EUROPEAN COMMISSION - EUROSTAT

Production of Methods and Tools for Improving Causes of Deaths Statistics at Codification Level

Gérard Pavillon*, Lars Age Johansson**

Contract n° 9351005

Final Report

August 2001

*** CépiDc - Inserm**
Centre d'épidémiologie sur les Causes Médicales de décès
Institut National de la Santé et de la Recherche Médicale

****Statistics Sweden**

Contents

Introduction

This project is based on the recommendations of the Eurostat final report on Automated Coding in Member States [Eurostat99]. This report includes several recommendations and guidelines for the Automated Coding System design and use. In particular, the recommendation 29 specifies that "In order to produce more comparable statistics at the European and international level, Member states should cooperate". In this context 4 tasks of common interest were proposed.

Specifications of a cause of death automated coding system. This part studies the automated translation of the text of diseases and injuries mentioned on death certificates into ICD10 codes. This coding function is very complex and can influence greatly the selection of the underlying cause used for mortality statistics and for international comparison [Eurostat01][Jougla98]. This study represents the major work load of this report for several reasons. Firstly, previous reports have not described the coding function in sufficient detail, especially not as regards the language-dependant aspects. Secondly, in order to describe this function clearly, it was important to examine several national procedures of coding: the specifications of this report are based on the French, Swedish and US coding procedures. Thirdly, it was important to develop a coding prototype to test the feasibility of some of the specifications.

Knowledge base on bridge coding studies. Bridge coding studies are essential to assess the shifts in mortality statistics due to the change of classification version or of coding methods. Most of the time these studies are not disseminated at international level. In the context of the ICD10 implementation and of the generalisation of Automated coding systems (ACS), the availability of a knowledge base on bridge coding studies including method description, data and papers is essential. This part is based on an international survey on bridge coding studies. It also gives recommendation for the building of a knowledge base on bridge coding studies.

Test deck. Test deck is mentioned in the report on ACS as important means to evaluate ACS. Some test decks exists at national level mainly for evaluating ACS upgrades. Similarly, international test decks could be used to assess the international comparability of mortality data but up to date no international test decks exist and there have not been any discussions on

developing international test decks either. This part defines more precisely the purposes of international test decks and makes preliminary suggestions on how to develop international test decks.

Workshop on automated coding system. In the context of this project, a 3 days workshop has been organised. This objective was to convene experienced countries in the field of ACS in order to produce recommendations for countries planning to implement such systems.

These 4 parts are developed in this report.

Part 1 Specifications of a cause of death Automated Coding System

1. Introduction

This study concerns the automated coding of causes of death with the International Classification of Diseases, Tenth revision (ICD10). Automated Coding Systems (ACS) used in the field of mortality data production, usually perform 4 functions [Eurostat99]:

- Coding
- Editing
- Selection of the underlying cause
- Multiple cause coding

In this study, the 2 first functions are studied. These functions must be done before the selection of the underlying cause step. If the Editing function is a straightforward function that consists in checking the ICD10 codes assigned to a given death against the sex and age of the decedent, the Coding function is much more complex. It consists in assigning ICD10 codes to each cause reported on the death certificate. Then it appears several problems.

- The ICD10 code assignment widely depends on the language used. For instance, the expressions 'hjärtinfarkt" in Swedish, "Myocardial infraction" in English and "Infarctus du myocarde" in French will be mapped to the same ICD10 code "I21.9". This situation obliges each country to develop its own coding module.

- The diagnoses are frequently mixed with other indications that must be taken into account for the coding. For instance the mention "renal failure (2 months) due to diabetes" includes 2 diseases (Renal failure and Diabetes), one time indication (2 months) and one causal relationship ("due to"). All these indications can be relevant both for the coding and the selection of the underlying cause. Thus they must be deciphered and coded.

- The ICD10 code assignment does not depend on the text of the specific diagnostic statement only. It must also take into account other information mentioned on the death certificate such as sex, age and time indications. For instance a Pelvic peritonitis is coded K65.0 for males and N73.5 for female. A non-traumatic subdural haemorrhage is coded

P528 for children and I620 otherwise. A myocardial infarction is coded I25.8 when reported with a duration of more than 4 weeks, I21.9 otherwise.

- The assignment of an ICD10 code also depends on the other diagnoses mentioned on the death certificate. For instance an haemorrhage can be coded T14.9 if there is an indication of violent death (mention of an external cause of death) on the same certificate or R58. The diagnosis relationships that determine the selection of a code can be very complex. For instance gangrene can influence the code of diabetes if and only if the diabetes is the cause of the gangrene. Diabetes alone is coded E149. But in the following certificate:

  a) Gangrene, diabetes

  b)

  the diabetes will be coded E14.5 (diabetes with peripheral circulatory complications), but in the following one:

  a) Gangrene, diabetes

  b) Arteriosclerosis

  the diabetes will be coded E14.9 since the arteriosclerosis is specified as the cause of the gangrene. Diabetes is not the only example and other conditions such as Aids, cancer, surgery or violent deaths oblige to take into account the relationships with all the others diagnoses mentioned on the death certificate to be coded correctly.

Since the coding function depends on the language used, each country must develop and/or adapt its own software to implement the coding function. The way this software is designed will widely determine the selection of the underlying cause and the international compatibility of the mortality statistics. In Europe, most of the non-English speaking countries using an automatic coding system developed their own software for the coding function while using the NCHS ACME system for the selection of the underlying cause of death. ACME is a very reliable system, but the underlying cause it selects will be influenced by the coding of the causes of death.

This leads to the conclusion that the Coding function must take into account the ICD10 coding rules, the ACME requirements and the language specificity. In this context, the first objective of this study is to formalise the requirements of the ICD code assignment and editing functions in order to produce a set of ICD10 codes compatible with the ACME coding system for the selection of the underlying cause. A second objective, is the description of a

language independent system, i.e. a system where the language aspects are isolated and can be expressed as parameters of a general system. This general system is specified and a prototype of the language dependant aspects with examples both in French and Swedish is developed.

2. Problem description

Causes of death are reported on a death certificate, which in most of the European countries, looks like the international form proposed by WHO [ICD92]:

CAUSE OF DEATH

Part I		Approximate interval between onset and death
Disease or condition directly leading to death*	a).....	.....
	due to (or as a consequence of)	
Antecedent causes Morbid conditions if any giving rise to the above cause, stating the underlying condition last	b).....	.....
	due to (or as a consequence of)	
	c).....	.....
	due to (or as a consequence of)	
	d).....	.....

Part II

Other significant conditions contributing to the death, but not related to the disease or condition causing it .....

** This does not mean the mode of dying, e.g. heart failure, respiratory failure. It means the disease, injury, or complication that caused death.*

This certificate includes several lines and 2 parts. The first part is devoted to the diseases or conditions directly leading to death and the second part to the other significant conditions that contribute to death. An example of properly completed death certificate is presented below:

I a) Bronchopneumonia
b) Chronic bronchitis
c)
d)

II Chronic myocarditis

This means that a "Bronchopneumonia" due to a "Chronic bronchitis" directly lead to death. The "Chronic myocarditis" mentioned in part II only contributed to death. In this example, the death certificate is quite well filled but very frequently several problem appears.

- The same cause can be expressed in many different ways. For instance the only code "I219", corresponding to "myocardial infarction", has been found under 93 different expressions in French death certificates and 96 in the Swedish ones. Sometimes there are several possible spelling for the same word (for instance in French, "cardio-respiratoire", "cardiorespiratoire" or cardio respiratoire" can be found).

