Data comparison of the Health Information System and regular health statistics.
III analysis

National Institute for Health Development
Department of Health Statistics

DATA COMPARISON OF THE HEALTH INFORMATION SYSTEM AND REGULAR HEALTH STATISTICS.

III analysis

Tallinn 2013

Mission of the Department of Health Statistics:

Public Health and Welfare through Better Statistics and Information

Authors:

Angela Poolakese
Hedi Liivlaid
Merike Rätsep
Natalja Eigo
Riina Tilk

When using or citing the data in this analysis, please refer to the source.

TABLE OF CONTENTS

List of tables

Definitions used

Adult – in health care statistics: a person aged 15 and more

Child – in health care statistics: a person aged 0–14

Discharge type – hospital checkouts, transferred or deceased

Entry – information about a single case or a stay at the department

Hospital admission – hospitalisation of a patient in an inpatient facility, involving a stay of at least for 24 hours

Hospital bed specialty – determined through the health care services provided in connection with the hospital bed

Patient status – recovered, deteriorated or unchanged

Treatment case – a single treatment case

Used abbreviations

EeHF –   Estonian e-Health Foundation

EHAK –   Classification of Estonian administrative units and settlements (*Eesti haldus- ja asustusjaotuse klassifikaator*)

e-Health –   Health Information System

ICD-10 –   the 10th revision of the International Statistical Classification of Diseases and Related Health Problems

DHS –   Department of Health Statistics of National Institute for Health Development

NCSP –   NOMESCO Classification of Surgical Procedures

NIHD –   National Institute for Health Development

INTRODUCTION

Created in 2008, the Health Information System (e-Health), managed and developed by the Estonian e-Health Foundation (EeHF), is a health sector cooperation model that incorporates various services, an important part of which is a database that is part of the state information system[1]. This is definitely one of the largest databases containing data necessary for health statistics. Therefore, it is logical to develop health statistics by starting using a system which allows the use of already collected detailed data for producing more accurate health statistics. Last year, the long-awaited statistics module of the e-Health was initiated, the main purpose of which is to introduce the treatment information of patients to be transferred to the e-Health for statistics. There are also expectations to reduce the administrative burden for both the health care providers and the respondents by transferring from reports to administrative data.

In recent years, the EeHF, the Ministry of Social Affairs, the Department of Health Statistics (DHS) of National Institute for Health Development (NIHD), the software developers, and the health care providers have agreed on the initial inpatient and day care document standards that are forwarded to the e-Health. The description of the information flow has been started, which would specify what kind of data and how to implement the statistical module. Also, some definitions were elaborated and quality control rules were developed. In summer 2013, it is planned to carry out a pilot project of the statistics module to evaluate the feasibility of collecting the agreed data.

By the second half of 2011, the number of medical documents in e-Health had exceeded a certain critical limit and the possibility of using existing information occurred. In order to improve the quality of documents received by the e-Health and prepare the transition from the regular statistics reports to the information sent to the e-Health, the HCD carried out two analyses to measure the quantity and quality of the epicrises submitted to the e-Health.

This analysis is the third in its line, providing an overview of the completion of inpatient epicrises received in the e-Health. This work focuses on the same period as in the second analysis, i.e. the data of the third quarter of 2011. The analysis observes the completion of data fields concerning diagnoses, surgical operations, patient's place of residence, and emergency of the case, i.e. characteristics that were not studied in the previous analysis. It is planned to implement the analysis results in the pilot project of the statistics module and consider it also in further developments.

The purpose of the analysis is:

- to analyse the completion of the characteristics of the inpatient epicrisis and compliance with the rules;
- to obtain information from definitions and concepts that require review and agreement between the target groups;
- to obtain information on how to prepare data extract and how the data set changes depending on the moment the extract is prepared;
- to identify problems that need to be addressed when implementing the statistics module.

---

[1] https://www.tai.ee/en/r-and-d/health-statistics/e-health-information-system

1. METHODOLOGY

For the purpose of the analysis, the EeHF provided the DHS with the inpatient epicrises received in the e-Health, where the patient's stay in a hospital remained in the period between 1 July 2011 and 30 September 2011, regardless of the date on which these are received in the e-Health system. The extract was made as of 27 April 2012.

The description and data set of the inpatient, or case epicrisis have been published at http://pub.e-tervis.ee/manuals/Meditsiinidokumentide%20kirjeldus/1.0 (*in Estonian*).

1.1 Characteristics analysed and the related lists/classifications

Table 1. **Characteristics, lists/classifications**

Characteristic	List/Classification	Link to the list/classification (*in Estonian*)
Patient's place of residence	Classification of Estonian administrative units and settlements (EHAK)	http://pub.etervis.ee/classifications/EHAK/2007v2
Hospital admission emergency code and name	Scheduled admission_emergency	http://pub.etervis.ee/classifications/P%C3%B6%C3%B6rdumise%20plaanilisus_erakorralisus/1
Code and name of the diagnosis of the principal disease	The 10th revision of the International Statistical Classification of Diseases and Related Health Problems (ICD-10)	http://pub.etervis.ee/classifications/RHK-10/5
Code and name of the statistical type of the diagnosis of the principal disease	Statistical type of diagnosis	http://pub.etervis.ee/classifications/Diagnoosi%20statistiline%20liik/2
Code and name of the complication of the principal disease	ICD-10	http://pub.etervis.ee/classifications/RHK-10/5
Code and name of the statistical type of the complication of the principal disease	Statistical type of diagnosis	http://pub.etervis.ee/classifications/Diagnoosi%20statistiline%20liik/2
Code and name of the intercurrent disease	ICD-10	http://pub.etervis.ee/classifications/RHK-10/5
Code and name of the statistical type of the intercurrent disease	Statistical type of diagnosis	http://pub.etervis.ee/classifications/Diagnoosi%20statistiline%20liik/2
Code and name of the external cause	ICD-10	http://pub.etervis.ee/classifications/RHK-10/5
Code and name of the statistical type of the	Statistical type of diagnosis	http://pub.etervis.ee/classifications/Diagnoosi%20statistiline%20liik/2

Characteristic	List/Classification	Link to the list/classification (*in Estonian*)
external cause		
Date of surgery	–	–
NCSP code and name	NOMESCO Classification of Surgical Procedures (NCSP) 2010 version	http://pub.etervis.ee/classifications/NCSP/2
Code and name of the health insurance fund	Price list of the Estonian Health Insurance Fund (EHIF)	http://pub.etervis.ee/classifications/Haigekassa%20hinnakiri/5
Code and name of the anaesthesia	Types of anaesthesia	http://pub.etervis.ee/classifications/Anesteesia%20liigid/2
Code and name of the additional instrument of the health insurance fund	Medical equipment of the 2008 list of the health insurance fund	http://pub.etervis.ee/classifications/EHK%202008a.%20hinnakirja%20meditsiiniseadmed/2

The characteristics that we compared in the context of an entry made with two different dates (December 2011 and April 2012) were the following: number of entries, number of cases, and the number of hospitals that presented the epicrises.

1.2 e-Health data editing

The file forwarded by the EeHF in April 2012 had a total of 183,823 entries with 54,074 cases. After the data editing, 178,301 entries and 50,721 cases remained.

Outpatient and day care treatment cases, as well as duplicate cases were deleted from the data file. Those cases where there was no information on the bed profile were considered to be outpatient. If the hospital bed specialty code was P49, these cases were considered to be day care treatment cases.

To identify duplicate cases, four data fields were compared to each other: the registry code of the health care institution, the beginning of the treatment case, the end of the case, and the patient ID number (PID).

