</b>  | Refusals (contact made, but refused to participate)  | 1,618  |
| <b>4.3</b>  | Unable to participate  | 148    |
| <b>4.4</b>  | Not included (i.e. partially filled) questionnaires  | 68     |
| <b>4.5</b>  | Other reasons  | 208    |
| <b>5</b>  | <b>Respondents (filled questionnaires)</b>   | 5,452  |

After removing incomplete entries, the final data set of EHIS2014 included 5,452 persons. The required minimal sample size of the European Health Interview Survey for Estonia was 4,270 completed personal questionnaires.

Response rate for the survey based on the initial sample size was 64% (5,452/8,500). Taking the frame error into account, the response rate was 67% (5,452/8,153).

## 2.2 Questionnaire

An individual's health status can be evaluated based on the occurrence of diseases and health problems depending on how much these problems limit the individual's ability to cope on functional and social level. Regarding the limitations, it is important whether they are temporary or long-term, whether and how much assistance is needed and received by the individual to cope with daily activities, and if there are shortcomings. In addition to physical health, mental health needs to be taken into consideration as well. Health status is mainly affected by chronic diseases and long-term health disorders, the causes of which are related to the social environment, economic well-being and health behaviour. Health behaviour can in turn be either health-maintaining and enhancing or health-damaging. The need for using and the use of medical care differs from person to person depending on their health and socio-economic status. The health survey looks at all these health-related fields together and in an interconnected way.

The mandatory part of the questionnaire was made up by the variables listed in Annex I (4) of regulation (EU) No 141/2013 of the Commission to be submitted to the Commission (Eurostat), which contained among other things the necessary socio-economic background variables (education, employment, living conditions, household, income) and technical variables of the questionnaire not listed in the standard questionnaire. The mandatory part was composed of the following modules:

1. European Background Module (EBM);
2. European Health Status Module (EHSM);
3. European Health Care Module (EHCM);
4. European Health Determinants Module (EHDM).

The questionnaire looked at the various aspects of the respondent's health status, such as chronic diseases and injuries, mental health and emotional state, functional limitations. It also looked at the use of medical care, inpatient care, daily coping with household and personal care activities, as well as the use of medicines and health behaviour.

The reference period for health-related events was generally 12 months or the last four weeks. However, with certain modules, such as mental health or use of medicines, the reference period was 2 weeks. As different time periods were used, the reference period was always named and emphasised in the question.

Parts of the questionnaire and their order were the following:

- A. Household
- B. Health status
- C. Accidents and injuries
- D. Mental health
- E. Emotional state
- F. Physical and sensory functional limitations
- G. Personal care activities
- H. Household activities
- I. Preventive actions
- J. Sexual behaviour
- K. Education and work
- L. Place of residence
- M. Childhood home
- N. Social support
- O. Use of outpatient care
- P. Use of inpatient and day care
- Q. Use of medicines
- R. Smoking, alcohol consumption
- S. Drug use
- T. Weight, height, physical activity and eating habits
- U. Attitudes

Compilation of EHIS2014 questionnaire had two main sources: mandatory questions of the second wave of the European Health Interview Survey and the questionnaire of the previous Estonian Health Interview Survey (EHIS2006). The aim was to get data comparable with other European Union Member States within the second wave of European Health Interview Surveys as well as with the data of the previous EHIS2006. To achieve comparability, the questionnaire needed to be adjusted, and eventually it contained more questions. In certain cases, simply adding follow-up questions or amending possible answers was not enough. For example, in the case of non-use of mental health and medical care, it was necessary to double the questions asked on the same topic because the indicators necessary for the output were inherently too different for integrating. Thereafter, both groups needed to be found a suitable place in the questionnaire. Therefore, the two sources made the preparation of the questionnaire more difficult and this resulted in a longer questionnaire.

The preparation of the questionnaire and instructional material in Estonian was followed by their translation into Russian. The translation was conducted by an expert who also participated in the preparation of the questionnaires for the two previous health surveys. This ensured the quality of the translation with attention to relevance and linguistic specificities.

All the instrumentation necessary for the interviews was prepared in two languages, Estonian and Russian. As two interviewing methods were used, the web version (CAWI) and the computer assisted personal version (CAPI), the wording of the questions as well as the wording of the short guidelines added to the questionnaire for assisting the respondent were unified so that they could be used for both interview methods.

SE deployed a completely new software for EHIS2014 for both interviewing and data processing, therefore, the technical preparation of the questionnaire was very time-consuming and more time was planned for everything starting from the technical description of the questions and answer options to the screen displays and design of the questions. SE and NIHD repeatedly checked and tested whether the redirections in the programmes work properly. Not everything that was desired (e.g. asking questions in the form of tables, calculations during the interview) could be realised in the questionnaires for technical reasons and compromises had to be made.

As it was a computer assisted interview, automatic checks between different questions, but also the minimums and maximums of the answers, links, filters and other rules were added to the questionnaire (CAWI, CAPI) to ensure the quality of the data set. The validation rules given by Eurostat for the data set to be submitted could not be implemented directly but needed to be adapted according to the questions used in the questionnaire. Also, these validation rules were for the variables to be submitted and coded for Eurostat only.

Seven variables were adapted compared to the mandatory questions of the European Health Interview Survey. For the European Health Interview Survey, these variables could not be formed directly from the answers to the questions in the questionnaire, but had to be calculated based on a combination of several variables or answers.

Therefore, the Estonian survey included separate questions about all types of accidents and the respective need for medical care. The mandatory part was limited to three types – traffic accidents, home and leisure accidents. Concerning received medical care, it was required to ask about one of the most serious case of the named accidents. EHIS2014 asked about received medical help for all types of injuries separately.

In the Estonian questionnaire, coping with daily personal care activities (eating, getting out of bed, getting dressed, washing the entire body, using the toilet) and the need for assistance were asked based on separate activities, just like in the previous 2006 survey, and not about all the activities in total. The same applied to questions about housework (preparing food, shopping, lighter or heavier household activities, using the telephone, taking medicines, managing daily finances).