- Several causes can be stated on the same line, separated or not by different signs (full stops, commas, hyphens, semicolons, sign +). In Swedish and in French, hyphens are ambiguous because they can be found as separators of compound words ("cardio-respiratoire") or as cause separators ("infarctus – diabète"). More difficult are the cases where two diagnoses are bound in the same expression. For instance in French "Cancer bronchique métastasé" includes 2 different causes: "Cancer bronchique" and "métastases". In other cases, conditions can be "factorised": in the English expression "vascular and renal failure", the failure is both renal and vascular.

- Relationships, causal or not, can be expressed by literal expressions and not just by the sequence on lines a, b, c, d in part I of the death certificate. These expressions will be called "connectors". Examples of connectors are "due to", "with", "and", "caused by" etc. The important thing is to make a distinction between causal and non-causal connectors. The connector "due to" for instance, is clearly causal. In the following certificate:

  I a) renal failure due to diabetes
  b)
  c)
  d)

  II Hypertension

  the diabetes is the cause of the renal failure and it should have been reported alone on the line b. The non-causal connector "and" is often used to associate 2 conditions on the same line without causal connection (e.g. "cancer of rectum and cardiac failure"). Some connectors can be causal or not depending on the diseases connected. This is the case with the conjunction "avec" in French.

- Different information can be mixed with conditions. This is the case with time and surgery indications that can take various forms in the text of the diagnoses. These pieces of

information are quite important because they may influence the coding. The problem is that they can take various forms. Time indications can be expressed as time interval (1week) or as dates (12 May 1987). Here are some examples:

Cerebrovascular disease since July
Breast cancer in 1985
Fracture of femur 1 week ago
Sepsis 7 months
Cardiac failure today

Indication of surgery can be stated separately or mixed together with diseases and/or time indications:

Choc cardiogénique post-opératoire
Prothèse de hanche
Artéritique pontage aorto-fémoral juin 98
Greffe cardiaque
Fracture du fémur (opération il y a 2 semaines)

In this context, the automated coding of information reported on the death certificate is not merely a mapping of discrete expressions into ICD10 codes. It requires a rather complex analysis of the text to separate the different types of information. This strongly depends on the language used and cannot be generalised. However it would be interesting to specify clearly the functions that such a system should perform and, when possible, to develop tools to perform these functions independent of the language.

3. System architecture

The automated coding of causes of death can be divided into 4 functions.

- **Text parsing.** This function aims at identifying the different categories of information relevant for coding: text of diagnoses, diagnoses connectors, time indications and surgery indications.

- **Diagnoses coding.** This function selects the ICD10 code(s) for each diagnosis mentioned on the death certificate. The diagnoses coding sometimes depends on certain conditions and the code selection takes into account pieces of information mentioned on the death certificate: sex, age, other diagnoses, time, surgery indications and other available information. Causal or non-causal condition connectors are also relevant to the coding. This function must save the respective place of the different elements (part of the certificate, line in part I, place in the line).

- **Editing.** Each ICD10 codes are checked according to sex and age. Consistency of time indications is also checked at this level.

- **Output standardisation.** ICD10 codes are delivered with the indication of their respective position on the certificate. This output can be used to select the underlying cause.

Several guidelines should be respected when developing an automated coding system.

- All the relevant information reported on the death certificate should be taken into account. In particular all the information necessary for the selection of the underlying cause should be kept.
- The system should be able to identify and to indicate ambiguous information.
- Editing should be as comprehensive as possible in order to increase the final quality of data.
- The output of the system should be compatible with the automated system used for the selection of the underlying cause.
- The interface should be as similar as possible to the death certificate.

- The system should run in interactive mode as well as in batch mode. The interactive mode is used to enter directly death certificates and the batch mode to code or recode set of death certificates already keyed in.

According to the functions defined in the previous chapter, the architecture of the system is synthesised in figure 3-1. Each function will be separated into general processes common to all languages and parameters dependent of a specific language. General processes will be transcribed into algorithms and parameters will be stored in a database.

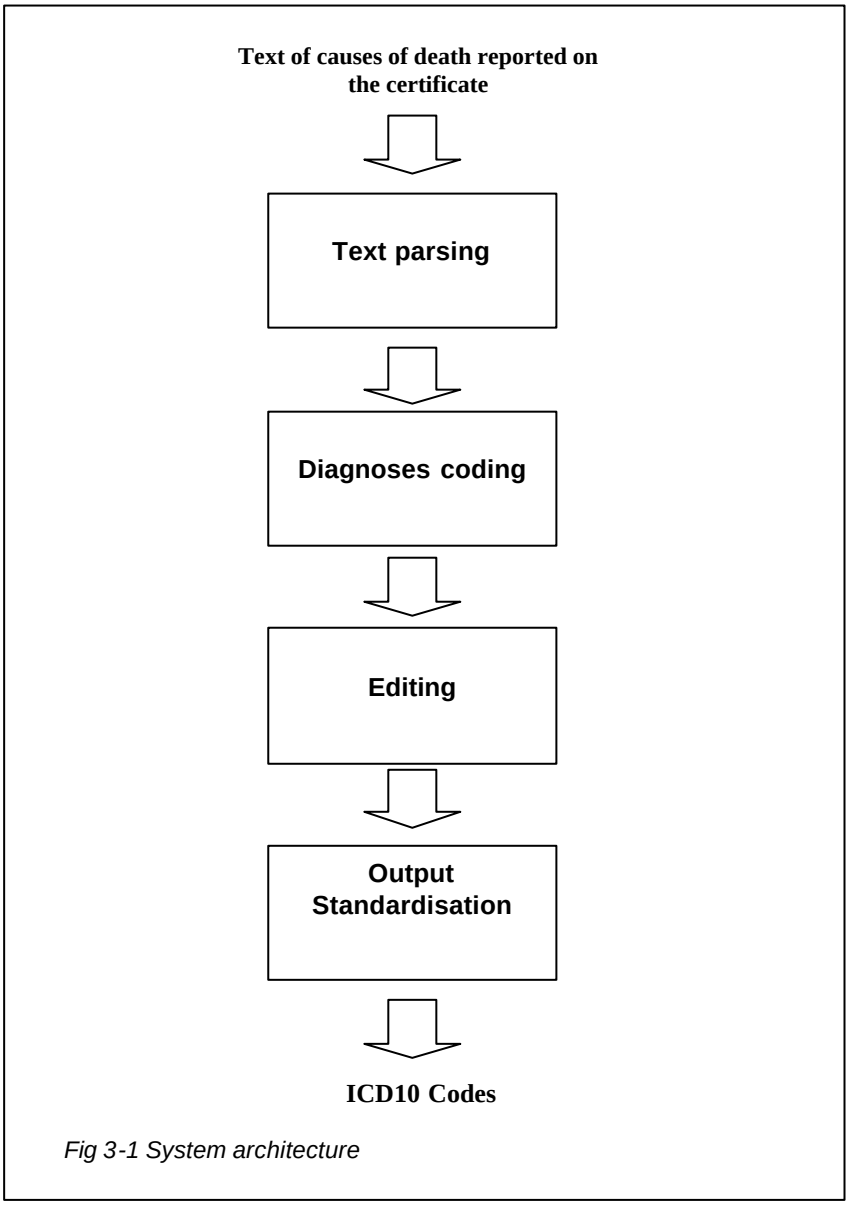


Fig 3-1 System architecture

4. System description

This chapter is an in-depth description of each of the functions presented in chapter 3.