Original file

183,823 entries,
54,071 cases

Deleted cases

Combination (registry code of facility, the beginning of the case, the end of the case, PID) repeats = duplicate cases (29)

Hospital bed specialty is empty = outpatient cases (2,311)

Hospital bed specialty is P49 = day care treatment cases (1,013)

Final file
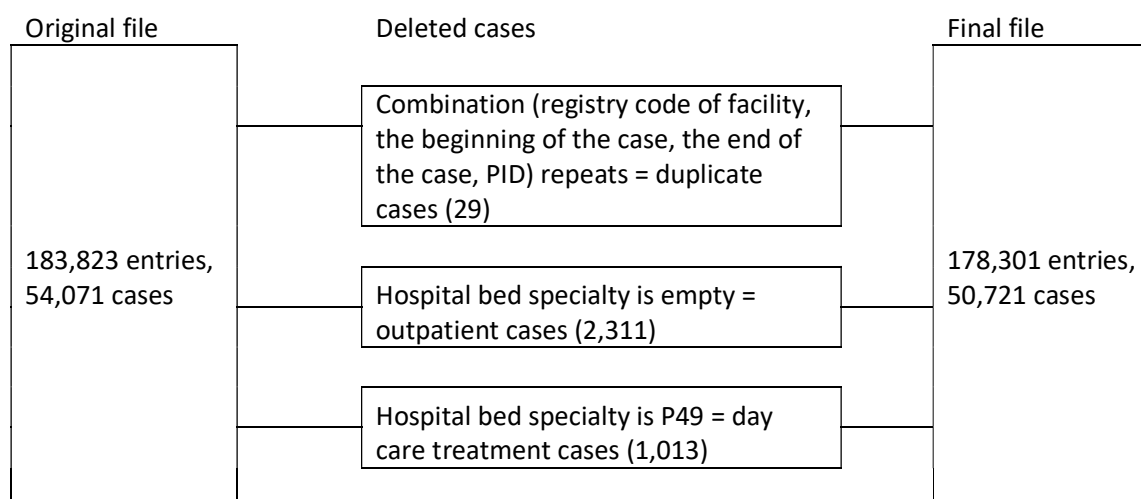
178,301 entries,
50,721 cases

Figure 1. **Editing of the e-Health datafile for the data analysis**

Most of the duplicate cases had a different document number. The e-Health instruction foresees that a new version of the same document should be submitted when making corrections. In three cases, the document number was the same, but the version number was also the same.

The documents are provided with numbers within the institution, therefore, documents from different hospitals had the same numbers, and they cannot be considered to be duplicate cases.

Similarly, to the previous analysis, there were also duplicate hospital bed specialties in the provided file (15,254 entries in the source file). These are entries where there are several entries on the same patient, the same time period, and the same hospital bed specialty within the same case. During the analysis, it became evident that these duplicate hospital bed specialty entries had different diagnoses, therefore, the entry as a whole was not a duplicate. Meaning that the analysis of the hospital bed specialty data required the use of different entries compared to the handling of diagnoses.

1.3 Completion of the e-Health characteristics, verification of the compliance to the provided format and the lists and classifications published in the EeHF

We compared the characteristics and their compliance with the standard. In case of major deviation, we consulted with the health care providers and the EeHF to identify whether this error could be eliminated, whether it is a temporary or continuous error. Absolute figures or percentages have been used to compare the indicators.

In the results section, the information about corresponding characteristics' standards, which has been published in the EeHF or used for health statistics is presented in *Italic* fond.

2. RESULTS

2.1 Comparability of data / compliance of an extract

When comparing the document and version numbers with the data in the previous extract, the document had a new version number in 23 cases. Documents with a new version number were submitted by five hospitals.

Compared to the documents received with the previous extract, the new extract did not include 137 documents anymore. During the consultation with the EeHF, it became evident that during 6–7 December 2011, there was a technical error in receiving the documents, which resulted in the fact that the processing of documents remained incomplete and the valid documents for some cases went missing. In three cases, a new document with the same inpatient care start and end dates about the same patient was already available.

More than half (56%) of the so-called missing documents came from one hospital. The share of the following two hospitals was 22% and 10%, respectively.

This extract included 2,622 documents, which were not present in the previous extract. These involved cases that were sent to the EeHF after the previous extract of the data.

Table 2. **Comparison of the extract as of December 2011 and April 2012**

	December 2011	**April 2012**
Entries	118,876	178,301
Cases	48,237	50,721
Hospitals	25	25

2.2 Determination of case

2.2.1 Definition of case as an inpatient case

In the health statistics, the following definition has been agreed on: Inpatient case – the patient is taken to a hospital (preparation of inpatient medical history) with the intention of providing treatment lasting at least one day (24 h). In case of death or transfer to another hospital, the case duration may be shorter.

In order to assess the correctness of the case definition, we compared the atypical cases, i.e. hospital admission and leaving from the hospital took place on the same day.

The same arrival and departure dates occurred in 4,877 cases, of which 141 were referred to another hospital and 92 ended with death. In 2,124 cases, where the date of arrival and the date of leaving were identical, contained the operation date and the NCSP code, with the most frequent operations including cataract surgeries, adenotonsillectomies, cervical abrasions, aborts, which are usually carried out in day care.

Only one health care provider had indicated both the admission and discharge times. One third of these cases lasted up to six hours (including one tenth to four hours).

10

During the consultation with the health care providers, it became evident that the inpatient epicrisis was used to transfer both hospital and day care data for technical reasons in 2011. To distinguish between inpatient and day care treatment case, the characteristic "case type" was used. Since the extract of the EeHF does not contain the "case type" data, it is neither possible to say the exact number of day care cases nor exclude them from the analysis.

Table 3. **Cases ended on the same day as the date of arrival by hospital**

	Share of the cases on the same day, %	Share of the same day cases, %		
		checked out	transferred	deceased
Hospital 1	25.0	96.5	2.9	0.6
Hospital 2	23.7	99.3	0.1	0.6
Hospital 3	19.4	96.3	2.6	1.1
Hospital 4	18.4	59.6	40.4	0.0
Hospital 5	17.0	93.9	3.2	2.8
Hospital 6	14.8	97.8	1.5	0.7
Hospital 7	13.1	96.0	3.0	1.0
Hospital 8	7,8	97.6	1.2	1.2
Hospital 9	5,6	81.4	16.3	2.3
Hospital 10	5.5	96.2	1.3	2.6
Hospital 11	3.6	88.8	7.7	3.6
Hospital 12	3.5	79.3	15.5	5.2
Hospital 13	3.2	69.7	24.7	5.6
Hospital 14	3.0	86.6	3.1	10.2
Hospital 15	2.8	85.7	14.3	0
Hospital 16	2.6	63.2	26.3	10.5
Hospital 17	1.9	52.9	41.2	5.9
Hospital 18	1.3	66.7	0	33.3
Hospital 19	0.9	0.0	100.0	0.0
Hospital 20	0.8	60.0	20.0	20.0
Hospital 21	0.0	-	-	-
Hospital 22	0.0	-	-	-
Hospital 23	0.0	-	-	-
Hospital 24	0.0	-	-	-
Hospital 25	0.0	-	-	-
Total	*9.6*	*95.2*	*2.9*	*1.9*

2.2.2 Specific cases for opening a new case

An epicrisis should be transferred to the e-Health when the case is completed. *The data dictionary of medical terms, which has been prepared as part of the digital health record project in 2006, has stated that the inpatient case involves the activities related to the examination and treatment of a patient in a single health care provider from hospitalisation to leaving the hospital, and long-term nursing care is a special case.*

Based on the health statistics reporting and hospital surveys, it can be assumed that a different approach is used to open a new case. Since the number of cases is an important indicator, the NIHD aims at harmonising the definition. As of the beginning of 2013, it has been agreed in the working group of the health statistics definition that a new inpatient case within the same hospital is opened in the e-Health only if treatment is

continued as part of long-term nursing care, taking into account that this includes the provision of nursing care.