In the questionnaire, alcohol consumption was asked based on each type of alcohol, i.e. the percentage of alcohol content, (light alcohol, strong beer, wine, spirits). The consumed amounts of alcohol on the basis of the type of alcohol had to be later added up by converting them into the required alcohol units (alcohol doses).

Smaller adjustments were also made for many other topics. In the part about physical activity, concerning transportation form place to place, either on foot or by bicycle, individuals were asked to name the time spent in minutes instead of time periods. The answer options to the question about the last flue vaccination were adjusted in such a way that the results could be compared to the results of the 2006 survey but would also contain the mandatory answer options. For this, all individuals who had received flue vaccination up to two years ago were asked the month and year of their last vaccination. EHIS2014 questionnaire posed questions to individuals about visits to their family physician within the last 4 weeks based on type of visit – consultation, home visit, consultations via phone and e-mail. Use of prescription and over-the-counter medicines were asked about not only for the last 2 weeks but for several time periods.

The questionnaires of the Estonian Health Interview Survey 2014 can be found on the website of NIHD (<https://www.tai.ee/et/tegevused/teadustoo/loppenud-uuringud/eesti-terviseuuring-2014>).

## 3 Data collection

Statistics Estonia conducted fieldwork for the survey from April to December 2014. As the results of health-related interview surveys depend on when the interview is conducted, the fieldwork of the survey was distributed over as long of a time period as possible to ensure a uniform workload for the interviewers and to lower seasonal effects on the results of the survey. Data was collected using two interview methods: web interview (CAWI) and a computer assisted personal interview (CAPI). Interviews were conducted in Estonian and in Russian. The Estonian and Russian questionnaires of the health survey and the interviewer manuals together with additional information for the respondents were developed at NIHD as described above.

For preparing the main survey and minimising later measurement errors, a pilot survey was conducted in October 2013. The questionnaire could be filled out between 1 October and 7 October 2013 in the electronic data collection environment of Statistics Estonia, eSTAT. Those who did not fill in the questionnaire during that time period were visited by a trained interviewer from Statistics Estonia between 8 October and 31 October 2013. Sample size for the pilot survey was 160, of whom 115 were respondents, with 14 using the CAWI option and 101 CAPI.

After the pilot survey, focus group interviews took place in order to determine possible deficiencies. In the collaboration between SE and NIHD, the results of the pilot survey were summarised by the end of January 2014. Then began updating the questions and instructions in the electronic environment to ensure timely readiness for the fieldwork.

### 3.1 Fieldwork and data collection

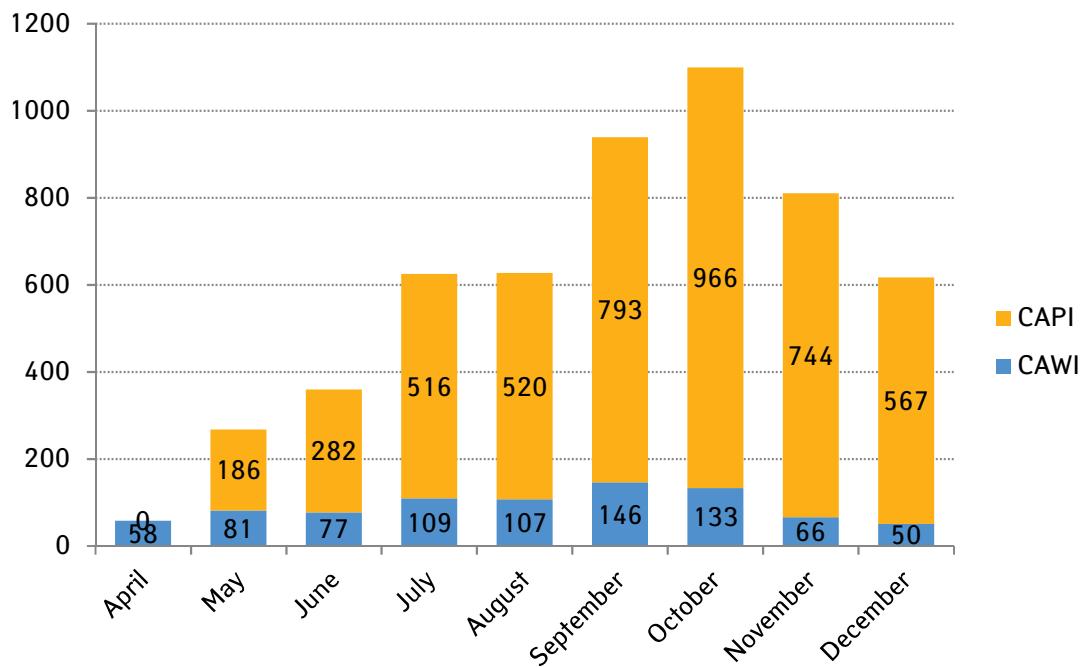
Before fieldwork began, NIHD and SE organised a survey-centred training for the interviewers and interview managers. NIHD was responsible for the part of the training concerning the content of the survey and SE handled technical issues.

Fieldwork for the survey began in April 2014 with the CAWI method, and a month later, in May 2014, CAPI began for those who had not answered in the CAWI. The sample of the survey was divided into seven parts and was sent out for answering in seven waves on a monthly basis, first as CAWI and then as CAPI.

| Month of survey | Sample size  |
|-----------------|--------------|
| 4               | 1,113        |
| 5               | 1,113        |
| 6               | 1,113        |
| 7               | 1,448        |
| 8               | 1,448        |
| 9               | 1,278        |
| 10              | 987          |
| <b>Total</b>    | <b>8,500</b> |

Fieldwork started and ended as planned and on time. Largest portion of answers was received in the fall of 2014, between September and November (Figure 1). Data inflow and correctness were continuously checked by SE. In total, 663 questionnaires (5,452 respondents or 12%) were filled out online, which was the amount expected based on the pilot survey. Compared to others, the web-questionnaire was filled out more by women, younger people, people with a higher education and

working people. Based on age groups, the largest difference was in the group aged 65 and older, but in the younger age group 15–25, there was no difference in the choice of interview method.