4.1 Text parsing

The figure 4-1 gives a diagram of the sequence of text parsing operations. The rectangles (Coding, Syntax standardisation 1 etc.) represent the different steps of the processing and the ellipsis represent the tables (language dependant parameters) needed. In interactive mode, the text parsing can be launched immediately after the diagnoses entered on the same line have been keyed in, in bacth mode the text parsing is launched whan diagnoses are available on the record. Before each step, the system attempts to codes all the diagnoses mentioned on the line. This attempt succeeds when all the diagnoses mentioned on the line are coded, otherwise it fails. When it fails, a new step is applied in order to try to standardise the text. The 5 standardisation steps are listed below.

- Standardisation 1. This step performs the most common and systematic transformations on the text. For instance the word "Probable" in French that never influence the coding will be eliminated in this step. This step also includes standardisation of spelling such as, in Swedish, the substitution of "k" for "ch" in some words ("cholangio" would be changed in "kolangio"). This aim at reducing the volume of the index.
- Separators substitution. This step deals with separators expressed with word such as "due to", "caused by" or "with". These separators are categorised as causal or non-causal.
- Time intervals. The time interval indications are transformed into duration according to the time of death.
- Surgery indications. Mentions of surgery are detected and standardised.
- Standardisation 2. During this last step a new attempt is done to eliminate words that do not influence systematically the coding. For instance the words "left" or "right" are eliminated here.

The steps of standardisation 1 and 2 aim at reducing the volume of the index by standardising the different expressions of the same diagnosis. Two options are possible: 1) all the different forms of the same diagnosis are entered in the index and the standardisation steps are minimum or 2) the standardisation steps are parameterised to convert the different forms of the diagnosis into a "canonical" form. The choice here depends mainly of the language; For instance in French there are about 100 different expressions for the ICD code I25.1

"Atherosclerotic heart disease" and it is valuable to enter all these form in the index. In Swedish, a preliminary test comprising 25,000 records contained 400 different expressions for the only expression "Coronary atherosclerosis" which is a subpart of the I25.1 code. In this case the standardisation steps help to reduce the index size.

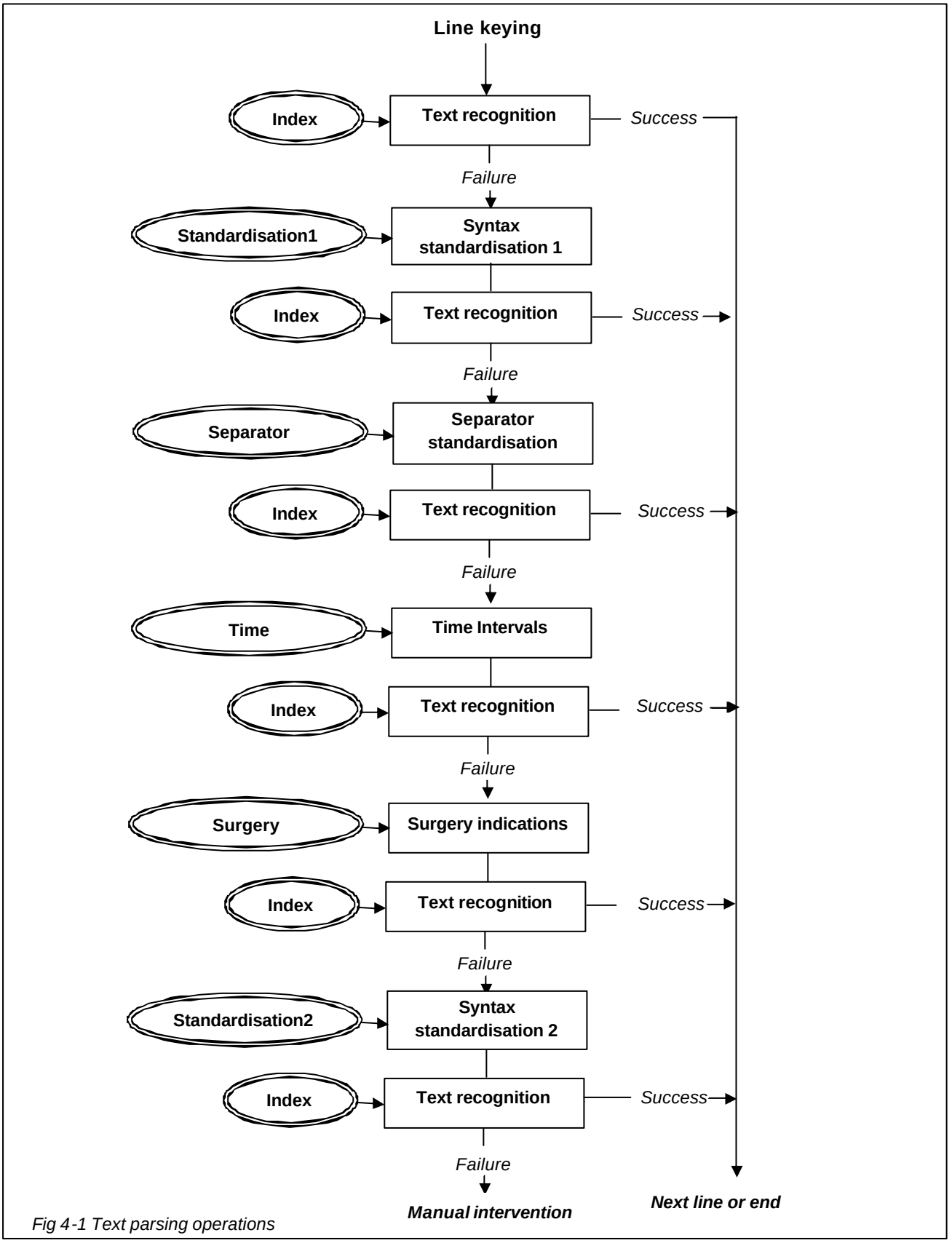


Fig 4-1 Text parsing operations

The next section clarifies the way causes of death are entered in the system. The following sections of this chapter explain in detail each step of the text parsing function. Each function is described and the tables used are listed. In addition examples are given. The full description of tables is provided in section 4.5.

4.1.1 Text keying

As the code of a diagnosis may depend on other diagnoses mentioned on the death certificate, all the reported diagnoses will need to be keyed in. The death certificate is keyed in line by line. Diagnoses mentioned on a line are captured together. The line is ended with a carriage return. Different diagnosis are separated by commas when there is no doubt, for instance:

Hypertension diabetes

will be keyed in as:

Hypertension, diabetes

However this work already supposes knowledge of the ICD coding and should not be done by typists but by experienced coders or by typists specially trained in medical terminology. Moreover, commas are sometimes part of a diagnoses when quantities are involved (e.g. "...duration 1,5 day" or "alpha 1,2").

When diagnoses are separated by expressions such as "with", "due to" etc., these expression are captured as they are mentioned. The blank is sometimes used as a separator between the diagnoses mentioned on the same line, but the automatic management of this separator would greatly complicate the program. Therefore, in this system, blank, as a diagnosis separator is manually replaced by a comma.

4.1.2 Text recognition

Description

This function checks that each text enclosed between 2 diagnosis separators can be coded. A diagnosis separator can be:

- the start of line
- the end of line
- a comma
- a standardised causal or non-causal separator
- a standardised mention of surgery

The output of this function is binary: if all the text can be coded according to the Index table, it succeeds otherwise it fails.

Table

The Index table (see description in 4.5) is used to search for the text of causes. If one or several records are found the function succeeds otherwise it fails. ICD10 codes provided by the Index table are not considered at this level.