A new case was opened in connection with the same patient in the same health care provider according to the end date of the previous case 851 times. Almost a third of these were cases of one hospital. A total of 328 cases (39%) involved transfer, the rest of them included checkout.

In half cases (446 cases, 52%), the principal diagnosis was the same. Also, 60 cases had the same principal diagnosis and hospital bed specialty.

In 96 cases, a new case was opened with the same hospital bed specialty, and 54 cases of them involved the nursing care beds. There were 342 cases where a new case was opened with the nursing care bed and the previous cases had ended with another profile. Among these, almost equal number of patients were checked out and transferred, 174 and 168, respectively.

There were 343 cases where a new case is opened with regard to the same patient in the same hospital on the next day. Of these, 96% ended with checkout and 4% with transfer. In 240 cases, a new case was opened with the same hospital bed specialty: in 72 cases, it involved long-term nursing care, 39 cases psychiatry, 19 gynaecology, 18 surgery, and 10 general medicine and pregnancy pathology hospital bed specialty.

For 207 (60%) cases out of 343, the principal diagnosis was the same. Both the hospital bed specialty and the principal diagnosis were the same in 48% of cases (165 cases).

While the long-term nursing care profile was generally started with a new case, three hospitals also showed other profiles alongside the long-term nursing care profile under one case. Such cases amounted to a total of 144 (in comparison, there were a total of 1,649 cases with the long-term nursing care profile, which were presented by 14 hospitals). Of these cases, 90% originated from one hospital.

2.3 Data fields related to patient's place of residence

The previous analysis verified the compliance with the county and state characteristics, which are not repeatedly described in this analysis.

2.3.1 Patient's place of residence: city – *Patient's alleged actual place of residence*
Text field. There are no instructions on whether it is required to use the name of the city or rural municipality in accordance with the Classification of Estonian administrative units and settlements *(EHAK).*

Epicrises did not only indicated the cities, but the data was also presented at the level of rural municipalities and towns. There were 436 different values presented in the file. The field had not been completed in 17,421 cases, i.e. 34% of cases.

2.3.2 Patient's place of residence according to the Classification of Estonian administrative units and settlements (EHAK)

EHAK is indented for use to mark a territorial location. The classification covers the following: state administration units – counties; local government administration units – rural municipalities and cities;

limited local government administration units – rural municipality districts and city districts; other settlements and settlement units – cities without municipal status, towns, small towns, villages and urban regions; others. Each classification object has been provided with a unique four-digit code, and the name of the object and its alternative place name have also been indicated (if available). Additional code-related characteristics (classification part) indicate the county and local government of each object, as well as its type (county, rural municipality, city, town, city without municipal status, rural municipality district, city district, small town, village or other settlement).

By the numeric value of the code, the following are distinguishable:
- level of county (code smaller than 100)
- level of local governments (code from 100 to 999)
- level of settlements (code greater than 1,000)

(See Classification of Estonian administrative units and settlements 2013v3, http://metaweb.stat.ee/)

There were 1,158 different codes in the file. The field had not been completed in 34,709 cases, i.e. 68.4% of cases. In all cases, the field was completed only by three health care providers. In some cases, the field was completed by five hospitals, and the rest of the health care providers had not completed the given field at all.

Although, according to the classification, the code of the EHAK has four digits, all codes indicated by the hospitals had either five or six digits. If there was a five-digit code, both the county and the rural municipality code were indicated according to the classification. If there was a six-digit code, also the code of the type of the place of residence (town, small town, village, city, city district, county, rural municipality, city without municipal status) was indicated in addition to the county code.

2.4 Hospital admission emergency

2.4.1 Hospital admission emergency (code)

Numeric field. The basis is the classification "Scheduled admission_emergency" ("Pöördumise plaanilisus_erakorralisus"). The classification provides three codes: 1, 2, 3.

The characteristic was completed for all cases. Only one hospital used all three admission types (however, based on the division, it may be assumed that the main difference lies between emergency and scheduled admission; in only few cases, the admission has been indicated as inevitable). Other hospitals used only two codes: either 1 and 2 or 1 and 3. In consultation with the hospitals, it has become clear that, in terms of content, only emergency admission is provided with the emergency and inevitable admission code.

2.4.2 Hospital admission emergency (name)

Text field. Textual interpretation of the previous indicator. According to the classification, corresponds to the codes in the previous clause: 1 – scheduled; 2 – inevitable; 3 – emergency.

The characteristic was completed for all cases. The analysis showed that hospitals mostly only distinguish between two admission types: scheduled and emergency, however, the corresponding admission code name differed by hospital. Three hospitals indicated the emergency admission with the inevitable admission

code. The rest of the hospitals used the emergency admission code with the emergency admission. In one hospital, there were few cases with the inevitable admission, and the rest were scheduled or emergency cases.

2.5 Diagnoses

In nine cases, no diagnosis was indicated (no principal or intercurrent diagnosis, complications, external causes).

2.5.1 Compliance of the ICD-10 code levels

ICD-10 codes are used to transfer the names of disease diagnoses and other health problems to alphanumeric codes[2] to allow systematic registration, analysis, interpretation and comparison of international mortality and morbidity data.

ICD-10 has codes with three, four and five digits. *(NB: The fourth digit in the code follows a decimal point, which is not referred to as a separate code digit.) In certain cases, the use of six-digit codes is also allowed; their use has not been agreed in Estonia. Mostly, the disease conditions are coded with four digits, but in most chapters, there are diagnoses that are coded with three digits. For chapters "Psychiatric and behavioural disorders" (Chapter V) and "External causes of illness and death" (Chapter XX), most codes have five digits.*

In health statistics reports, the disease states are usually indicated by a three- or four-digit code, the exceptions include "Psychiatric and behavioural disorders" (Chapter V) and "External causes of illness and death" (Chapter XX), which are encoded with five digits by the psychiatrists. In the Health Information System, the encoding is planned at the lowest level, i.e. if there is a five-digit code in the ICD-10, the diagnosis should be encoded with five digits.

There were used a total of 10,526 diagnosis codes on the epicrises, which have been presented with five digits in the ICD-10, and a total of 121,654 diagnosis codes, which have been presented with four digits. A total of 99% of the four-digit codes were correctly indicated. Slightly less than half of the five-digit codes were correctly indicated.

Table 4. **Indication of diagnosis codes according to the ICD-10 levels**

	Five code places in the ICD	**Four code places in the ICD**
Total codes	10,526	121,654
Has been indicated as the four-digit code	4,995	120,792
Share, %	47.5	99.3
Has been indicated as the three-digit code	196	862
Share, %	1.9	0.7

In 4,995 cases, where should be five-digit ICD-10 code, the diagnosis code included only four digits. These codes are included in Chapter V (F00–F99), Chapter XIX (S00–T98), and Chapter XX (V01–Y98) of the ICD-10. Such incomplete codes were presented by almost all hospitals.

---

[2] Code containing letters and numbers.

Also, there were 196 cases where the codes in the ICD had five digits, but only three digits were used. Eight hospitals used the three-digit code instead of the five-digit code. The same number of hospitals also used the three-digit code instead of the four-digit code.

In few cases, the diagnosis code that required four digits according to the ICD-10 was indicated with the five-digit code. In most cases, this concerned the conditions presented in Chapter F.