**Figure 1.** Receipt of filled in questionnaires of the Estonian Health Interview Survey 2014 from April to December 2014 by interview method

Respondents were contacted by post as well as by e-mail. As a reward, respondents received reflectors and keychains.

Interviews lasted on average 1 hour and 46 minutes with the CAWI method and 1 hour and 21 minutes with the CAPI method. The minimum length of a CAWI interview was 5 minutes, the minimum length of a CAPI interview also 5 minutes. However, there were only 6 interviews that lasted 5 to 10 minutes, 5 of these were online. The maximum length of CAWI was 16 hours and 53 minutes and CAPI 19 hours and 9 minutes. With questionnaires filled out online, in some cases, the slowness of filling the questionnaire (which was complained about) should be taken into account, also there is no information about periods of inactivity while the questionnaire was still open. In the case of CAPI, the length of the interview was affected by the completion of the interviewer section of the questionnaire, which marked the end time of the interview. Also, very short, one-minute questionnaires were created when the "Complete" button was not pressed on the questionnaire. Therefore, data on the length of the interview is not very accurate.

There was only one case, where another member of the household (final stage of a disease) answered the questionnaire, which was initially classified as answered by a member of the household (Proxy interview). Based on the person's data and comments, it was decided that this was a situation where the person was not able to fill the questionnaire on the computer but answered the questions themselves while a member of the household entered the answers on the computer. So there are no Proxy interviews in the data set. In 97 cases, it was noted that the respondent was aided – in these cases the respondent was very sick, could not see or hear or remember very well anymore. This also included young respondents whose mother helped answer questions concerning the household.

## 3.2 Work of interviewers

**Kristiina Saar, Head of Interviewer Network at Statistics Estonia**

In the health survey, the usual number of contacts via home visits for an interviewer was required – 3 in rural areas and 5 in cities before the respondent was classified under non-participation. The interviewer also left a note with their name in the mailbox every time, where they noted down the expected time of the next visit. On average, 4 contact attempts were made before the respondent was recognised as non-participating.

If the respondent had a telephone number, the interviewer could call it if they had not managed to contact the individual during a home visit. The interview could not be conducted over the telephone, the phone could only be used for setting up a visitation time.

A notification letter was sent to all individuals selected for the survey by post at the beginning of the survey. The letter introduced the objectives, content and organisation of the survey, and data protection rules. The letter contained the contact information of the interviewer and links to the web interview together with instructions for logging in.

With the health survey, reminders about the web interview (CAWI) were sent to those who had not completed the questionnaire yet and whose e-mail address SE had. These were sent out three times, once in September and twice in October. In 2014, sending reminders was not usual practice yet, which is the reason why reminders were not implemented from the beginning.

After the end of fieldwork, SE sent the respondents a feedback questionnaire with nine questions for evaluating the work of the interviewers. The health survey was the first time the feedback letter was sent as a link by e-mail, which is why more letters could be sent than normal. The feedback letter was sent to all respondents whose e-mail SE had, and a physical letter was sent to a selection of respondents whose e-mail address was not available.

522 letters were sent out by post for receiving feedback on the interview. 228 replies were received, i.e. 44%. 85% of the individuals who replied were satisfied with the work of interviewers, 8% somewhat satisfied, the remaining 7% had not answered this question. 1,751 letters were sent by e-mail and 461 or 26% received replies. The share of replies received by e-mail was smaller also because many of the e-mail addresses were invalid. 85% of the respondents were satisfied with the work of the interviewer, 13% were somewhat satisfied but 11 or 2.4% were not satisfied. In total, 689 answers were received to the feedback letters.

Quality control checked whether the interview was conducted according to the methodological requirements. It was discovered that 32 of the individuals who received the letter (6.7%) said that the interview was conducted by phone. The interviewers were asked about all these cases and some of the claims were confirmed, most of the times the respondents did not have time for a meeting or they were ill, but not always.

The respondents could add comments to the letter about the survey. Over 150 comments were added in total, mainly specifying answers (e.g. about the location of the interview if it took place in a cafe or elsewhere out of the home of the respondent) or assessing the work of the interviewer. The latter also included positive feedback. Negative comments were about the length of the questionnaire (too long) and about questions (too personal).

The interviewers felt uncomfortable asking questions related to sexual behaviour – they were difficult to ask, especially from older people, or on the contrary, from younger participants. With the computer interview, it was possible to offer the option for the respondent to fill in the answers themselves, but this is probably more suitable for younger people and less for the older people. Respondents complained the questionnaire was too long – the interview often took 1.5 hours and sometimes even more. Questions about childhood were sometimes difficult to remember – the

example that was mentioned was a question about the number of books at the respondent's childhood home. In the feedback letters, people complained overall that the questions were too personal to answer in front of the interviewer, although the option for answering online was not used very actively. The answer options for some questions were thought too rigid, incl. questions on religion did not have a suitable answer option.

Media coverage on the survey was not sufficient – this feedback was received from both the respondents and the interviewers. Wider coverage was probably expected because of the involvement of NIHD. Four health-related interview surveys involving NIHD began in spring of the same year and the notifications for different surveys might have caused more confusion for the respondents. SE sent a notification about the survey and later e-mails with reminders if the questionnaire was not filled online.

Most unanswered questions were about how many teeth the individual was missing, time of the last flue vaccination and blood sugar measurement; individuals were also reluctant to reveal their age during their first sexual intercourse, and could not assess the interest and involvement of others in the respondent's activities.

## 4 Data processing and data set assembly

The data set was cleaned, coded, weighed and otherwise prepared for the further use of the data by Statistics Estonia. The checking and cleaning of the health survey data began in June 2014 after the receipt of the first questionnaires and ended in April 2015. When necessary, NIHD was consulted.

Data errors are minimised by data control and clean-up, which are meant to ensure that answers are logical, consistent with each other and that the questionnaires do not contain entry errors. One of the biggest advantages of the personal interview filled on a laptop compared to an interview with a paper questionnaire is data clean-up during the interview. There are logical checks in the electronic environment for this purpose which notify of a conflict right after answering the last question the check is related to. The interviewer must check these logical conflicts immediately during the interview, and depending on the problem, either correct the answer or add an explanatory note to the question in the questionnaire. The data clean-up operations for the submitted questionnaires have been described below.