4.1.3 Syntax standardisation 1

Description

This step performs systematic transformations on the text of causes. First, multiple blanks in the line are replaced by a simple blank. Then the text is modified according to the filters provided by the Standardisation1 table.

Table

The Standardisation1 table has the following structure (see 4.5):

Attribute	Type (length)
Key	Text (6)
OrderNum	Integer
FilterIn	Text (100)
FilterOut	Text (100)
Action	Text (10)
Likelihood	Text (1)
DateTo	Date - dd/mm/aaaa
DateFrom	Date - dd/mm/aaaa
Comment	Text (100)

The fields "FilterIn" and "FilterOut" are filters used to modify the line content: if the input text matches with the input filter, the text is modified according to the output filter. The other fields are common to most of the tables and are explained in section 4.5.1.

The syntax used for FilterIn and FilterOut is shown below.

A **filter** is a sequence of any combination of strings and symbols.

Strings are any combination of characters others than brackets [ ];

Symbols are letters enclosed with brackets [ ]. Each symbol stands for a specific type of string:

Symbol	stands for
c	one character
S	a string (a set of characters)
l	one letter a-z, A-Z
L	a set of letters a-z, A-Z
d	one digit 0-9
N	an integer number (series of digit 0-9)

Symbols can be individualised by a digit, for instance the filter [S1] [S2] makes a difference between 2 different strings S1 and S2 separated by a blank.

If the input filter (FilterIn) match to the line content, the output filter (FilterOut) is applied to the line and results in a new line.

Examples

Example of filters

1 Input filter: [S](probable)
Output filter: [S]
line content: infarction (probable)
new line: infarction

2 Input filter: [S1] [l][d][d].[d][S2]
Output filter: [S1][S2]
line content: Diabetes E42.1, infarction
new line: Diabetes, infarction

Example of table Standardization1 (only the relevant fields are shown in this example):

Key	Order	OrigString	NewString
10	10	[S1].[S2]	[S1],[S2]
10	20	[S1];[S2]	[S1],[S2]
10	30	[S1]&[S2]	[S1],[S2]
10	40	[S1]+[S2]	[S1],[S2]
20	10	cancer of[S1]	[S1]cancer
20	20	malignant tumour of[S1]	[S1]cancer
20	30	malignant[S1]tumour	[S1]cancer
30	10	metastases of [S1]and[S2]	[S1] metastasis, [S2] metastasis

The "Key" and "Order" fields aim at ordering and grouping the filters (see Table description in 4.5). The filters devoted to the same job are grouped with the same Key and differentiated with the "Order" number. Key and Order are numbered with an increment of 10 to allow future updates.

Filters 10 (grouped with Key=10), aim at standardising the diagnoses separators : when ".;&=" are used they are replaced by a comma.

Filters 20 standardise the expression of cancers in order to reduce the volume of the index: the diagnosis "Cancer of liver", "Malignant tumour of liver" and "malignant liver tumour" will be mapped to the unique expression "liver cancer".

Filter 30 transforms the expression "metastases of x and y" in 2 diagnosis separated by a comma "x metastasis, y metastasis".

4.1.4 Causal and non-causal separator standardisation

Description

This step aims at separating the diagnoses including separators others than the comma: causal connection ("due to", "caused by") or non-causal connection ("with"). These separators are identified and replaced by one of the 3 standard separators:

[SepC<] right separator: the condition on the left is due to the condition on the right

[SepC>] causal left separator: the condition on the left causes to the condition on the right

[SepE] non-causal separator: the 2 concomitant conditions are related

Table

The identification of the different kind of separators is made with the Separators table (see 4.5) described below:

Attribute	Type (length)
Key	Text (6)
OrderNum	Integer
OrigSep	Text (100)
Type	Text (1)
Action	Text (10)
Likelihood	Text (1)
DateTo	Date - dd/mm/aaaa
DateFrom	Date - dd/mm/aaaa
Comment	Text (100)

The fields "OrigSep" is the text of the separator to be standardised (e.g. "due to"). "Type" is the nature of the separator: non-causal, causal left or causal right depending on the direction of the causality. When it is mentioned, "Action" is the name of a specific procedure to be performed.

Example

Example of Separators table (only the relevant fields are shown in this example):

Key	Order	OrigSep	Type	Action	likelihood
10	10	due to	3		C
20	10	giving rise to	2		C
30	10	with	1		C

"Due to" is a causal connector where the right part causes the left part. "giving rise to " is causal connector where the right part causes the left part. "with" is a non-causal connector.

4.1.5 Time interval

Description

Time indications mentioned with certain diagnoses or in the column on the right of the death certificate (Approximate interval between onset and death) are mapped into time intervals from death. A time indication mentioned with a given diagnosis is associated with this diagnosis only. When time indications are mentioned in the column on the right of the death certificate, it is associated with the diagnosis reported on the line. If several diagnoses are mentioned on the line, or if there are conflicts between time indications, the death certificate is pointed out.

Time indications can be time intervals (e.g. "2 weeks ago") or absolute time (e.g. "Stroke on 21/4/1999"). In this case the time interval from death is computed. In both cases, this step standardises the time indication under the form of a specific syntax.

Table

Time indications are detected with the Time interval table (see 4.5.5) with the following description:

Attribute	Type (length)
Key	Text (6)
OrderNum	Integer
TimeExp	Text (100)
Action	Text (10)
Likelihood	Text (1)
DateTo	Date - dd/mm/aaaa
DateFrom	Date - dd/mm/aaaa
Comment	Text (100)

"TimeExp" is a string with the same filter syntax defined in the Standardisation1 table (see 4.2.2). When this filter matches the input text, the specific action mentioned in the field "Action" is performed. This action return a standard time indicator of the form [Int*iu*] where Int is a constant that identify an interval, i is an integer giving the elapsed interval of time and u is a time unit (m: minutes, h: hours, D: days, W: weeks; M: months; Y: years).

Example

part of a TimeIntervals table (only the relevant fields are shown in this example):

Key	Order	TimeExp	Action	LikelyHood
10	10	[s1][n] weeks ago[s2]	W	C
10	20	[s1][n] weeks ago	W	C
20	10	[s1]([n] weeks)[s2]	W	C
20	20	[s1]([n] weeks)	W	C
30	10	[s1]on [n]/[n]/[n][s2]	D1	C

The four first lines correspond to the indication of a time interval expressed in weeks. They all call the action "W" which will return an expression of the form [Int*n*W]. The last line expresses an absolute time indication, which will be transformed into a time interval of the form [Int *iu*].

4.1.6 Surgery indications

Description

Management of surgery is similar to separator management.

Table

The table Surgery (see 4.5.6) is as below:

Attribute	Type (length)
Key	Text (6)
OrderNum	Integer
SurgExp	Text (100)
Direction	Text (1)
Action	Text (10)
Likelihood	Text (1)
DateTo	Date - dd/mm/aaaa
DateFrom	Date - dd/mm/aaaa
Comment	Text (100)

"SurgExp" is a string that indicates the mentioned of a surgery act (e.g. "suite à opération"). This expression is associated with a direction indicating if the expression affects the left or right diagnosis. When the surgery expression is found in a text, it is standardised under the form [<S] or [>S] depending whether the direction is left or right.

Example

Here are 2 records of a surgery table:

Key	Order	SurgExp	Direction	Action	likelihood
10	10	(operation	L		P
10	20	operation of	R		C

The surgery expression is just mentioned in the field SurgExp and the direction is specified as "L" if the expression affects the diagnosis on the left or "R" if it affects the diagnoses on the left side. The program standardises the expressions under the form [<S] or [>S] respectively.