2.5.2 Code of the diagnosis of the principal disease

A text field that is based on the ICD. In the e-Health, there is a rule that each case must have one principal diagnosis.

The diagnosis of the principal disease was missing in 27 cases. In nine of these cases, there were no diagnoses in the epicrisis marked. In other cases, the external cause (12 cases), intercurrent disease (3 cases) or intercurrent and external cause (2 cases) had been completed. Two thirds of the entries came from one hospital (18).

One principal diagnosis was marked according to the e-Health rule in 93% of cases. Approximately 7% of the cases had more than one principal diagnosis.

Table 5. **Number of the diagnoses of the principal disease**

	Number of the diagnoses of the principal disease													Total
	0	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**12**	**13**	
Number of cases	27	47,185	1,897	911	410	180	52	33	16	6	1	2	1	**50,721**

There were also some not existing ICD codes (F06.40, F41.00, F43.00, F44.90, F60.40, F61.00, F71.10, F73.10, M46.24) presented in the principal disease section at the epicrisis. Most of these codes were from Chapter F.

2.5.2.1 Cases with more than one principal diagnosis

A total of 3,509 cases had more than one principal diagnosis. In one third of these cases, only principal diagnoses were marked. In other cases, several principal diagnoses were also accompanied by the intercurrent diagnoses, complications or external causes.

Between hospitals, there were great differences in completing the single principal diagnosis rule. Four hospitals had more than one principal diagnosis in more than ten percent of cases, including more than half of the cases in one hospital. The other hospitals had indicated one principal diagnosis, or there was minimal deviation from the rule. In consultation with the health care providers, it came out that the presence of several principle diagnoses was mainly due to an information technology problem in one hospital, where the complications from the principal disease ended up under the principal diagnosis. The e-Health extract as of December 2012 showed that compliance with the principal diagnosis rule has improved. Of all epicrises submitted to the e-Health (inpatient, day care, and outpatient), only 219 cases had several principal diagnoses, including a maximum of three diagnosis codes per case.

The ICD allows multiple codes to be used to describe the condition in some circumstances, therefore, more than one diagnosis code can be shown as a principal diagnosis by a physician. Such cases include:

- **Use of the cross and asterisk system**

For more complete information on the condition, it is allowed to encode certain conditions using two codes – in addition to the principal disease, the organ or site is encoded. In this case, the principal disease in the ICD-10 is marked with a cross and an additional code with an asterisk. The code with a cross is the initial code that should be always used. One should never use a code with an asterisk alone.

In most hospitals, inaccuracy in tagging an asterisk was noted, although the number of cases was not high. The code with the asterisk was the principal diagnosis in the 220 cases listed first. Of these, in 210 cases, the asterisk code was the only principal diagnosis. More frequently, i.e. one third of cases, dementia was incorrectly encoded, in one fifth of the cases, secondary anaemia was incorrectly encoded, and in 13% of cases, psoriatic arthropathies were incorrectly encoded.

- **Bacteria, viruses and other infectious agents (B95–B97)**

According to ICD rules, these sub-divisions must never be used for primary coding, but as further or additional codes to identify pathogens classified elsewhere.

The codes B95–B97 were indicated as the only principal disease in three cases, including one case showed the actual principal disease as the intercurrent disease.

- **Multiple injury**

When multiple injuries/sites occur at the same time, it is recommended to indicate an injury that is clearly more difficult and resource-intensive as the principal diagnosis in the statistics. If no condition dominates in its seriousness, the multiple injury code can be registered as the main condition. Not to lose the information, physicians may have reported more than one injury under the principal diagnosis.

In 104 cases, more than one diagnosis code for injury (Chapter XIX) was reported under the principal disease. In one third of the cases, along with injury codes, there were also intercurrent diseases or complications. In the case of the rest, it was possible to encode most of them by more serious injuries.

In the ICD, it is permitted to use two codes to indicate psychiatric conditions with a specified condition and tumours with functional activity with the specification of the function type did not occur in the data extract.

2.5.3 ICD code (principal disease)

A text field that shows which ICD version hospital is using to classify diseases.

For all cases, where the principal disease code was completed, the ICD code was *ICD-10*.

2.5.4 Name of the principal disease diagnosis

A text field that is based on the ICD code name and related to the field *diagnosis code of the principal disease.*

While there were 3,950 different diagnosis codes of the principal disease, there were also 10,065 different diagnosis names of the principal disease. When looking at the code and the name in combination, there

were 10,099 different combinations, which means that several names corresponded to one code, and there were also names with multiple codes.

Only for 1,864 codes there was one verbal match. The largest number of different names occurred for the code C61, a total of 67. There were 66 different names within one hospital. When excluding this hospital, there were only two different names for the code C61. The multiplicity of different names within one code was due to the fact that the hospital had added additional text to the classification name.

The difference in the name of the classification was also due to the use of abbreviations and the addition of the title of the chapter to the sub-code and its name.

Cases where two codes corresponded to one name were 31 in the extract. In addition, three different codes were used for one name. When looking at the classification, the same sub-sections have the same names. For example, the counterpart for F70.1, F71.1, and F72.1 in the classification is "Significant behavioural disorder", i.e. there is no specification of the higher levels.

2.5.5 Statistical type of the principal diagnosis (code)

A text field that should be completed for each principal diagnosis code. The field is based on the classification *Statistical type of diagnosis*. numbers 1, 2, and 3 are used as the classification codes, and encodings "+", "−", and "0" are used as short names.

The statistical type of diagnosis was completed in 42,126 cases. In 8,568 cases where the main diagnosis existed, the statistical type of diagnosis was not indicated. A total of 99.8% of cases with no statistical type of diagnosis occurred in one hospital (8,555 cases).

Excluding those cases where there was no code for the statistical type of diagnosis, the statistical type of diagnosis was indicated as much as the diagnosis of the principal disease. This means that, for example, if there were nine diagnoses of the principal disease, the same number of codes and names of the statistical type of diagnosis was also found.

Some hospitals used numeric encoding 1, 2, and 3, but others the symbols "+" and "−".

Table 6. **Statistical type of the principal diagnosis**

Code	Short name	Name	Long name
1	+	first illness	first illness
2	−	recurring illness	recurring illness
3	0	initial diagnosis	initial diagnosis

2.5.6 Statistical type of the principal diagnosis (name)

A text field that should be completed for each principal diagnosis code. The field is based on the classification Statistical type of diagnosis and serves as the textual explanation of the previous field. The guide to the epicrisis does not specify whether a short or long name should be used.

Table 7. **Using the code and name of the statistical type of the principal diagnosis**

Code	Name	Frequency
-	repetitive case in life	7
-	recurring	2,892
+	first case in life	6
+	first time	5,202
1	+	4,214
1	first illness	14,913
2	-	3,892
2	recurring illness	12,197
3	0	379
3	initial diagnosis	245

In all cases where the statistical type code field of the principal diagnosis was completed, there was also a name available. Similarly, to the code field, there is also no uniform completion between the hospitals. One health care provider used the short name of classification ("+", "−", "0"), which was also used for completing the code field by some other hospital. Some hospitals used the long name (first illness, recurring illness, initial diagnosis) and some used names are not based on the classification at all.

2.5.7 Code of the complication from the principal disease

A text field that is based on the ICD classification. If completed, there may be several codes.

The code of the complication from the principal disease was indicated for 7,223 cases and 11,401 times. A total of 1,342 different codes have been used from Chapter A to Chapter Y.

The code of the complication from the principal disease was not indicated in 85.8% of cases. Of those cases, where the complication code was available, one complication code was indicated in 64% of cases, two codes in 23% of cases, and three codes in 8% of cases. The maximum indicated number of complications in a case was 14.

