

# Quality Improvement of Causes of Death Statistics by Automated Coding in Estonia, 2011

---

**Technical Implementation report, Grant agreement nr 10501.2009.002-2009.461**

## **Introduction**

The grant agreement between the European Commission and National Institute for Health Development was signed on 20.11.2009. Expected results of the action were (1) to produce and publish quality indicators for 1989-2009 along with analysis, (2) to implement IRIS as quality control tool on 2009's data and (3) to attend workshop on IRIS organized by Eurostat. As abovementioned workshop took place on 31<sup>st</sup> January – 4<sup>th</sup> February 2011 in Bonn and it was attended by NIHD staff other parts of this project were also shifted to 2011: quality indicators produced for 1989-2010 and IRIS (V4.0.28 25.01.11) was implemented on 2010's data already available in the first quarter of 2011.

## **I. Workshop on IRIS**

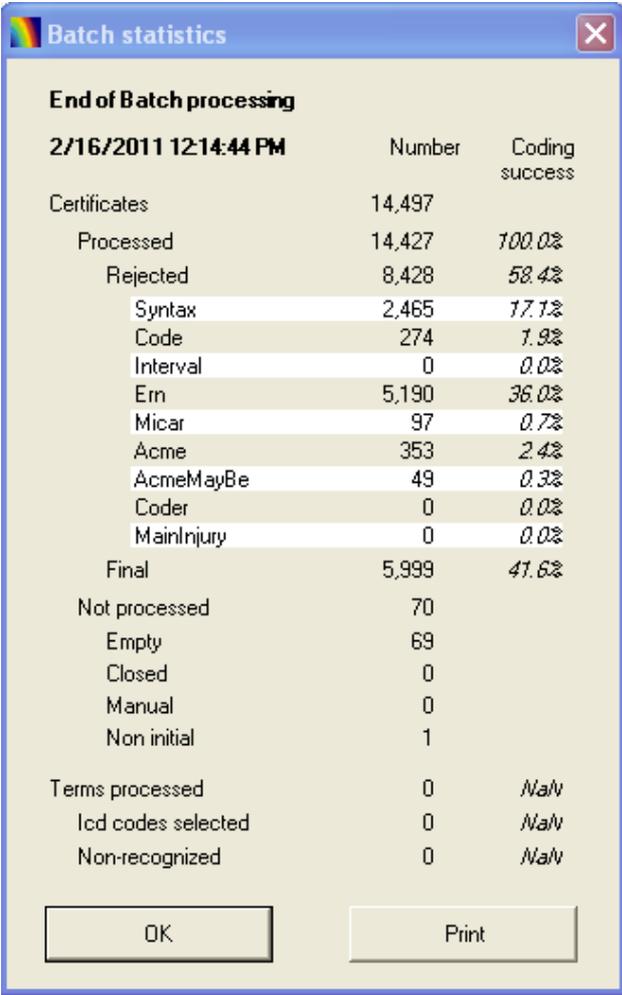
Workshop on IRIS held in Bonn on 31<sup>st</sup> January – 4<sup>th</sup> February 2011 under framework of another Eurostat-funded project was attended by Gleb Denissov, Head of Estonian Causes of Death registry. There were four trainers and enough time to discover and understand all features of IRIS. A half of closing day was designated for questions and answers, so everyone could catch up or even investigate context how this software was created, discuss specific issues and propose own solutions to its very authors. An advantage of this software is that it is developed by people working very long time in mortality coding, perfectly in position to understand users' needs. Workshop has given us competence to continue with work on IRIS, including activities under this grant agreement with European Commission.

## **II. Preparation of IRIS file and processing**

There were two extractions from database, the first one was made on 14th February 2011 and included the majority of 2010's cases - 14565, another one on 25th March 2011 to cover 1236 cases entered into database after the first extraction, totally 15801 cases. As ICD codes were stored in separate cells MS Excel function CONCATENATE were used to merge contents of those cells into lines and insert commas and blanks according to IRIS file format. Cells containing only commas and blanks were then deleted by Replace All feature in MS Excel. "Code only" column was filled with "1" to indicate that there is no diagnose text, default values in Ident table and line numbers were entered according to IRIS file format. Then data were copied into MS Access database provided with IRIS.

The first attempt to process the first extraction with IRIS resulted in a large number rejects – 58.4% of which 17.1% were syntax and 36.0% were ERN rejects (Fig.1).

Fig. 1. The first processing – large number of rejects



The reason for syntax reject were excessive commas and blanks, recognized by IRIS as errors, ERN rejects were largely due to so-called created codes required by the program and not used in manual coding. Excessive commas and blanks were deleted by Replace All feature in MS Access, last character “9” was added to simulate created codes to ICD codes I500, I219, I633, I634, I635, I639, G20, I64, I618, I613, I615, I619. Thus the number of rejects in the first extraction was reduced to 2908 or 20.1% (Figs 2 and 3). A few cases were not processed by IRIS as they lacked some data required by IRIS (e.g. exact date of birth).