4.1.7 Syntax standardisation 2

This step is identical to the first step of syntax standardisation 1 excepted that it occurs at the end of the text parsing function. The Standardisation2 table is used in this step (see 4.5).

4.2 Diagnoses coding

The goal of the diagnoses coding is to select the appropriate codes for all the diagnoses reported on the death certificate. Diagnosis coding is launched when text parsing ended with success, i.e. when the text of causes of death mentioned on the certificate has been keyed in and when all the parts of this text have been identified as diagnoses, separators, time intervals or surgery indications.

Diagnosis coding is made up of 3 steps as shown in figure 4-2

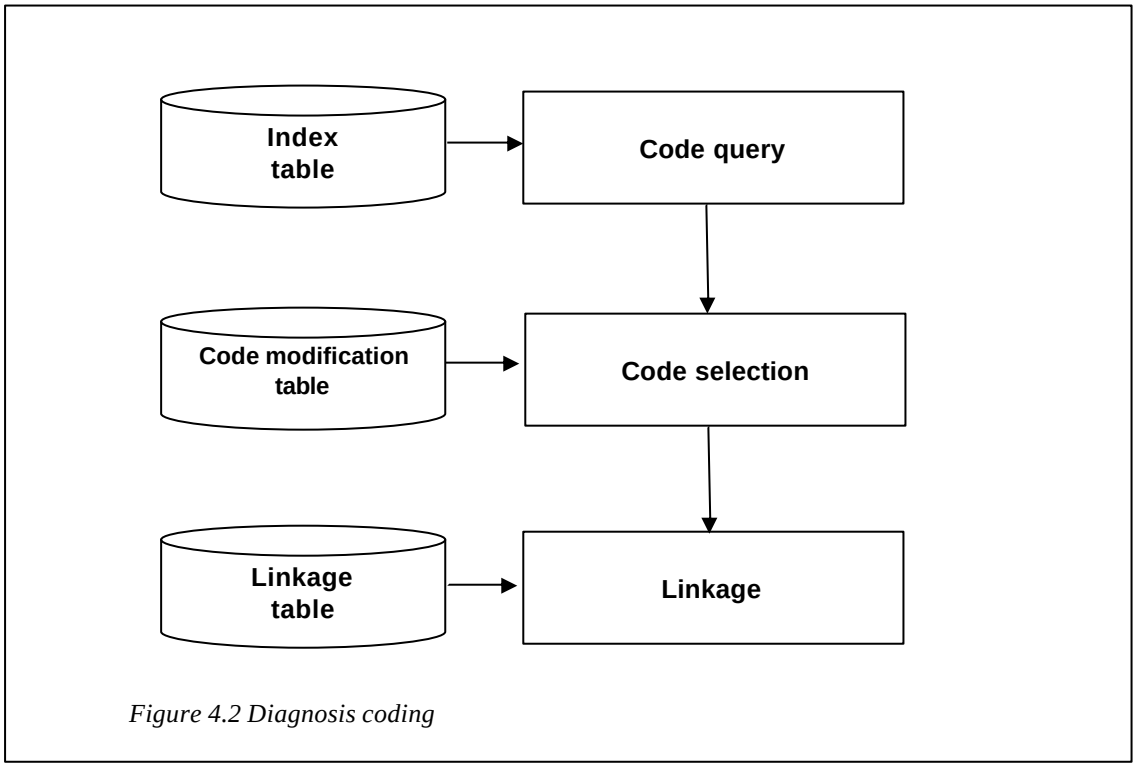


Figure 4.2 Diagnosis coding

The Code query step provides a tentative ICD10 code for the diagnosis. The code selection step can modify this code according to sex, age or duration of the disease. Finally the Linkage step can modify the code according to the other codes mentioned on the death certificate. The following chapter describes these 3 steps in detail.

4.2.1 Code Query

Description

Code Query maps the diagnosis text to an ICD10 code.

Table

The Code Query step uses the Index table to search for the ICD10 code(s) associated with the diagnoses reported on the death certificate. Each diagnosis should have one and only one entry in the Index table. Text parsing has already checked that there is at least one entry for the diagnosis. The field IcdCode1 is mandatory, it is a tentative ICD10 code that may be modified in the next step.

Examples

In the example below, a Subdural haemorrhage is associated with ICD10 code I620.

Index table	
Key	
OrderNum	
Diagnosis	Subdural haemorrhage
IcdCode1	I620
IcdCode2	
Likelihood	C

The field IcdCode2 is optional, it is used in the case where 2 codes have to be associated with a single diagnosis. For instance:

Index table	
Key	
OrderNum	
Diagnosis	Metastatic Kaposi's of brain
IcdCode1	C469
IcdCode2	C793
Likelihood	C

the Metastatic Kaposi's of brain corresponds in fact to 2 different diagnoses, the Kaposi's sarcoma (C469) and the brain metastases (C793).

4.2.2 Code selection

Description

The Code selection step modifies, when necessary, the codes provided by the Code query step.

Table

This function is based on the Code modification table. The modifications are done according to sex, age and/or disease duration mentioned on the death certificate.

Examples

In the following example the code K650 (Acute peritonitis) is modified into code N73.5 (Female pelvic peritonitis) for female only.

Code modification table	
IcdCode	K650
OrderNum	
Sex	2
Age	
Time	
NewCode	N735
Likelihood	C

In this example, code I62.0 (non-traumatic subdural haemorrhage) id modified into code P52.8 for children under 13 months.

Code modification table	
IcdCode	I620
OrderNum	
Sex	
Age	0-12M
Time	
NewCode	P528
Likelihood	C

In this example, the code I21.9 (Acute myocardial infarction) is modified into I25.8 (Chronic ischaemic heart disease) when the interval from death is over 28 days.

Code modification table	
IcdCode	I219
OrderNum	
Sex	
Age	
Time	28-J
NewCode	I258
Likelihood	C

In this last example, the code I51.9 (heart disease) is modified into Q24.9 (congenital malformation of heart) when the duration is equal to the age of the decedent. This modification is not certain (likelihood = 'P') thus the operator will be prompted to confirm.

Code modification table	
IcdCode	I519
OrderNum	
Sex	
Age	
Time	Age
NewCode	Q249
Likelihood	P

4.2.3 Linkage

Description

This step is complex since it involves all the diagnoses reported on the death certificate. Two main operations are realised.

- Code modification. A code is modified depending on the other codes reported. For instance a cancer of lung can be coded as primary (C349) or secondary (C780). It is coded as secondary if it is mentioned alone or with another cancer which is not a common site of metastasis (ICD10, vol. 2, page 76). Otherwise, it is coded primary. Many codes can be modified in this way: all the cancers that are common sites of metastases, surgery complications, condition that can be both due to a external cause or not (e.g. Haemorrhage)...

- Code addition. A new code is added depending on the other codes. For example, the ICD10 obliges to code the disease that motivated a surgery instead of the surgery itself. However, the disease implicated is not always mentioned or is merely suggested in the name of the operation (e.g. Gastrectomy, heart surgery...). Thus the code of the disease should be added to the code of the surgery (e.g. stomach disease, heart disease). According to ICD10, this addition is not needed, but ICD10 only deals with the underlying cause coding, not with multiple cause. In addition, ACME, the NCHS automated coding system for the underlying cause selection, requires the 2 codes.