- The logical checks occurred during the interview and explanations added by the interviewers were reviewed. Sufficient explanations were accepted, in the case of mistakes or insufficient explanations, the interviewer asked more information. If necessary, the interviewer contacted the respondent again.
- Interviewer-added notes were reviewed. If necessary, answers were corrected or the interviewer was asked for more information. If necessary, the interviewer contacted the respondent again.
- The explanations added to the answer option "other" were reviewed. If necessary, answers were corrected or the interviewer was asked for more information. If necessary, the interviewer contacted the respondent again.

Mistakes were also found in numbers with decimal places (e.g. in the weight variable). The discovered mistakes were corrected.

The following classifications were used in the data processing of the health survey for coding answers: classification of occupations (ISCO-08), classification of administrative and settlement units of Estonia (EHAK), classification of languages, country of citizenship list, classification of nationality, international statistical classification of diseases and related health problems (ICD-10),

international standard classification of education, classification of states and territories, national standard classification of education, list of religions, classification of curricula and fields of training. The answers written under the option “Other” were reviewed and coded under the existing answer options, if possible.

Preparation of the data set intended for data users started in May 2015 and ended with the formation of a data file for Eurostat in September 2015.

The microdata set of the survey was submitted to Eurostat on schedule, at the beginning of October 2015. Estonian data set was validated by Eurostat in February 2016 and control protocols were sent to SE for correcting the mistakes found. Statistics Estonia sent the amended data file back to Eurostat in July 2016. The mistakes turned out to be mainly coding errors – the given coding requirements and filters had not been sufficiently considered in preparing the dataset. Eurostat also checked non-response rates by variables. The household income variable was missing for 26% of respondents. The interview language variable was missing for 13% of the respondents, 12% of these were related to the fact that this information was not collected about the web interview (CAWI).

## 4.1 Weight calculation

***Maiki Ilves, Head of Data Processing Development Department at Statistics Estonia***

The population of the survey included of individuals living in Estonia as of 1 January 2014 who are at least 15 years of age and not living in an institutional household. By weighing the results are extended to the described target population.

Based on the final result, the 8,500 individuals in the sample divided as follows:

- respondents – 5,452 (64% of sample),
- non-response – 2,701 (32% of sample),
- frame error – 347 (4% of sample).

Weights are calculated in three stages: design weights, weights adjusting non-response and calibrated weights, which are final weights.

### 1. Design weights

Design weights are calculated based on the design of the sample, which in the case of this survey is a systematic choice in the strata. There are 4 strata in total and the design weight is constant within one stratum. Design weight is calculated as the ratio of the number of sample objects in a stratum to the number of objects in a stratum in a frame. Design weight is calculated for all individuals in the sample.

### 2. Weight adjusting non-response

Logistic regression has been used to account for non-response. The probability of response was calculated using logistic regression, where the variables used for explaining response were: sex\*5-year age group indicator, county (Tallinn separately), rural-urban variable calculated on the basis of the individual's place of residence. There are certainly more variables affecting non-response, but for using them in the model, it is necessary to have the value of the variable for all individuals in the sample.

The weight adjusting non-response MVKAAL is calculated as design weight divided by the probability of response. The weight adjusting non-response is only calculated for individuals who responded.

### 3. Calibrated weight

As the weight adjusting non-response does not take into account coverage problems or other variables affecting response that were not taken into account in the previous stage.

Calibration means the adjustment of weights with a certain factor so that upon adding up the weights we get the given calibration sums, e.g. size of population. The size of the population based on the Statistics Estonia database as of 1 January 2014 has been selected for the calibration sum. Summary amounts consider the surveyed target population (individuals aged 15+) excluding the individuals living at institutions. If the number of institutionalised individuals in a surveyed group is not known, the number of institutionalised individuals in this group has been evaluated based on the principle of proportionality. The following variables were used for calibration: sex\*5-year age group, county (Tallinn separately), nationality, education (ISCED 2011) groups 0–2, 3–4, 5–8, education unknown). Education variable was linked to the EHIS2014 data set using the SE education database variable, which is used for calculating the size of population based on education and for publication in the SE public database (see SE database table RV0231) (10).

Calibration was done by SAS macro CALMAR and resulted in the calibrated weight LKAAL. Calibrated weight is calculated only for the respondents.

“European Health Interview Survey Manual” recommended trimming the weights if they vary too much. The need for trimming was checked in the case of both MVKAAL as well as LKAAL and there was no need for trimming the weights. Value 3 was used as the trimming factor as was recommended in the manual.

## 4.2 Results

Eurostat's quality report, which reported the results of the EHIS2014 survey, presented the distribution of respondents based on main variables (table 1) and main health questions (table 2).