Fig.2. Getting rid of syntax rejects

End of Batch processing		
2/16/2011 1:26:40 PM		
	Number	Coding success
Certificates	14,497	
Processed	14,428	100.0%
Rejected	6,461	44.8%
Syntax	0	0.0%
Code	320	2.2%
Interval	0	0.0%
Ern	5,566	38.6%
Micar	105	0.7%
Acme	393	2.7%
AcmeMayBe	77	0.5%
Coder	0	0.0%
MainInjury	0	0.0%
Final	7,967	55.2%
Not processed	69	
Empty	69	
Closed	0	
Manual	0	
Non initial	0	
Terms processed	0	NaN
Icd codes selected	0	NaN
Non-recognized	0	NaN

Fig 3. Reducing number of other rejects

End of Batch processing		
2/18/2011 12:55:34 PM		
	Number	Coding success
Certificates	6,461	
Processed	6,461	100.0%
Rejected	2,908	45.0%
Syntax	0	0.0%
Code	383	5.9%
Interval	0	0.0%
Em	1,924	29.8%
Micar	107	1.7%
Acme	404	6.3%
AcmeMaybe	90	1.4%
Coder	0	0.0%
MainInjury	0	0.0%
Final	3,553	55.0%
Not processed	0	
Empty	0	
Closed	0	
Manual	0	
Non initial	0	
Terms processed	0	NaN
Icd codes selected	0	NaN
Non-recognized	0	NaN

Then manually assigned underlying cause codes were extracted from database along with some variables not available in IRIS files (e.g. age in years) and imported into MS Access database. A query was designed and run in MS Access to find out cases with disagreement between IRIS and manually assigned underlying cause of death. There were 1405 such cases in the first extraction and 125 in the second one, totally 1530. It is worth noting that sometimes IRIS is able to propose underlying cause for rejected cases, these cases were also included into analysis. Additionally, 408 rejects due to code were extracted into separate file along with “ToDoList” field of IRIS data file explaining the reason for reject.

### III. Coding analysis

#### ICD versions

Some ICD-10 updates were not really implemented in Estonia, IRIS has discovered many C80 codes, which category was split into C800 and C809 in 2010, and few C141 codes which was removed by 1995's corrigenda, however is present in Estonian ICD-10 books and software. On the other hand IRIS did not accept J09 – the code for avian influenza introduced with 2005's ICD-10 update.

#### External causes

External causes of death are known to be difficult to ACS-s, however there are simple cases that could be processed automatically. For instance, all certificates with T510 - toxic effect of alcohol in line Ia and X45 – accidental poisoning with alcohol in line Ie were rejected as T510 is not valid for underlying cause of death and the program could not link it to X45, having applied instead Rule 2. Suicides by hanging are reported in a similar manner - T71 (asphyxia as nature of injury) due to X70 (suicide by hanging). In this case IRIS has inserted W84 (asphyxia as external cause) into the sequence and rejected the record as there is “no ERN for ICD code W84”.

#### General Principle as a problem

Some sequences reported on death certificates were not found in General Principle tables, for instance peritonitis K650 due to thyroid cancer C73 and IRIS selected K650 as underlying cause of death in this case. However this sequence could be accepted as cancer is a very serious disease and peritonitis is more like a symptom. Other sequences not accepted by IRIS were I839->I260, G951->J851 and I714->I770.

#### Unjustified replacements by MICAR

Sometimes MICAR, a component of MMDS and IRIS software replaces specific conditions with less specific ones, for instance A46, which is a specific bacterial infection was replaced with B948 – sequelae of bacterial infection and the latter code was selected as underlying cause of death. In a similar way G002 (specific meningitis) was replaced with G09 (a sequelae code). However, the opposite approach is also detected: I749 and I829 (thrombosis of unspecified arteries and veins respectively) were replaced with I219 (acute myocardial infarction) and I351 (unspecified aortic valve insufficiency) was replaced with I061 (rheumatic aortic valve insufficiency). A separate issue is state of drunkenness F100, which is replaced by MICAR with F101 – harmful use of alcohol.

### Picking underlying cause from Part II

Sometimes there is a valid sequence in Part I and a number of conditions in Part II, which according to certifier have not caused death directly. For instance, there was a certificate with cerebral infarction due to hypertension in Part I, diabetes and Z955 - presence of heart implant in Part II. Z955 was replaced by IRIS with ischemic heart disease I259, which was selected as underlying cause of death (Fig 4). However certifier did not report any problem with heart implant or ischemic heart disease in this case. Another example was pneumonia in Part I, “HIV-positive” and hepatitis B in Part II. IRIS has selected B24 – AIDS as underlying cause of death, which was clearly not intended by certifier. This particular case was certified by a forensic medic after autopsy.