Linkages also depends on the relationships between codes. In the previous example on lung cancer, the modification of the code does not require a specific relationship between lung cancer and the other cancer. If a stomach cancer is mentioned together with a lung cancer, the lung cancer will be considered as secondary whatever the relationship is and wherever they are mentioned on the death certificate (part I or part II, on the same line, on 2 different lines...). On the contrary, the relationship between diabetes (E14.9) and gangrene (R02) must be causal to modify diabetes into diabetes with peripheral circulatory complications (E14.5).

Linkages may be dependent on the time elapsed between the conditions. For instance, heart failure can be coded I50.9 or I97.1 if it is a complication of cardiac surgery. However, if the surgery occurred after the heart failure or more than 4 weeks before, heart failure will not be considered as a complication of the surgery.

Tables

This step uses 2 tables: the linkage table linkage table and the group table.

1) The **linkage table** allows to define the conditions of linkage and the resulting code. Each record defines the linkage condition for one ICD10 code. The table is described below.

Attribute	Type (length)
IcdCode	Text (6)
OrderNum	Integer
IcdCodePos	
LnkCodes1	Text (100)
LnkRel1	Text (2)
Time	text(100)
LnkCodePos1	
LnkCodes2	Text (100)
LnkRel2	Text (2)
LnkCodePos2	
ResultCode1	Text (6)
ResultCode2	Text (6)
IcdCodeAction	Text (1)
LnkAction	Text(1)
Likelihood	Text (1)
DateTo	Date - dd/mm/aaaa
DateFrom	Date - dd/mm/aaaa
Comment	Text (100)

IcdCode and OrderNum are the primary key. IcdCode is a code given by the Index and Code modification tables. OrderNum allows to group several record for the same ICD code.

The field LnkAction defines if all the possible linkages of a group or only one linkage are to be performed.

LnkCode1 define the set of ICD codes linked with IcdCode. The syntax allows expressing isolated codes and intervals. For instance "C73, C760-C779" stands for codes C73, C760, C761, C762, ... , C779. The minus sign can also be used to excludes a subset of codes: "(C000-C969) –(C760-C789)" stands for the set of codes "C000-C759, C810-C969 ". Also symbols already defined in the Code Groups table can use (see Code Group table below).

LnkRel1 defines the condition of relationship that must exist between the IcdCode and the LnkCode.

LnkCode2 and LnkRel2 define a second set of link codes.

ResultCode is the resulting code if the linkage conditions are satisfied.

IcdCodeAction tells if the linkage is a code modification or a code addition. In this last case, it specifies how to add the new code.
Time is the condition of delay that must exists between the IcdCode and the linked code.

2) The Code groups table. This table allows to define only once the set of codes that are frequently use to define the linked codes in the Linkage table (fields LnkCode1 et LnkCode2). A name can be mapped to the set of codes and used afterward. The table is as follow.

Attribute	Type (length)	Key/index	Comments
Name	Text (20)	Primary key	Symbolic name of the code group
CodeGroup	Text (100)		Expression of the Icd10 codes belonging to the group
DateTo	Date - dd/mm/aaaa		Time until when the record has been used
DateFrom	Date - dd/mm/aaaa		Time from when the record has been used
Comment	Text (100)		Free text concerning the record

Name is the name of the set of codes. For instance, "MPTumor" could refer to Malignant primitive tumour. CodeGroup is the definition of the group. For instance, the expression "C000-C759, C810-C969, C97" defines the malignant tumours.

Algorithm

Each code is examined for linkage. Linkages involving causal relationships are realised first. This implies that, in part I, codes are examined bottom-up from the last line up to line a. If a code is modified, new linkages may be applied. The process is followed until there is no more linkage. An error in linkage table can lead to an infinite loop. The easier way to prevent this is to limit the number of linkages and to keep track of the records for debugging.

Examples

Case 1

I a) Stomach cancer C169
b) Lung cancer C349
c)
d)

II

Final codes: Lung cancer C780
Stomach cancer unchanged

Comment: The stomach cancer is nor a common site of metastases. In this condition the lung cancer is considered as secondary and the code is C780. There is no relationship condition. The symbol MpTumour and Csm are defined in the group table. The expression MpTumour-Csm means "all the malignant tumour excepted the common site of metastases".

Linkage table	
IcdCode	C349
OrderNum	1
ExclusiveGroup	
LnkCodes1	MpTumor - Csm
LnkRel1	
Time	
LnkCodes2	
LnkRel2	
ResultCode1	C798
IcdCodeAction	R
LinkageAction	
Likelihood	

The 2 following records define the symbols MpTumor and Csm in the Code Group table (for the complete definition of CSM, refer to ICD10, vol 2, 4.2.3, G)

Group table	
Name	MpTumor
OrderNum	1
CodeGroupExp	C000-C759, C810-C969, C97

Group table	
Name	CSM
OrderNum	1
CodeGroupExp	C400-C419, C690-C729 ...

case 2

Certificate

I a) CVA I64
b) Thrombosis I829
c) Atherosclerotic heart disease I251
d)

II

Final codes: unchanged

Comment: The thrombosis can be linked with Cerebrovacular accident and also with Atherosclerotic heart disease when it is reported alone with one of these conditions. However, if it is reported with these two conditions, there is no linkage. The following records show how to manage this situation. The first record identifies the situation where the thrombosis is reported with both the Atherosclerotic heart disease and the CVA. In this case the same code is kept and the field LnkAction stops the subsequent linkages for the code I829.

Linkage table	
IcdCode	I829
OrderNum	1
LnkCodes1	I251
LnkRel1	**due to**
LnkCodes2	I64
LnkRel2	Cause of
ResultCode1	I829
LnkAction	C

Linkage table	
IcdCode	I829
OrderNum	2
LnkCodes1	I251
LnkRel1	due to
ResultCode1	I240

Linkage table	
IcdCode	I829
OrderNum	3
LnkCodes1	I64
LnkRel1	cause of
ResultCode	I633
IcdCodeAction	R

case 3

Certificate

I	a) Pulmonary embolism	I269	T817
	b) Open heart surgery	ZZ1	Y839
	c)		
	d)		

II

Final codes: Addition of the code I519, heart disease

Comment: This certificate brings 3 problems. 1) there is no specific code in the ICD10 for Open-heart surgery. The code tentative ZZ1 is made up. 2) The pulmonary embolism is a complication of the surgery and the code I269 should be change into T817 and the open heart surgery with complications should be coded Y839 3) According to ICD10 (Vol2, 4.2.6), the disease that imposed the surgery should be added. This disease will be selected as the underlying cause.

Linkage table	
IcdCode	ZZ1
OrderNum	1
LnkCodes1	
LnkRel1	last used line
ResultCode1	Y839
ResultCode2	I519
Likelihood	

Linkage table	
IcdCode	I269
OrderNum	1
ExclusiveGroup	
LnkCodes1	Y830-Y849
LnkRel1	due to
ResultCode1	T817
Likelihood	

case 4

Certificate

I a) Gangrene, Diabetes R02, E149
b) Atherosclerosis I709
c)
d)

II

Final codes: Atherosclerosis I702
Other codes unchanged

Comment: The problem is that both atherosclerosis and diabetes can link with gangrene. Since Atherosclerosis is in a causal relationship with gangrene and will link with it, diabetes should not. The first record of the linkage table links Atherosclerosis and Gangrene resulting in Atherosclerotic gangrene I702. The second record prevents Diabetes from linking with Gangrene in presence of I702. and stops the linkage of Diabetes in this group. The third record links Diabetes and Gangrene when there is no Atherosclerotic gangrene.