One code that was not present in the ICD classification was also used to indicate the complication of the principal disease – F04.11.

Table 8. **Number of complications from the principal disease in a case**

	Number of codes in a case														Total
	0	1	2	3	4	5	6	7	8	9	10	11	12	14	
Number of cases	43,498	4,643	1,653	574	202	72	44	17	5	4	4	1	2	2	**50,721**

2.5.8 ICD code (Complication from the principal disease)

A text field showing which classification has been used to encode the complication from the principal disease.

For all cases, where the code of the complication from the principal disease was completed, the field value was *ICD-10*.

2.5.9 Complication from the principal disease (name)

A text field that is based on the ICD code name and related to the field code of the complication from the principal disease.

There were 2,662 different the ICD code names used in the epicrises. Given that there were 1,342 different codes, it can be seen that several different names corresponded to the same ICD code. There were 2,667 different combinations of code and name, meaning that also several different codes corresponding to some names.

2.5.10 Statistical type of the complication from the principal disease (code)

A text field that should be completed for each complication code. The field is based on the classification *Statistical type of diagnosis*. The classification provides that numbers 1, 2, and 3 are used as codes, and encodings "+", "–", and "0" are used as short names.

Two different encodings were used. In one case, there are numbers 1, 2, and 3, and in other case, there are symbols "+" and "–".

The field was empty in 2,182 times and in 1,329 cases where the code of the complication from the principal disease was completed. While the statistical type of the principal disease was predominantly not completed by one hospital, there were more facilities that had indicated the complication code but not the statistical type of the complication. Six hospitals had not indicated the statistical type of complication at all.

2.5.11 Statistical type of the complication from the principal disease (name)

A text field that should be completed for each complication code. The field is based on the classification *Statistical type of diagnosis and serves as the textual explanation of the previous field.*

In all cases where the field "statistical type code" was completed, also the name was present. Eight different names were used in the epicrises. In addition to verbal explanations, this column also included "+" and "–" and "0".

2.5.12 Intercurrent disease (code)

A text field that is based on the ICD classification. If the field is completed, several codes may be indicated.

There were 3,940 different values indicated.

For few cases, diagnostic codes that are not included in the ICD were used (F03.00, F06.70, F07.00, F19.60, F25.10, F41.10, F45.00, F45.80, F60.20, F60.40, F70.10, F84.10, M42.96, M47.85, M48.02, M48.95, M54.29, M80.00). These codes are presented in ICD not with four but five digits.

Table 9. **Number of intercurrent diseases in a case**

	Intercurrent disease codes																		Total
	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	
Number of cases	20,932	11,139	7,168	4,841	2,974	1,692	976	472	251	133	71	28	18	11	8	4	2	1	**50,721**

2.5.13 ICD code (intercurrent disease)

A text field showing which classification has been used to encode the intercurrent disease.

For all cases, where the intercurrent disease code was completed, the field value was *ICD-10*.

2.5.14 Intercurrent disease (name)

A text field that is based on the ICD code name and related to the field *diagnosis code of the intercurrent disease.*

There were 9,271 different names, and 9,314 different code and name combinations. Based on the above, several names corresponded to one code and, at the same time, there were names that corresponded to several codes.

2.5.15 Intercurrent disease (statistical type code)

A text field that should be completed for each intercurrent disease code. The field is based on the classification *"Statistical type of diagnosis"*. The classification provides that numbers 1, 2, and 3 are used as codes, and encodings "+", "−", and "0" are used as short names.

However, two different encodings were used. In one case, there were numbers 1, 2, and 3, and in other case, there were symbols "+" and "−".

The statistical type of the intercurrent disease was not indicated in 4,923 cases where the intercurrent disease code had been indicated. Here was a similar picture of the type of code for the diagnosis of the principal disease diagnosis: the field was mostly empty, and four hospitals had few uncompleted cases.

2.5.16 Intercurrent disease (statistical type name)

A text field that should be completed for each intercurrent disease code. The field is based on the classification *Statistical type of diagnosis and serves as the textual explanation of the previous field.*

There were 10 different names used. In addition to verbal expressions, this column also had symbols "+" and "−" and "0". In all cases where the previous field was completed, also the name was presented.

2.5.17 External cause (code)

Text field. If the field is completed, there can be only one code.

The field is based on the ICD. The first three codes (one letter code and two number codes) indicate the type of external cause, the fourth digit marks W00–Y34 (except for Y06, Y07); in case of external causes to be encoded, the number code of the site and the fifth digit V01–Y34 shows the number code of the activities of the external causes to be encoded.

There were 1,337 different codes in the external cause code field.

There were altogether 4,511 cases of external causes (codes V01–Y98). In 38 of these cases, the external cause was mentioned as the principal disease, in four cases as the complication from the principal disease, and in 20 cases as the intercurrent disease (a total of 1.4% of cases).

A total of 4,724 cases with an external cause were filled. Cases where the external cause field was filled but it did not contain external cause codes amounted to 265 (about 6%). Of these, the diagnoses of injury, poisoning and the consequences of certain other external causes (codes S00–T98) were reported in 32 cases

(12%). Other diagnoses of A00–R99 amounted to 154 cases and the codes Z00–Z99 were used in 96 cases (58% and 36%, respectively).

Of the mentioned 265 cases, where the external cause field did not contain external cause codes, 254, i.e. 96%, were submitted by one hospital.

Indication of external causes in the event of cases including codes S00–T98 in Chapter XIX of ICD-10

The diagnoses of injury, poisoning and the consequences of certain other external causes (codes S00–T98) were reported in a total of 4,878 cases. Of these, 4,376, i.e. 90% also had the code of external cause.

Of 502 cases where the external cause code was unmarked, 386 cases, or 77% were reported by two hospitals (49% and 28%, respectively).

In the event of 502 cases without an external cause, the codes of injury, poisoning and other consequences of external causes (Chapter XIX) were indicated as follows:

Table 10. **Cases of injury without external cause (Chapter XIX S00–T98) by diagnosis**

ICD-10 code	Name	Number of cases
S00–S99	Injuries to different areas of the body	163
T00–T75	Injury, poisoning and certain other consequences of external causes	42
T78	Adverse effects, not elsewhere classified	5
T79	Certain early complications of trauma	3
T80–T88	Complications of surgical and medical care, not elsewhere classified	212
T90–T98	Sequelae of injuries, of poisoning and of other consequences of external causes	77

Of all 4,878 cases with the injury and poisoning diagnosis (S00–T98), in 4,227 cases (i.e. 87%) the injury code was indicated in the column of the principal diagnosis. Of these, 363 cases (about 9%) lacked an external cause. The code of S00–T98 was the only code under the principal disease in 3,804 cases (i.e. 90% of the cases of injury indicated in the principal diagnosis column), of which 192 cases used the so-called combined injury (the section S00–S99 codes, where the fourth digit is ".7" and the section T00–T05 codes).

Cases with several external cause code

Of 4,459 cases, where the external cause code is under the external cause, 65 cases (i.e. 1.5%) include several marked codes. Of these, only seven cases included two codes of external cause (V01–Y98). A total of 40 cases include in addition to an external cause also the codes S00–T98, and the event of 28 cases, also other codes were entered (A00–R99, Z00–Z99).

Table 11. **Number of external causes in a case**

	Number of codes indicated in the external cause field								Total
	1	2	3	4	5	6	7	8	
Number of cases	4,394	38	11	8	3	4	0	1	4,459

In four cases out of seven, there were two codes of external cause (W00–Y34), in one case, there were two codes of complications of other medical procedures (Y84), in one case, there was a code of medical device malfunction (Y79) added to falling (W01), and in only one case, the alcohol poisoning code (X45) was accompanied by the relation to alcohol verified by determining the level of poisoning (Y91).