**Table 1.** Distribution of respondents based on main variables, EHIS2014

|  | Target population |             | Sample       |             | Responses    |             | Non-responses |             |
|--|-------------------|-------------|--------------|-------------|--------------|-------------|---------------|-------------|
|  | Number            | %           | Number       | %           | Number       | %           | Number        | %           |
| <b>Total</b>                           | <b>1,107,791</b>  | <b>100%</b> | <b>8,500</b> | <b>100%</b> | <b>5,452</b> | <b>100%</b> | <b>3,028</b>  | <b>100%</b> |
| Women                                  | 599,788           | 54%         | 4,622        | 54%         | 3,138        | 58%         | 1,564         | 51%         |
| Men                                    | 508,003           | 46%         | 3,878        | 46%         | 2,314        | 42%         | 1,484         | 49%         |
| <b>Women (age)</b>                     |                   |             |              |             |              |             |               |             |
| 15-24                                  | 71,158            | 12%         | 489          | 11%         | 290          | 9%          | 199           | 13%         |
| 25-34                                  | 91,780            | 15%         | 702          | 15%         | 449          | 14%         | 253           | 17%         |
| 35-44                                  | 89,812            | 15%         | 678          | 15%         | 444          | 14%         | 234           | 16%         |
| 45-54                                  | 91,119            | 15%         | 697          | 15%         | 491          | 16%         | 206           | 14%         |
| 55-64                                  | 95,161            | 16%         | 738          | 16%         | 526          | 17%         | 212           | 14%         |
| 65-74                                  | 77,014            | 13%         | 611          | 13%         | 467          | 15%         | 144           | 10%         |
| 75-84                                  | 61,993            | 10%         | 504          | 11%         | 353          | 11%         | 151           | 10%         |
| 85+                                    | 21,751            | 4%          | 203          | 4%          | 118          | 4%          | 85            | 6%          |
| <b>Men (age)</b>                       |                   |             |              |             |              |             |               |             |
| 15-24                                  | 75,804            | 15%         | 526          | 14%         | 311          | 13%         | 215           | 14%         |
| 25-34                                  | 97,804            | 19%         | 732          | 19%         | 386          | 17%         | 346           | 22%         |
| 35-44                                  | 92,307            | 18%         | 693          | 18%         | 385          | 17%         | 308           | 20%         |
| 45-54                                  | 85,519            | 17%         | 651          | 17%         | 376          | 16%         | 275           | 18%         |
| 55-64                                  | 75,935            | 15%         | 593          | 15%         | 391          | 17%         | 202           | 13%         |
| 65-74                                  | 48,237            | 10%         | 397          | 10%         | 279          | 12%         | 118           | 8%          |
| 75-84                                  | 27,026            | 5%          | 229          | 6%          | 154          | 7%          | 75            | 5%          |
| 85+                                    | 5,672             | 1%          | 57           | 1%          | 32           | 1%          | 25            | 2%          |
| <b>Level of education (ISCED 2011)</b> |                   |             |              |             |              |             |               |             |
| ISCED 0-2                              | 249,289           | 23%         | 1,982        | 24%         | 1,255        | 24%         | 727           | 24%         |
| ISCED 3-4                              | 469,679           | 44%         | 3,618        | 44%         | 2,297        | 43%         | 1,321         | 44%         |
| ISCED 5-8                              | 351,228           | 32%         | 2,682        | 32%         | 1,787        | 33%         | 895           | 32%         |
| <b>Status of employment</b>            |                   |             |              |             |              |             |               |             |
| Employed                               | ..                | ..          | ..           | ..          | 2,938        | 54.0%       | ..            | ..          |
| Unemployed                             | ..                | ..          | ..           | ..          | 11           | 0.2%        | ..            | ..          |
| Pensioner                              | ..                | ..          | ..           | ..          | 1,418        | 26.0%       | ..            | ..          |
| Student                                | ..                | ..          | ..           | ..          | 360          | 7.0%        | ..            | ..          |
| Other non-active                       | ..                | ..          | ..           | ..          | 485          | 9.0%        | ..            | ..          |
| <b>Population density</b>              |                   |             |              |             |              |             |               |             |
| Medium-density region                  | 762,638           | 69%         | 5,641        | 66%         | 3,429        | 63%         | 2,212         | 73%         |
| High-density region                    | 345,153           | 31%         | 2,859        | 34%         | 2,023        | 37%         | 836           | 27%         |
| <b>Size of household</b>               |                   |             |              |             |              |             |               |             |
| 1                                      | ..                | ..          | ..           | ..          | 1,429        | 26%         | ..            | ..          |
| 2                                      | ..                | ..          | ..           | ..          | 1,876        | 34%         | ..            | ..          |
| 3                                      | ..                | ..          | ..           | ..          | 1,019        | 19%         | ..            | ..          |
| 4                                      | ..                | ..          | ..           | ..          | 715          | 13%         | ..            | ..          |
| 5+                                     | ..                | ..          | ..           | ..          | 413          | 8%          | ..            | ..          |

**Table 2.** Distribution of respondents based on health-related questions, EHIS2014

| Indicator or answer to a question  | Number of respondents | Assessment (weighted) | Standard error of assessment | Lower and upper confidence limits of the 95% confidence interval |
|--|-----------------------|-----------------------|------------------------------|--|
| <b>Respondents with good and very good health (HS1)</b>                                    |                       |                       |                              |  |
| <b>Total</b>   | <b>3,116</b>          | <b>0.591</b>          | <b>0.007</b>                 | <b>0.577; 0.604</b>  |
| Women  | 1,728                 | 0.624                 | 0.010                        | 0.603; 0.644   |
| Men  | 1,388                 | 0.563                 | 0.009                        | 0.545; 0.581   |
| <b>Respondents with a long-term disease or health problem (HS2)</b>                        |                       |                       |                              |  |
| <b>Total</b>   | <b>3,582</b>          | <b>0.649</b>          | <b>0.007</b>                 | <b>0.636; 0.663</b>  |
| Women  | 2,180                 | 0.600                 | 0.011                        | 0.579; 0.621   |
| Men  | 1,402                 | 0.691                 | 0.009                        | 0.674; 0.708   |
| <b>Respondents with significant health-related long-term limitations on activity (HS3)</b> |                       |                       |                              |  |
| <b>Total</b>   | <b>597</b>            | <b>0.105</b>          | <b>0.004</b>                 | <b>0.097; 0.113</b>  |
| Women  | 384                   | 0.089                 | 0.006                        | 0.077; 0.100   |
| Men  | 213                   | 0.119                 | 0.006                        | 0.107; 0.130   |
| <b>Respondents who had been hospitalised during the last 12 months (H01)</b>               | <b>549</b>            | <b>0.098</b>          | <b>0.004</b>                 | <b>0.090; 0.106</b>  |
| <b>Respondents aged 18+ who were obese (body mass index <math>\geq 30</math>)</b>          | <b>1,108</b>          | <b>0.196</b>          | <b>0.006</b>                 | <b>0.185; 0.206</b>  |

### 4.3 Complementation of data set with data from administrative sources

The health survey usually asks individuals for information that cannot be obtained from other sources. The questionnaire of EHIS2014 was capacious, including both the mandatory part of the European Health Interview Survey as well as the questions necessary for comparisons with data from the previous health surveys conducted in Estonia. When designing the survey, linking with health data from administrative data sources was added to the plans. SE had implemented the linking of registry data to social surveys for decreasing the volume of survey instrument and lowering expenses. This option was also applied for the health survey to enrich the data set.