Linkage table	
IcdCode	I709
OrderNum	1
LnkCodes1	R02
LnkRel1	
LnkCodes2	
LnkRel2	
ResultCode	I702
IcdCodeAction	Y

Linkage table	
IcdCode	E149
OrderNum	1
LnkCodes1	R02
LnkRel1	
LnkCodes2	I702
LnkRel2	
ResultCode	E149
LnkAction	G

Linkage table	
IcdCode	E149
OrderNum	1
LnkCodes1	R02
LnkRel1	
LnkCodes2	
LnkRel2	
ResultCode	E145

Case 5

Certificate

I a) Heart failure	2 weeks	I509
II Surgery for fracture of femur	3 weeks	Z924, S729

Final codes: Heart failure I978

Comment: The heart failure is first coded to I509. But if the surgery was performed before the heart failure occured, and there were less than four weeks between the surgery and the death, then the heart failure should be coded to I978. This is what is expressed in the following record (the time condition specified is the time elapsed between the ICD code I50.9 and the linked code Z92.4).

Linkage table	
IcdCode	I509
OrderNum	
IcdCodePos	
LnkCodes1	Z924, Y830-Y849
LnkRel1	
Time	0-4W
LnkCodePos1	
LnkCodes2	
LnkRel2	
LnkCodePos2	
ResultCode1	I978
ResultCode2	
IcdCodeAction	
LnkAction	
Likelihood	
DateTo	
DateFrom	
Comment	

case 6

Certificate

I a) Abdominal carcinosis C762
b) Pancreatic tumour D377
c)
d)

II

Final codes: C798 for Abdominal carcinosis
C259 for Pancreatic tumour

Comment: The pancreatic tumour is not specified as Malignant or benign and is first coded D377. The Abdominal carcinosis is not specified as primary or secondary site and is first coded C762. Since the pancreatic tumour causes the abdominal carcinosis, the pancreatic tumour is malignant. This linkage modifies the D377 code into C259. Now the Abdominal carcinosis can be stated as secondary and is modified into C798.

Linkage table	
IcdCode	D377
OrderNum	1
ExclusiveGroup	
LnkCodes1	C762
LnkRel1	
LnkCodes2	
LnkRel2	
ResultCode	C798
IcdCodeAction	Y
Time	
Likelihood	

Linkage table	
IcdCode	C762
OrderNum	1
ExclusiveGroup	
LnkCodes1	C798
LnkRel1	
LnkCodes2	
LnkRel2	
ResultCode	C259
IcdCodeAction	Y
Time	
Likelihood	

4.3 Editing

This step checks that the codes assigned are consistent with sex and age condition. Codes with usage restricted to one sex (e.g. malignant neoplasm of prostate) or to a precise age (e.g. antepartum haemorrhage) are listed in the Edit table (see 4.5) with the conditions of restriction. The records listed below show some examples of the Edit table.

CodeInf	CodeSup	Order Num	Sex	Age Inf	Age Sup	Age Unit	Edit Type	Cause Status	Message
C600	C639	1	2				I	A	Code not used for males
P000	P999	1		1	10	Y	R	A	Rare code for children between 1 and 10 years old
P000	P999	2		11	999	Y	R	A	Very rare code over 10 years old
P546	P546	1	1				I	A	Code not used for males

Line 1 specifies that codes for malignant neoplasms of male organs are restricted to males. Line 2 and 3 show than Chapter XVI (Certain conditions originating in the perinatal period) are seldom used for children between 1 and 10 years old and very seldom used for people over 10 years old. On line 4, code P546 is only for females.

4.4 Output standardisation

Ouput standardisation consists in presenting the ICD10 codes for the program that will select the underlying cause of death. As most of the time, this program is the NCHS system ACME, the output standardisation is based on a code representation compatible with ACME. The standardisation deals with 2 aspects: the code representation and the ampersand flag placement.

4.4.1 Code representation

The ICD10 codes must be presented in such a way that their place on the death certificate is kept because the selection of the underlying cause depends widely on this topologic information. The convention is the following:

. codes are separated by a space

. lines are separated by a "/"

. parts are separated by a "*"

The followings death certificate and codes:

I a) sepsis, cardiorespiratory arrest	B99, R092
b) hip fracture	S720
c)	
d) fall	W19
II Hypertension, cardiac insufficiency	I10, I509

will be represented by the following record:

B99 R092/S720/&W19*I10 I509

4.4.2 Ampersand flag placement

The ampersand sign (&) is used by ACME to select the underlying cause when an external cause of death is involved. Rules to assign ampersands are given in [NCHS01] pages 200 and 289. This procedure is quite complex to be done manually and the coding program will help. In the algorithm below, instructions page 220 are called C1 through C5 and the instructions page 289 are called R1 though R6.

Begin

C1 – If there is only one code from chapter XX, precede it with an ampersand.
If there are several codes from chapter XX, ask for the more specific external cause. Precede it with an ampersand

C2 - If a code Y40-Y84 has an ampersand, perform R1 through R6

C3 - If there is no condition requiring the medical or surgical care and if the complication or adverse effect is coded to chapter I-XVIII, precede the complication with an ampersand.

C4 - If there is a code X45-X49 or W78-W80 with an ampersand, ask if there a localised effect of this code, if yes, precede it with an ampersand.

C5 - If there is a code W78-W80 with an ampersand, check that the immediate cause (first causes mentioned on the first line) is coded T179. If nor unampersand the code W78-W80.

End

R1 - Ask for the underlying condition that necessitated the medical or surgical care. Precede it with an ampersand.

R2 - Ask for the complication, if it is NOT coded T80-T88, delete the ampersand on the Y code.

R3 - If there is no condition requiring the medical or surgical care and if the complication is coded to chapter I-XVIII, precede the complication with an ampersand.

Ask for the condition that necessitated the medical or surgical care.

R4 - If there is only one condition and if the medical or surgical care was administered for an injury, precede the code of the injury with an ampersand.

R5 - If there are several conditions, precede the first code with an ampersand.

R6 - If the medical care was administered for diagnostic purposes, precede the code for the underlying or most definite condition that was found or confirmed by the diagnostic finding with an ampersand.

4.5 Table description

4.5.1 General structure of the tables

The coding system uses a database to store all the data needed by the functions. This database contains several specific tables. Some fields are common to all the tables.

- In each table, records are identified by a **Key** field common to a group of records. This field is of text type, so any letter or digit can be used to build an explicit key. When necessary, an additional **Order** field gives the sequence of application of the records: records with the same Key value must have a different Order value. **Order** is an integer.

- When it is mentioned, the field **Action** contains the name of a specific procedure to be performed. This field gives additional possibilities to program specific procedures for complex cases. However the specific procedure must be developed in the software.

- A **Likelihood** field is always present to indicate whether the operation defined by the record is certain ('C') or only probable ('P'). In this last case the modification is done but a message is sent to the operator to point out the operation.

- Fields **DateTo** and **DateFrom** are used for update purpose: DateTo indicates the date until which the record was used, DateFrom indicates the date from which the record was used. Only the records with null DateTo are used.

- Finally a text field **Comment** can be used to add additional information concerning the record.

4.5.2 Standardisation1 and standardisation2 tables

These 2 tables have the same structure but not the same content. They are used in the Syntax standardisation 1 (4.1.3) and Syntax Standardisation 2 (4.1.7) functions. The content of these tables depends of the language used.