Use of additional factors related to the causes of illness and death classified elsewhere (external cause codes Y90–Y98)

Of all the treatment cases with an external cause, the codes of the additional factors of external causes were indicated as follows:

Table 12. **Diagnoses of the additional factors of external causes (Y90–Y98)**

ICD-10 code	Name	Number of cases
Y90	Relation to alcohol that has been verified by determining the blood alcohol content	4
Y91	Relation to alcohol that has been verified by determining the poisoning strength	2
Y95	Hospital conditions	2
Y96	Occupational conditions	30

Of the 38 additional factors presented, 34 were marked as the only code under the external cause. Three hospitals had submitted relations with alcohol (Y90–Y91). One hospital had submitted 27 cases with the code concerning conditions related to professional work.

Compliance of the external cause code levels

Of 4,459 treatment cases with an external cause, the latter was marked with a five-digit code in 2,856 cases (i.e. 64%), with a four-digit code in 1,400 cases (31%), and with a three-digit code in 203 cases (5%). Of the four-digit codes, 975 (or 70%) belonged to the code group V00–Y34, meaning that these should be five-digit codes. Of the three-digit codes, 155 cases (or 76%) should have five digits (V00–Y34) and three cases four digits (Y83, Y79). The total number of cases, where external cause was indicated with three or four digits but should be indicated with a five-digit or four-digit external cause code, amounted to 1,133 (i.e. 25% of all cases related to external causes). These cases were presented by 20 hospitals out of 22 health care institutions who had submitted cases with external causes, and 60%, or 679 of cases were presented by four hospitals.

Indication of complications of surgical or other medical assistance

In 212 cases, the fields of the principal disease, the intercurrent disease, the complication from the principal disease or external cause (125, 33, 36, and 18 cases, respectively) were classified as "Complications of surgical and medical care, not elsewhere classified" (codes T80–T88), however, there was no external cause code in the field of external cause. Similarly, 77 cases included the codes T90–T98 referring to injury, poisoning and other consequences of external causes (59 principal diseases, one case of the complication from the principal disease, and 17 intercurrent diseases), however, the field of external cause did not have the code of external cause.

2.5.18 ICD code (external cause)

A text field showing which classification has been used to encode the external cause.

For all cases, where the external cause code was completed, the field value was ICD-10.

2.5.19 External cause (name)

A text field that is based on the ICD code name and related to the field code of the external cause.

There were used 1,863 different names. There were 1,874 different code and name combinations. Based on the above, several names corresponded to one code and, at the same time, there were names that corresponded to several codes.

2.5.20 External cause (statistical type code)

A text field that should be completed for each external cause code. The field is based on the classification *Statistical type of diagnosis*. The classification provides that numbers 1, 2, and 3 are used as codes, and encodings "+", "–", and "0" are used as short names.

Two different encodings were used, in one case, numbers 1, 2, and 3, and in other case, the symbol "+".

The statistical type of external cause was not mentioned in 2,267 treatment cases when the external cause code was indicated.

2.5.21 External cause (statistical type name)

A text field that should be completed for each external cause code. The field is based on the classification *Statistical type of diagnosis and serves as the textual explanation of the previous field.*

There were seven different names. In addition to verbal expressions, this column also had symbols "+" and "–" and "0". In all cases where the previous field was completed, also the name was present.

2.5.22 Use of certain diagnoses

The practice of using some certain diagnoses is described below. In cases where there was more than one diagnosis under the principal diagnosis, the first diagnosis is taken as principal diagnosis.

2.5.22.1 R00–R99 "Symptoms, signs and abnormal clinical and laboratory findings, not elsewhere classified"

R was the principal disease in 540 cases, and the intercurrent disease has not been mentioned in 311 cases. The average length of stay was 2.41 days. In four cases, where the principal disease was R code, death was indicated as the discharge type.

Table 13. **More frequent R codes as the principal disease and the average number of bed days**

ICD-10 code	Name	Number of cases	Average number of days of inpatient care
R10.4	Other and unspecified abdominal pain	81	1.69
R49.0	Voice disorder, or dysphonia	41	0.00
R04.0	Nose bleeding, or epistaxis	41	2.76
R11	Nausea and vomiting	28	1.50
R55	Syncope, or fainting and collapse, or unannounced loss of postural tone	28	3.29
R06.5	Mouth breathing	24	0.75
R56.8	Other and unspecified afebrile convulsions	23	2.22
R56.0	Fever strokes	20	1.20
R10.3	Localised pain in other areas of lower abdomen	15	1.27
R31	Unspecified haematuria, or blood in urine	12	4.58
R50.9	Unspecified fever	12	5.92
R07.3	Other chest pain	10	2.30
R33	Urinary retention	10	4.10
R42	Vertigo, or dizziness	10	5.00

2.5.22.2 Z00.1 – regular medical check of a child

As a rule, a newborn's inpatient epicrisis is not sent to the e-Health, a birth epicrisis is prepared instead. Based on the ICD-10, healthy newborns are encoded under Z38 "Liveborn infants according to place of birth".

The code Z00.1 was used to first indicate the principal disease in 675 cases. All cases occurred in one hospital. The number of bed days was 0–5, the average number of bed days was 2.02. Obviously, the newborn babies get also a medical history (date of birth July–September 2011). In six cases, the intercurrent disease was indicated – Chapter Q.

2.5.22.3 Z49.0 – preparatory care for dialysis, Z49.1 – extracorporeal dialysis, Z49.2 – other dialysis

The codes Z49.1 and Z49.2 were used to indicate the principal disease in 50 cases, where more than one code was marked under the principal disease. These codes have not been mentioned first.
In seven cases, the intercurrent disease was Z49.1 and Z49.2. There were no differences between the hospitals, three of them had indicated Z49.1 or Z49.2 as the intercurrent disease.
The external cause was Z49.0, Z49.1 and Z49.2 in eight cases. All cases occurred in a single hospital.
In 38 cases, the complication was indicated as Z49.1 and Z49.2. As there were so few cases, it could not be said that there was a difference in the indications.

2.5.22.4 Z51.0 – radiotherapy session, Z51.1 – Chemotherapy session for neoplasm, Z51.2 – Other chemotherapy

The code Z51.1 was indicated as the principal diseases in 75 cases, but only in nine cases the code had been mentioned first. If Z51.1 was mentioned first, all had Chapter C (tumours) referred to as the intercurrent disease. All cases indicated as the first diagnosis code were from one hospital.
In 779 cases, the intercurrent disease was Z51.0 and Z51.1. In most cases (546 times) the principal disease was indicated in Chapter C. Most of the cases were concerned with one hospital.

The codes Z51.0 and Z51.1 were used to indicate external causes in 31 cases. The principal disease was always Chapter C. Most of the cases belonged to one hospital, only one case was from another hospital. The code Z51 was not indicated as compliance.

2.6 Surgical operations
The concept of operation has not been defined in the e-Health. Health statistics indicates all surgical procedures encoded on the basis of the NCSP main chapters (Chapters A–Q), small surgical procedures (Chapter T), and taking organs for transplantation (Chapter Y).

All hospitals who had sent the inpatient epicrises and surgical operations data (i.e. the operation date, and the NCSP or health insurance fund code was completed) to the e-Health also showed haven inpatient surgery in the health statistics report, except for one.

A total of 12,037 inpatient epicrises with a surgical operation were sent to the e-Health. Of these, 2,124 cases included checkout on the same day, and it was likely that these cases included largely the day surgery service cases (see part 2.2.1).