Fig 4. “Presence of heart implant” in Part II converted to I259

The screenshot shows a software interface for medical coding. It features a table with columns for 'Interval', 'Code only', 'ICD-10 codes', and 'Line coded'. The table contains several rows, with the following ICD-10 codes and line coded status:

Interval	Code only	ICD-10 codes	Line coded
	<input type="checkbox"/>		
	<input checked="" type="checkbox"/>	I6399	«
	<input checked="" type="checkbox"/>	I110	«
	<input type="checkbox"/>		
	<input type="checkbox"/>		
	<input checked="" type="checkbox"/>	E118, Z955	«

Below the table, there is a 'Coding' section with the following elements:

- Underlying cause:** I259
- Automatic:**
- Final:**
- Coder reject:**
- Acme codes:** I639/I110\*E118 I259

### When artificial intellect fails

There were cases difficult for formal applications of rules, for instance there was a case with several circulatory conditions and G412 – state of epilepsy in between. After trying to establish links between these conditions IRIS applied Rule 2 and selected cerebral infarction I635 reported in line Ia as underlying causes of death (Fig. 5). However this was not right decision as epileptic symptoms are likely to be due to brain damage caused by vascular diseases. When G412 was removed IRIS linked cerebral infarction to hypertension and selected I110 – hypertensive heart disease with congestive heart failure as underlying cause of death.

Fig.5 Circulatory conditions and epilepsy

Code only	ICD-10 codes	Line coded
<input type="checkbox"/>		
<input checked="" type="checkbox"/>	I6359	«
<input checked="" type="checkbox"/>	G412	«
<input checked="" type="checkbox"/>	I119	«
<input type="checkbox"/>		
<input checked="" type="checkbox"/>	I48, I5009	«

Underlying cause:   
 Automatic:

```

Explain window
Iris  Micar  Acme
Acme .AMS file:
2010AL00000101 I635/G412/I119*I48 I500
2010AL00000102 Intentional Reject = '' Place of Inj = ''
2010AL00000103 Is I635 due to I119? YES
2010AL00000104 Is G412 due to I119? NO
2010AL00000105 No TUC by GenPrin - Apply Rule1
2010AL00000106 Is I635 due to G412? NO
2010AL00000107 Select TUC = I635 - Rule2
2010AL00000108 MANUAL UC:  ACME UC: I635
    
```

**Changes made**

As a result of coding comparison underlying cause of death codes were changed in 901 cases (5.7% of the sample), this number does not include changes to invalid codes (rejects due to code) as those changes involved not only underlying, but also multiple cause codes. The majority of changed underlying cause codes (492) were I11.- hypertensive heart disease, which means that cases with multiple circulatory conditions are somewhat difficult for manual coding. Statistics on the most frequently changed manually assigned codes (n>4) and most frequent resulting codes (n>14) are presented in Table 1.

Table 1. Changes to underlying cause coding

Changed codes	Codes to change to								
	Total	I110	I258	C800	I251	I635	I634	I132	I119
Total	901	382	67	58	40	40	38	20	15
I119	415	353	16		5	1	18	7	3
I110	77		30		17	1	11	12	1
C80	61			58					
I672	49				1	29	2		
I635	21	3	2		1	1			5
I259	14	1	7		5		1		
J159	14	1							
I509	13	6	1						
F019	11	1							3
I461	11				1		1		1
C97	10								
I258	10				4	1	2		
I519	10	2			3				
J448	9								
I132	7		4				1		
I251	6								
I260	5	2							

#### IV. Conclusions – IRIS as a quality control tool

1. Good to monitor implementation of ICD updates, to find out typing mistakes
2. Good for circulatory diseases
3. Bad for external causes of death
4. Much more user-friendly than ACME as data file can be easily edited/processed with MS Access and IRIS “Explain window” interface explains actions by all IRIS/MMDS components
5. Updates to ACME decision tables are needed regarding General Principle and linkage
6. Some code replacements by MICAR do not seem justified

## V. Quality indicators

The source of methodology to define quality indicators was “European Statistics on Causes of Death – COD” a document issued by Eurostat on 31<sup>st</sup> May 2007 to provide the Member States with guidelines to implement recommendations of the ad-hoc Task Force on Quality Control. With regard to some of these indicators the situation in Estonia did not change in 1989-2010. Mortality data refer to residence concept; data on foreign residents died in Estonia are collected, but not published or sent to international organizations and not used to calculate quality indicators. In case of autopsy medical death certificate is filled by forensic or hospital pathologist performed it, so 100% of autopsy results are used for certification and coding. Medical doctors are trained in certification before graduation as a part of obligatory training in pathological anatomy and forensic medicine. After graduation guidelines and occasional training courses in certification are provided by mortality data collection unit, frequently in co-operation with the Ministry of Social Affairs. Only medical doctors have been allowed to fill in medical death certificate since 1994, before a small number of deaths were certified by assistant physicians (feldscher).