A direct need for linking data arose already for obtaining the highest completed level of education. It is known that the Estonian Education Information System holds data on education acquired in Estonia since 2005, but nothing before that nor on education acquired elsewhere. Great amount of work was done by SE for the 2011 Population and Housing Census in order to obtain the education variable for individuals. Individuals were asked for very detailed information in order to achieve a result with the required accuracy. It was decided that the results of the population census should be used for obtaining the EHIS2014 education variable, which was supplemented with Education Information System data from the years following the census. Only questions about education completed abroad and one control question with basic education levels were added to the questionnaire.

Obtaining the income of the respondent and their household from registry data was also considered. It is known that people do not want to answer questions about income. Upon designing EHIS2014, it was known that registry data provide results of sufficient quality on only some types of income. A possibility was discussed to add the personal identification codes of the household members to the questionnaire and to use them later to find the income of the household based on the data in registers. However, it was decided that the personal ID codes would not be asked, as it was foreseen that respondents would oppose or would not know the codes by heart. For identifying the household members when compiling the data, SE used a conventional procedure as in all social surveys, that is based on the sex, date of birth and composition of the household. There were a total of 132,324 household members, 129 of whom were left unidentified.

NIHD compiled a list of linkable data sources and registry variables, and SE ordered the desired variables. Compilation of the list was based on the possibilities of existing registry data. The additional need for health as well as socio-economic background variables was taken into account. For obtaining background variables, it was decided that a selection of data from the last population census would also be used. However, some questions were left in the questionnaire about which registry data was collected from administrative sources, in order to assess later the suitability of using registry data instead of the data asked in the questionnaire by comparing these two, and thus reducing the load on the respondents in the next survey.

The asked extra variables for EHIS2014 linked from the registers were the following.

**Estonian Unemployment Insurance Fund (as of end of 2014):**

- 1) Period of unemployment insurance benefit payments – start and end date, 1<sup>st</sup> and 2<sup>nd</sup> daily rate, starting date for the 2<sup>nd</sup> daily rate (paid during 2014)
- 2) Sum of redundancy payment and date it was awarded (paid during 2014)
- 3) Unemployment benefit payment period during 2014 – start and end date
- 4) Periods of registration as unemployed in 2008–2014 start and end date

**Estonian Health Insurance Fund (data of 2014):**

- 1) Sickness benefits by type (sickness benefit, care allowance, maternity benefit) – paid amount, number of days compensated
- 2) Having the certificate of incapacity for work during 2014 by type of incapacity for work (certificate for sick leave, for care leave, etc.)
- 3) Number of periods of incapacity for work (certificates) and total number of days – all days irrespective of payment in 2014
- 4) Existence or lack of health insurance as of the beginning of 2014, insurance basis (employee, child/student, pensioner, etc.)
- 5) Number of days of inpatient and day care, number of ambulatory care cases in 2014, Estonian Health Insurance Fund medical expenses
- 6) Dental care benefit
- 7) Prosthesis compensation

**Estonian Education Information System (EHIS):**

- 1) Completing a level of education in 2012–2014 (for obtaining the higher level of education the period is 1 January 2012 to 31 December 2014)
- 2) Current studies – students at the beginning and end of 2014 and the respective level (different academic year)

**Estonian Tax and Customs Board (data of 2014):**

- 1) Person's income from employment (gross)
- 2) Sickness benefits (i.e. the part of benefits covered by the employer, the rest comes from the Health insurance Fund)

**Estonian National Social Insurance Board (sums paid during 2014 as of the end of 2014 by type):**

- 1) Old-age pension
- 2) Pension for incapacity for work
- 3) Survivor's pension
- 4) Child allowance
- 5) Childbirth allowance
- 6) Parental benefit (parental wages)
- 7) Childcare allowance
- 8) Single parent's child allowance
- 9) Conscript's child allowance
- 10) Allowance for a parent with seven or more children
- 11) Adoption allowance
- 12) Adoption benefit
- 13) Foster care allowance
- 14) Start in independent life allowance for a child in a children's home
- 15) National maintenance allowance for a child
- 16) Type and severity of disability (incapacity for work valid or established during 2014)
- 17) Percentage of permanent incapacity for work (as of end of 2014 and changes during 2014)
- 18) Time of initial establishment of disability

**Defence Resources Agency (as of end of 2014)**

- 1) Has been declared fit for conscript service (variable = "yes")
- 2) Has been declared unfit for conscript service (variable = "no")
- 3) Has been declared temporarily unfit for conscript service

**Digital Prescription Centre (data of 2014)**

- 1) Digital prescription number
- 2) ATC code
- 3) Active pharmaceutical ingredient
- 4) Code of the package sold
- 5) Name of the package sold
- 6) Date of purchase of medicine
- 7) Number of packages sold
- 8) Person's expenses on a prescription
- 9) Expenses of the Health Insurance Fund on prescriptions per person

**Data from the 2011 population census:**

- 1) Highest completed level of education at the end of 2011 (according to the ISCED 2011 classification)
- 2) Permanent residence at the time of the census at local government level (E01). If the place of residence is the same at the time of the interview as it was during the census, the information characterising this living space (E02, E15–E26)
- 3) Place of residence at the time of the previous census in 2000, at the local government level
- 4) Location of principal job (A47) at the local government level
- 5) Birthplaces of parents and grandparents (A14, A14C)
- 6) Country and place of birth (A13, A13A)
- 7) Number of children given birth to until 31 December 2011 (A11A, for women aged 15 or older). Add data on children born in 2012–2014 to the data from the census
- 8) Religion A21, A21A
- 9) Nationality A17
- 10) Native language A19
- 11) Health status A50, A51
- 12) Main source of subsistence in 2011 A34

13) Language skills other than the native language A20, A20A

14) Growing agricultural and horticultural produce and keeping animals for self-consumption E28

#### **Population register**

1) Citizenship

2) Place of residence at the local government level

Therefore, the data set of EHIS2014 includes, in addition to the data collected with the interview, a selection of additional data linked from registers in a previously agreed volume. After the initial review of the linked data for the survey, inaccuracies were discovered, i.e. the data did not always contain the content asked for and shortcomings were detected in the descriptions. SE, as the holder of the data sets, was asked to correct the mistakes found.