2.6.1 Date of operation

The field format is YYYYMMDD (hhmmss).

In total, the operation date was performed in 12,157 cases and 14,903 times. In 7,138 cases, the operation time was with clock accuracy.

The extract includes 22 cases in which the operation date was not in the time framework of the case. Among them, there were 12 cases in one hospital.

In event of 144 cases, there were more operation dates indicated than the NCSP codes. Of these, 120 cases lacked both the NCSP code and the health insurance fund code. The additional instrument code was present in eight cases. The type of anaesthesia was reported in 138 cases out of 144. The availability of the operation date without the NCSP code was a problem of a couple of hospitals. In these cases, the most common diagnosis was spontaneous vertex delivery, acute lymphoblastic leukaemia, and acute maxillary sinusitis.

In 10,290 (84.6%) cases, there was one operation date. And there was a maximum of 22 operation dates.

2.6.2 NCSP code

A text field that is based on the NCSP. When completed correctly, it consists of three characters and two numbers.

In the event of 12,037 cases, at least one NCSP code was completed. This amounted to 23.7% of all cases. In four cases, there was the health insurance fund code indicated instead of the NCSP code. The extract includes one code that is not present in the classification (NHL41). In the rest of the cases, the NCSP codes complied with the presented format and classification.

A total of 1,438 different NCSP codes were used. Of hospitals, who submit a statistical report on inpatient surgery to DHS, not a single one had shown any codes. In most cases (84.7%) when the NCSP code was presented, the case had one NCSP code. The maximum indicated number of NCSP codes in a case was 22.

In all cases where either the health insurance fund code or the additional instrument code of the operation was marked, the NCSP code was also filled in. The cases where the NCSP code was present but the health insurance fund code or the additional instrument code was missing amounted to 3,361, or 28% of cases.

Use of the codes of Chapter Z (definers in all other chapters)
The NCSP rule says: *The procedure codes given in Chapter Z are the general procedure definers that provide additional information on the basic procedure codes. They are always included in the basic procedure codes and can never be used independently.*

The Z code was marked as the principal code in 270 cases, of which 252 were the only NCSP code. A total of 14 hospitals were shown as the main operation of the Z code, however, half of the cases were from a single hospital.

There was a total of 106 cases where the codes of Chapter T (small surgical procedures) were indicated. As a basic procedure, T codes were reported for 63 cases. In six cases with more than one NCSP code, the T code

was incorrectly indicated as the principal code. In the remaining 57 cases, the T code was the only NCSP code. Most of these were maternal episiotomies.

The Chapter U (transluminal endoscopy) code was the only code that was indicated 23 times and five times with other NCSP code.

The Chapter X (surgical examination procedures) NCSP code was the only operation code that was indicated in one case.

2.6.3 NCSP

A text field showing which classification has been used to encode the NCSP.

In all cases, where the NCSP code field was completed, the classification name was NCSP.

2.6.4 NCSP name

A text field that provides the field *NCSP code* with verbal match.

In all cases where the NCSP code was filled, the name was also filled in. There were 1,530 different names. There were 1,532 combinations of code and name, i.e. a couple of names correspond to a different code. Most codes corresponded to one name, but there were also codes with multiple matches. This was due to the presentation of letters with different spelling (š, ž and also <, >).

2.6.5 Health insurance fund code

A text field based on the list of the health insurance fund services, which is available on the website of the health insurance fund and also in Regulation No. 32 of the Government of the Republic "List of health care services of the Estonian Health Insurance Fund".

The health insurance fund code was filled in 8,090 cases, and it was left empty in 6,216 cases, when the NCSP code was present in the case. There were 600 different values presented. Both number and text codes were indicated.

Of hospitals, five facilities, which had presented the NCSP code, failed to submit the health insurance fund code. Two hospitals had copied the NCSP code to the health insurance fund code field (in three cases the main code was the Z… code and the actual main operation NCSP code under the health insurance fund).

In hospitals, the health insurance fund code started with zero, although the health insurance fund does not have zero in its own list.

2.6.6 Health insurance fund code (name)

A text field based on the list of the health insurance fund services, which is available on the website of the health insurance fund and also in Regulation No. 32 of the Government of the Republic "List of health care services of the Estonian Health Insurance Fund".

There were 650 different values presented. There were 749 different code and name combinations. This shows that, unlike the above indicators, there were more cases where many different codes corresponded

to one name. There were fewer cases where several different names corresponded to one code. The presence of different names was due to the fact that different lists were used.

2.6.7 Anaesthesia code

A text field, which is based on the classification "Type of anaesthesia". The field was completed in 10,050 cases; there were 32 different values. Two encodings were used. Most of the codes were based on the first version of the classification "Type of anaesthesia", which was valid until 1 January 2013. According to this, the codes were a letter (and number combinations). The other encoding used was similar to the new version of classification, according to which the field is encoded with numbers 1, 1.1...2,2.1...3.2. However, not all codes indicated by the hospitals are included in this version (4; 4.10; 5; 5.10; 10; 23) and the explanations of the codes in the classification do not correspond to the classification. This is probably another classification, not the second version of the classification "Type of anaesthesia" used by several different hospitals.

A total of 19 hospitals had submitted the data on anaesthesia. There was no code for anaesthesia in case of six hospitals in the extract.

2.6.8 Name of anaesthesia

A text field, the value of which must match the code in the previous field. There were 44 different names in the extract. There were 45 different combinations of code and name, while there were cases where the code was unmarked, but the name existed (5 times, 4 cases). One hospital had adjusted the names in the classification.

In total, the field was completed in 10,055 times.

2.6.9 Code of the additional resource of the health insurance fund

A text field that is based on the classification "Medical equipment of the 2008 list of the Estonian Health Insurance Fund". Correctly completed field consists of four numbers with the letter "L" at the end.

The field was filled 3,800 times, of which in 673 cases the hospital has not indicated the NCSP code or the health insurance fund code. There were 94 different values. Three of them did not comply with the format. Furthermore, neither the first nor the second version of the classification includes 13 codes used by the hospitals: 2623L, 2626L, 2750L, 2751L, 2756L, 2757L, 2759L, 2760L, 2862L, 2863L, 2901L, 2903L, 2904L.

The additional instrument code of the health insurance fund was shown by 17 hospitals.

2.6.10 Code name of the additional instrument of the health insurance fund

A text field that is based on the classification "Medical equipment of the 2008 list of the Estonian Health Insurance Fund".

The field value must match the code in the previous field. There were almost half more names than different codes (177 vs 94), which means that several different names corresponded to one code. The differences were due to the fact that the hospitals had clarified the explanations of the classification (included specific operation codes), shortened or used a different name from the classification. The differences were also caused by character differences. There were 178 different combinations of code and name, i.e. except in one case, one code corresponded to one name.

DISCUSSION

When reading the analysis, it should be taken into account that the data for 2011, which were submitted by April 2012, at the latest, have been observed. The technical problems that occurred then may be solved by now. In consultation with hospitals, it has emerged that IT systems have changed in several hospitals, quality controls have been added, and new documents have been introduced, such as the day care epicrisis.

One of the objectives of the analysis was to learn to prepare a case-by-case extract and to determine the characteristics that can be used to exclude duplicate and faulty cases. This goal was also met, but it is not clear for the DHS at this time whether this is sufficient and whether the given experience is helpful in the implementation of the statistical module. Some of the peculiarities of creating a datafile would need further clarification, such as how to make a single entry for a case in a more comfortable manner for a user (that the user would not make a single case of 13 entries if there are 13 diagnoses).