Eight of recommended quality indicators are presented in Table 2. The most visible impact on data quality is attributed to recruitment of new (second) mortality coder in 2003, resulted in sharp decrease of certificates with one coded line (from 60.8 to 17.9%). Steadily growing share of persons died in hospital as well as decreasing share of cases certified by pathologists reflects changes in Estonian health care system. Introduction of ACME in 2005 and IRIS in 2010 as quality control tool did not have significant impact on these indicators. The most relevant indicator to evaluate impact of automated coding systems would be apparently percentage of changed underlying cause codes, whereas indicators proposed by TF on Quality Control are to assess quality of certification.

## VI. Publication of results

As it is requested by Grant Agreement quality indicators and analysis were published on NIHD website in English and Estonian:

[http://pxweb.tai.ee/esf/pxweb2008/Database\\_en/Population/04Deaths/SDinfo\\_en.htm#quality](http://pxweb.tai.ee/esf/pxweb2008/Database_en/Population/04Deaths/SDinfo_en.htm#quality)

<http://pxweb.tai.ee/esf/pxweb2008/Database/Rahvastik/04Surmad/SDinfo.htm#kvaliteet>

A presentation was made by Gleb Denissov on Technical Group on Causes of Death meeting in Luxembourg on 6<sup>th</sup> May 2011, slides and working document are available on Circa site.

Table 2. Quality indicators, per cent of cases

Year	Profession of certifier				Died in hospital	Just one line filled in (database)	R00-R99	R54	R98-99	X59 of all deaths	X59 of external causes
	Physician	Forensic	Pathologist	Other							
1989	63,6	12,6	20,8	2,9	37,3	n/a	0,5	0,3	n/a	n/a	n/a
1990	65,1	14,1	18,0	2,8	35,6	n/a	2,2	1,8	n/a	n/a	n/a
1991	69,8	14,1	13,5	2,6	36,7	n/a	3,3	2,9	n/a	n/a	n/a
1992	64,4	15,1	14,8	2,5	37,6	n/a	4,0	3,6	n/a	n/a	n/a
1993	64,0	17,1	15,0	2,5	38,0	n/a	4,5	4,0	n/a	n/a	n/a
1994	64,9	20,6	14,4	n/a	35,6	n/a	3,8	3,0	n/a	n/a	n/a
1995	65,4	18,6	15,9	n/a	36,7	n/a	4,4	3,3	n/a	n/a	n/a
1996	67,2	17,4	15,4	n/a	36,8	n/a	4,7	3,7	n/a	n/a	n/a
1997	67,0	17,7	15,4	n/a	38,2	n/a	4,5	3,5	0,9	0,1	1,1
1998	66,4	18,4	15,2	n/a	39,3	n/a	4,3	3,2	1,0	0,1	1,1
1999	65,5	19,0	15,5	n/a	40,1	n/a	4,5	3,3	1,1	0,1	1,0
2000	66,8	18,0	15,3	n/a	41,2	n/a	4,5	3,3	1,2	0,2	1,4
2001	66,8	19,4	13,8	n/a	41,6	60,3	3,9	2,8	1,1	0,4	3,3
2002	68,7	18,6	12,7	n/a	43,8	60,8	4,2	2,8	1,4	0,2	2,0
2003	69,2	17,5	13,3	n/a	45,5	17,9	4,0	2,7	1,2	0,2	2,0
2004	70,7	16,4	12,9	n/a	47,8	16,6	3,5	2,4	1,0	0,2	2,0
2005	72,5	15,7	11,8	n/a	48,7	17,6	3,6	2,6	1,0	0,2	1,8
2006	74,0	14,0	11,9	n/a	48,9	17,6	3,6	2,6	0,9	0,2	1,7
2007	73,0	15,3	11,7	n/a	50,9	18,3	3,9	2,6	1,2	0,1	1,6
2008	75,8	13,2	11,0	n/a	52,8	15,2	2,0	1,2	0,7	0,1	1,3
2009	77,8	11,2	11,0	n/a	53,9	15,2	1,8	1,0	0,8	0,1	0,9
2010	79,2	10,7	10,1	n/a	54,7	15,9	2,0	1,0	1,0	0,1	1,5