SE, as the holder of the data set, has prepared the data for users with a little user-friendliness. The order of variables in the file does not always correspond to that in the added metadata (questionnaire and description), some inaccuracies exist. As the data sets are capacious, the survey data materials contain numerous different codes, names and redirections. User have to be careful when using the data and not to make mistakes. There are also similarly named variables on the same topic, which is why before using them, it is necessary to verify that you have chosen the desired variable. The requests to use the data must be submitted to the holder of the data sets, to Statistics Estonia.

## **5 Data dissemination**

After fieldwork had ended at the beginning of 2015, the cooperation between SE and NIHD continued in order to compile the data sets of the survey and make the database and registry data available. Two meetings took place (on 27 February and 13 April) between the representatives of NIHD and SE, where substantive issues and activities were agreed on. Unfortunately, the agreements were not realised and NIHD could not start processing the data in 2015. NIHD was granted remote access to the data set of EHIS2014 a year later, at the end of March 2016. Only then NIHD was able to get acquainted with the data set of the survey for the first time.

Eurostat published the first indicators of health surveys conducted in the Member States at the end of summer 2016. This was preceded by repeated enquiries by Eurostat to SE asking to correct the submitted data file. As was later revealed, the main problem was mistakes in the coding of variables in the microdata file, which was caused by inadequate precision in following the coding instructions. Correction of those errors was delayed from February to July and there was a risk that Estonian data would be left out of the first publication round. However, SE managed to implement the necessary changes in dataset and sent the improved microdata set to Eurostat. After that, NIHD was also given remote access to the microdata set submitted to Eurostat. Estonia's health data was published together with the data of other Member States in autumn 2016 in the databases of Eurostat and OECD and in publications (incl. "Health at a Glance: Europe 2016. State of health in the EU cycle").

At the end of 2016, NIHD published the first selection of the comparable indicators of the 2006 and 2014 surveys in the Health Statistics and Health Research Database. In 2017, the selection was supplemented to some extent. The indicators based on the income of a household member were not included in the selection published in 2016–2017, also the new indicators collected for the first time in 2014 were not published.

Solving some issues took longer than expected. The question of the imputation necessary for compensating for the non-responses to the income variable was raised in February 2016 based on the Eurostat microdata control protocols, which disclosed, among other things, summary reports of non-responses by variables. Variables related to the income of a household are important social

background variables and they are used together with health issues to highlight inequality. At first, SE postponed imputing in 2016, but in 2017 gave up due to the lack of an employee. Imputation of the income variables for the data set of EHIS2014 was completed by A. Võrk with the support of the Ministry of Social Affairs in 2018.

104 tables in 13 sub-divisions were published on the data of the 2006 health survey in the Health Statistics and Health Research Database of NIHD. Furthermore, 115 tables were released in the collection of tables based on 15 background variables, separately for men and women. The publication of the first part of the 2014 survey results was limited to 30 tables in NIHD's public Health Statistics and Health Research Database (11).

## 6 Comparability of data from 2006 and 2014 health surveys

Users of the Health Interview Survey data should keep in mind that even if a question had the same wording in the 2006 and the 2014 survey or the change in it was minimal, the results may differ significantly. It is known that several factors influence the answer given, including the placement of the question in the questionnaire and previous questions (i.e. the order of the questions) and the theme of the survey. In addition to the subjective assessment of the respondent and their state of being during the interview, answers are also influenced by the response rate of the survey and the used survey method, among many other reasons. In the case of a computer-assisted interview, a lot depends on the used validation rules. The prior health survey in 2006 was conducted using the traditional paper questionnaire with an interviewer.

It is not possible always to assess the impact of the differences of the asked questions to the survey results. In case of the two questions (self-perceived general health and long-standing health problem) of the Minimum European Health Module, the necessary comparative material existed in the time series of the Estonian Social Survey (EU-SILC). The extent of the impact was assessed by NIHD analyst Anne Selart in the respective analysis based on Estonian Health Interview Survey 2014 and 2006 data completed at the end of 2016 (12).

The only difference between the questions about self-perceived health in the 2006 and 2014 surveys was the wording of the intermediate answer category between good and poor health. In 2006, the expression "average" was used for characterising the health status, in 2014, it was replaced by the more neutral version "neither good nor bad". The analysis showed that the choice between "good" and the intermediate option is influenced by the wording in the case of respondents aged 45 and older. In the 45+ age group, the chances of choosing the intermediate option instead of "good" in the case of the wording "average" is about 1.4 times higher than with the wording "neither good nor bad". When looking at the options "bad" and the intermediate option, the influence of wording becomes apparent with women under 45 of age. The chance of choosing the intermediate option instead of "bad" is lower with the wording "average" than with the wording "neither good nor bad". (12)

When looking at the question concerning long-standing health problem, it was noted that in the results of the 2014 survey, the chances of the long-standing health problem being noted in the age groups 16–44 are 1.3–1.5 times higher compared to the 2006 survey. The impact on the answers was smaller in the older age groups. (12) The reason was the (automatic) validation rule added to the 2014 survey questionnaire, which forced the respondent to return to the question if they said later in the questionnaire that they have any of the asked chronic diseases if they had initially said that they have no long-standing health problems. The 2006 survey used a paper questionnaire and the results were not corrected later based on the answers to the question about diseases, although a similar control between the two questions was made. In 2014, 66% of people aged 16+ had a long-standing disease or health problem, in 2006 that percentage was 50%.