It is important that the same source conditions always have the same extract. As documents are corrected when new information becomes available, verifying this condition is complicated. In this analysis, an extract on the epicrises of hospital discharges was taken on the same period with a couple of months interval, and the total number of cases differed already by five percent, which is quite a big discrepancy. Documents have gone missing and have been added to the extract. It can be expected that technical problems will be solved and reduced by now. The analysis revealed the need for exact description and archiving of the inquiry and extract. It can be assumed that data will not change as the e-Health stabilises, but it must be agreed how often it is reasonable to update the output tables.

One of the aims of the analysis was to check the appropriateness of the rules / quality controls planned for the statistical module. For example, it has been agreed that there may be only one principal diagnosis per case. This rule is recognised by hospitals as necessary and considering the abundance of errors in the 2011 epicrises, when indicating the principal diagnosis, it is essential to add the control link. However, the ICD-10 allows certain codes to be encoded with multiple codes, of which only one code must be entered in the statistics. These states are, for example, using cross and asterisk system codes, multiple injuries, disease states with specific infectious agents, specification of functional activity of tumours. Although the allowed numerical multi-coding was often not used numerically (the problem might have been just the ability to encode such states). It would be necessary to discuss whether the requirement for a single principal diagnosis would not cause a situation where too little information is received on the diagnosis. For example, in statistics it is sufficient to have one principal diagnosis in case of multiple injuries in order to submit the inpatient care diagnoses, but less is known if this applies to planning injury prevention and treatment.

It is difficult to ensure the submission of data to the e-Health based on the common case definition, as physicians/hospitals submit the data to parties with different needs (health insurance fund, other physicians, statistics, etc.). Even in health statistics a treatment case is defined in many different ways. In statistics, there may be a need to adjust data. For example, in order to provide data based on an international definition, it may be necessary to add or subtract multiple cases into different episodes. Based on the analysis of the pilot project data, it is necessary to decide according to which rules the different cases should be combined (the length of period between various cases, how to consider the discharge type,

diagnosis, hospital bed specialty, etc.). It is also necessary to find a solution on how the data is presented in the tables of the e-Health statistical module reports.

As the health care institution provide data to different parties, the latter must monitor and agree on a common use of classifications.

In the statistical module, not only different case records, but also substantive errors can be problematic. Most quality controls have a deterrent effect, as it is often not possible to impose strict controls due to exceptions. If the health care personnel do not correct an obvious error, the question arises as to how to behave in this case? Whether to exclude the data from the report or make a statistical conversion? In this case, the different statistical results can cause confusion.

CONCLUSIONS

Preparation of data extract from e-Health

Approximately six percent of inpatient epicrises did not provide inpatient care data. As not all cases of day care were possible to distinguish and remove from the analysis, the number of inpatient cases was somewhat higher than in reality.

- In order to identify double cases, not only the combination of the document number and the identification facility's registry code should be used, but also the patient ID (PID) and the start and end of the inpatient treatment must be used to compare data.
- An inpatient epicrisis is not sufficient to be a definer for identifying an inpatient case. It is also necessary to verify the compliance between the case type and hospital bed specialty.

Completion of the characteristics of inpatient epicrisis and compliance with the rules

- The code of EHAK of the patient's place of residence was completed in one-third of cases, marked only by three hospitals. It should be noted that the codes are not presented as four digits, as provided by the classification, but the codes provided by the hospitals consist of two parts: the first two digits indicate the county; the last three or four digits show the rural municipality, city or town.
- Emergency of hospital admission was completed in all cases, but mostly two of the three lists were used. By excluding one hospital, other hospitals did not use the name of inevitable admission. The codes included scheduled, emergency and inevitable admission, however, in hospitals, only emergency or inevitable admission were indicated in addition to scheduled admission.
- Quantitative completion of the diagnosis characteristic was good, with only nine out of 50,721 cases being without any diagnosis.
- A common mistake in completing the diagnosis was failure to comply with one of the principal diagnosis rules, i.e. about 7% of the cases were reported with more than one principal diagnosis. Some of them were due to technical problems.
- Another common mistake in completing the diagnosis was marking the diagnosis code with insufficient level. A total of 99% of the four-digit codes in the ICD were correctly indicated. Diagnoses with five digits in the ICD were presented in a significantly poorer manner (i.e. diagnoses belonging to the chapter of mental disorders, injuries and external causes). Only half of the five-digit diagnoses in the ICD were correctly coded.
- In case of external causes, in addition to insufficient indication of the activity and site (25% of cases) the diagnoses according to the ICD-10 Chapter XIX "Injury, poisoning and the consequences of certain other external causes" were unspecified with an external cause (10%), or the codes of other conditions were entered in the field of external causes (6%).
- Some hospitals used changed names for the ICD-10 diagnoses, either by modifying the diagnosis or using abbreviations. For some cases the diagnosis code did not correspond to the ICD-10. In these cases, the conditions submitted with four-digit codes were encoded with five-digit codes in the classification.

- A total of 99% of cases, which included the operation date, also indicated at least one NCSP code. However, in 252 cases, the only NCSP code was the code of Chapter Z and in 23 cases, the code of transluminal endoscopy, and in one case, surgical examination procedure – Chapter X.
- The NCSP codes were indicated better than the codes of the health insurance fund. The Health Insurance Fund code was marked only in half of the cases where the NCSP code was completed. Five hospitals, which had marked the NCSP code, had no health insurance fund code at all.
- While the NCSP codes (except for one code) complied with the classification and there were few differences between the names and the classification, then different service list versions of the health insurance fund codes were used, and not all codes of additional resources could be found in the current list of health care services.

Established rules and instructions

- There are not always rules in place or they are not always sufficiently described to fill the epicrisis. It has not been indicated how to mark the code of EHAK, which name to use upon indicating the statistical type of diagnosis, and when it is a surgical operation.
- The rules for the summary notification of the epicrisis and statistics are not always compliant. For example, the epicrisis has one allowed external cause, but the summary notification has several ones.
- The analysis revealed a relatively large number of errors of which it could be possible to inform the person who has submitted the epicrisis by implementing the verification rules of the document/quality, for example, with regard to the number of principal diagnoses, the diagnoses not allowed to be a principal diagnosis, the codes allowed in case of external causes, the codes not allowed as a principal procedure in case of operations, the exclusion of the operation date from the case, etc.

Lists and classifications

- Not all published lists/characteristics function and require either revising or modifying of the user manual, for example, "scheduled admission_emergency" ("pöördumise plaanilisus_erakorralisus") or "ICD code" ("RHK kood").
- International/nationally used classifications, which are published online, are not identical. For example, the online version of the ICD-10 published on the website of the Ministry of Social Affairs, XML and the updated file differ in terms of some codes.

FURTHER ACTIVITIES

- Discussion with the EeHF about the feasibility of the functioning and adding documents or quality controls in the e-Health and the need to modify the instructions.
- Awareness of the need for integrating classifications in e-Health and health care providers data-entry-programmes.
- Providing feedback to the Ministry of Social Affairs as the body responsible for the ICD-10 about the differences in the ICD-10 versions, as well as to RESTA for development the pilot project of the statistics module.
- A working group has been set up by the EeHF to review the data composition and rules of outpatient epicrises, and the work is planned to be completed by autumn 2013. In order to harmonise the definitions of health statistics, another working group was established by the Ministry of Social Affairs in 2012, which has determined the definitions of inpatient and day care. The data set and definitions of both published (inpatient and day care) and unpublished (outpatient) aggregate statistical reports are planned to be revised according to the results of the pilot project of the statistical module and the decisions of the outpatient epicrisis working group.
- Organisation of regular trainings on the rules/instructions of the completion of the e-Health documents and use of the list/classifications.

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

- **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