In the case of the question about long-standing limitations in activities people usually do because of health problems., there were differences in the wording of the question, but the answer options were the same. When in 2006 people were asked about limitations in "daily activities", then in 2014, for the purpose of following more accurately the concept, instead the wording "in activities people usually do" was used. The 2006 survey also used the expression "in relation to a long-term disease or health problem", in 2014 the expression was "due to some health problems" without any reference to a long-term disease. With this question, it was unfortunately not possible to determine the impact caused by the amended wording and due to the time passed. Because the change in the wording of the Estonian Social Survey was perhaps even greater. In 2006 the question was "Are activities limited?", in 2014 the question was "To what extent are activities limited?" It should be

kept in mind that the question in the Estonian Social Survey about health-related long-standing limitations in activities is also the basis for calculating the number of healthy life years (HLY), which is used by Eurostat as well as OECD.

The questions of the Minimum European Health Module have been used in many health-related surveys. In the PHC 2011, the two last questions were asked, the first one concerning self-perceived health was excluded. This was thought to be too subjective to be allowed to answer by someone else, which is common in a population census. The purpose of adding health-related questions to a census was to obtain information on health status on the local government level and about various smaller regions that interview surveys do not enable. The health status data of the census is published in the Statistics Estonia database by local governments, for Tallinn in addition by city districts. When interview surveys are conducted about limited age groups (e.g. 16–64, 15–74), the advantage of a census is that all age groups are covered, incl. children aged 14 and younger. Even though some time has passed, this is the best available data on population self-perceived health status for the smaller regions.

## References

1. Leinsalu M, Grintšak M, Noorkõiv R. Eesti terviseuuring: tabelid. Estonian health interview survey: tables. Tallinn: Eksperimentaalse ja Kliinilise Meditsiini Instituut; 1999.
2. Oja L, Matsi A. Eesti Terviseuuring 2006. Tabelid. Estonian Health Interview survey 2006. Tables. Tallinn: National Institute for Health Development; 2009. <https://www.tai.ee/en/publications/download/148>.
3. Regulation (EC) No 1338/2008 of the European Parliament and of the Council of 16 December 2008 on Community statistics on public health and health and safety at work. L 354/70 (31.12.2008). <https://eur-lex.europa.eu/legal-content/EN/TXT/?qid=1564571329050&uri=CELEX:32008R1338>.
4. Commission regulation (EU) No 141/2013 of 19 February 2013 implementing Regulation (EC) No 1338/2008 of the European Parliament and of the Council on Community statistics on public health and health and safety at work, as regards statistics based on the European Health Interview Survey (EHIS). L 47/20 (20.02.2013). <https://eur-lex.europa.eu/legal-content/EN/TXT/?qid=1564571450579&uri=CELEX:32013R0141>.
5. Eurostat. European Health Interview Survey (EHIS wave 2). Methodological manual 2013 edition. Luxembourg: Publications Office of the European Union; 2013. <https://ec.europa.eu/eurostat/web/products-manuals-and-guidelines/-/KS-RA-13-018>.
6. European Commision. European Core Health Indicators (ECHI). [https://ec.europa.eu/health/indicators\\_data/echi\\_en](https://ec.europa.eu/health/indicators_data/echi_en).
7. Oja L, Matsi A, Leinsalu M. Eesti Terviseuuring 2006. Metodoloogiaülevaade. Estonian Health Interview survey 2016. Methodological Report. Tallinn: National Institute for Health Development; 2008. <https://www.tai.ee/en/health-data/research-reports/download/333>.
8. Aluoja A, Shlik J, Vasar V, et al. Development and psychometric properties of the Emotional State Questionnaire, a self-report questionnaire for depression and anxiety. Nordic Journal of Psychiatry 1999;53(6):443–449. <http://dx.doi.org/10.1080/080394899427692>.
9. Eurostat. Quality report of the second wave of the European Health Interview survey. Luxembourg: Publications Office of the European Union; 2018. <https://ec.europa.eu/eurostat/web/products-statistical-reports/-/KS-FT-18-003?inheritRedirect=true&redirect=%2FEurostat%2Fpublications%2Fstatistical-reports>.
10. Statistics Estonia. Statistical database: Population – Population indicators and composition. [http://pub.stat.ee/px-web.2001/I\\_Databas/Population/01Population\\_indicators\\_and\\_composition/04Population\\_figure\\_and\\_composition/04Population\\_figure\\_and\\_composition.asp](http://pub.stat.ee/px-web.2001/I_Databas/Population/01Population_indicators_and_composition/04Population_figure_and_composition/04Population_figure_and_composition.asp). Viewed on 4 May 2018.
11. National Institute for Health Development. Health Statistics and Health Research Database. Health and health behaviour: Estonian Health Interview Survey. <http://pxweb.tai.ee/PXWeb2015/pxweb/en/05Uuringud>. Viewed on 3 January 2019.
12. Selart A. Tervise enesehinnang ja pikaajalise haiguse esinemine Eesti Terviseuuring 2014 ja 2006 andmetel. Tallinn: National Institute for Health Development; 2016. <https://www.tai.ee/en/health-data/research-reports/download/479>.

## **ANNEXES**

1. Questionnaire in Estonian
2. Questionnaire in Russian
3. Interviewer manual in Estonian
4. Interviewer manual in Russian

The Estonian and Russian questionnaires and interviewer manuals of EHIS2014 are published on the website of National Institute for Health Development

<https://www.tai.ee/et/tegevused/teadustoo/loppenud-uuringud/eesti-terviseuuring-2014>.

## **Health and health care statistics:**

- **Health statistics and health research database**  
<http://www.tai.ee/tstua>
- **Website of Health Statistics Department of National Institute for Health Development**  
<http://www.tai.ee/en/r-and-d/health-statistics/activities>
- **Dataquery to National Institute for Health Development**  
[tai@tai.ee](mailto:tai@tai.ee)
- **Database of Statistics Estonia**  
<http://www.stat.ee/en>
- **Statistics of European Union**  
<http://ec.europa.eu/eurostat>
- **European health for all database (HFA-DB)**  
<http://data.euro.who.int/hfadb/>
- **OECD's statistical databases (OECD.Stat)**  
[http://stats.oecd.org/index.aspx?DataSetCode=HEALTH\\_STAT](http://stats.oecd.org/index.aspx?DataSetCode=HEALTH_STAT)