Structure

Attribute	Type (length)	Key/index	Comments
Key	Text (6)	Primary key	
OrderNum	Integer	Primary key	Sequence number
FilterIn	Text (100)		Description of the input text to be found
FilterOut	Text (100)		Description of the output text
Action	Text (10)		Name of a specific action to be performed in the program of the function
Likelihood	Text (1)		"C" : Certain "P" : Probable
DateTo	Date - dd/mm/aaaa		Time until when the record has been used
DateFrom	Date - dd/mm/aaaa		Time from when the record has been used
Comment	Text (100)		Free text concerning the record

Comments

"Key" and "OrderNum" are used to group and arrange the records. The fields "FilterIn" and "FilterOut" are filters used to modify the line content: if the input text matches with the input filter, the text is modified according to the output filter. (see 4.1 for filter syntax). When it is mentioned, "Action" is the name of a specific procedure to be performed.

4.5.3 Separator table

The Separator table is used with the Causal or non-causal separator standardisation function. Its content depends on the language used

Structure

Attribute	Type (length)	Key/index	Comments
Key	Text (6)	Primary key	
OrderNum	Integer	Primary key	Sequence number
OrigSep	Text (100)		Filter describing the input text to be found
Type	Text (1)		Type of connector: 1: non-causal connection; 2: causal (left causes right) 3: causal (right causes left)
Action	Text (10)		Name of a specific action to be performed in the program of the function
Likelihood	Text (1)		"C" : Certain "P" : Probable
DateTo	Date - dd/mm/aaaa		Time until when the record has been used
DateFrom	Date - dd/mm/aaaa		Time from when the record has been used
Comment	Text (100)		Free text concerning the record

Comments

"Key" and "OrderNum" are used to group and arrange the records.

The fields "OrigSep" is the text of the separator to be standardised (e.g. "due to").

"Type" is the nature of the separator: non-causal, causal left or causal right depending on the direction of the causality.

When it is mentioned, "Action" is the name of a specific procedure to be performed.

4.5.4 Time interval table

The Time Interval table is used in the Time interval function (4.1.5). This table depends on the language used.

Structure

Attribute	Type (length)	Key/index	Comments
Key	Text (6)	Primary key	
OrderNum	Integer	Primary key	Sequence number
TimeExp	Text (100)		Filter describing the input text to be found
Action	Text (10)		Name of a specific action to be performed in the program of the function
Likelihood	Text (1)		"C" : Certain "P" : Probable
DateTo	Date - dd/mm/aaaa		Time until when the record has been used
DateFrom	Date - dd/mm/aaaa		Time from when the record has been used
Comment	Text (100)		Free text concerning the record

Comments

"Key" and "OrderNum" are used to group and arrange the records.

"TimeExp" is a string with the same filter syntax defined in the Standardisation1 table (see 4.1). When this filter matches the input text, the specific action mentioned in the field "Action" is performed. This action return a standard time indicator of the form [Int*iu*] where Int is a constant that identify an interval, i is an integer and u is a time unit (m: minutes, h: hours, D: days, W: weeks; M: months; Y: years).

4.5.5 Surgery table

This table is used in the Surgery function (4.1.6). It depends on the language used.

Relation structure

Attribute	Type (length)	Key/index	Comments
Key	Text (6)	Primary key	
OrderNum	Integer	Primary key	Sequence number
SurgExp	Text (100)		String indicating surgery
Direction	Text (1)		"U" unknown; "L" Left; "R "right"
Action	Text (10)		Name of a specific action to be performed in the program of the function
Likelihood	Text (1)		"C" : Certain "P" : Probable
DateTo	Date - dd/mm/aaaa		Time until when the record has been used
DateFrom	Date - dd/mm/aaaa		Time from when the record has been used
Comment	Text (100)		Free text concerning the record

Comments

"Key" and "OrderNum" are used to group and arrange the records.

"SurgExp" is a string that indicates the mention of a surgery act (e.g. "suite à opération"). This expression is associated with a direction indicating if the expression affects the left or right diagnosis. When the surgery expression is found in a text, it is standardised under the form [<S] or [>S] depending whether the direction is left or right.

When it is mentioned, "Action" is the name of a specific procedure to be performed.

4.5.6 Index table

The Index table gives maps the text of a diagnosis to 1 or 2 ICD10 codes. It is used both in the Text recognition function (4.1.2) and Code Query functions(4.2.1). This table depends on the language used.

Structure

Attribute	Type (length)	Key/index	Comments
Key	Text (6)	Primary key	
OrderNum	Integer	Primary key	Sequence number
Diagnosis	Text (100)	Indexed	Text of the diagnosis
IcdCode1	Text (6)	Indexed	1rst corresponding code
IcdCode2	Text (6)	Indexed	2nd corresponding code
Likelihood	Text (1)		"C" : Certain "P" : Probable
DateTo	Date - dd/mm/aaaa		Time until when the record has been used
DateFrom	Date - dd/mm/aaaa		Time from when the record has been used
Comment	Text (100)		Free text concerning the record

Comments

"Key" can be any combination of letter or digit. OrderNum is a sequence number. Diagnosis is the text of a disease, which corresponds to IcdCode1 and IcdCode2. Diagnosis, IcdCode1 and IcdCode2 fields should be indexed in order to fasten research.

4.5.7 Code Modification table

The Code Modification table is used with the code Selection function (4.2.2). This table is language independent.

Structure

Attribute	Type (length)	Key/index	Comments
IcdCode	Text (6)	Primary key	
OrderNum	Integer	Primary key	Sequence number
Sex	Text (1)		Sex condition 1: code for male only 2: code for female only
Age	Text(100)		Age condition
Time	Text(100)		Time interval condition
NewCode	Text (6)		Resulting code
Likelihood	Text (1)		'C' : Certain 'P' : Probable
DateTo	Date - dd/mm/aaaa		Time until when the record has been used
DateFrom	Date - dd/mm/aaaa		Time from when the record has been used
Comment	Text (100)		Free text concerning the record

Comments

IcdCode is associated with one of the fields "IcdCode1" or "IcdCode2" of the index table. **OrderNum** is added to IcdCode to build a primary key and to order this subgroup of records.

Age and **Time** fields state a time interval of the form "s-eU" where s and e are integers indicating the beginning and the end of the interval and where U is the unit of time (M: minutes; H: hours; D: days; W: weeks; Y: year). Months are not used as a unit since it does not corresponds to a precise lapse of time. If e is not specified, it is considered as infinite. The following expressions are valid Age or time conditions:

0-28D	from 0 to 28 days included
1-14Y	from 1 to 14 years
0-2M	from 0 to 2 minutes
4-W	from 4 weeks (e not specified)

For the Time field, a symbolic name 'Age' can be used to indicate a delay equal to the age of the decedent.

NewCode field contains the ICD code that will replace the value of IcdCode in the record if the sex, age and time conditions stated are satisfied.

4.5.8 Code linkage table

The Code Linkage table is used in the Linkage function (4.2.3). This table is language independent.

Structure

Attribute	Type (length)	Key/index	Comments
IcdCode	Text (6)	Primary key	
OrderNum	Integer	Primary key	Sequence number
LnkCodes1	Text (100)		Linkage condition, lower ICD10 code Null: no linkage condition
LnkRel1	Text (2)		Relationship with the Linked code. See comments below
Time	text(100)		Time condition, minimum time interval between IcdCode and LnkCode1
LnkCodes2	Text (100)		Linkage condition, lower ICD10 code Null: no linkage condition
LnkRel2	Text (2)		Relationship with the Linked code. See comments below
ResultCode1	Text (6)		Icd10 code resulting from the linkage. If null the resulting code is the linked code.
ResultCode2	Text (6)		Icd10 code resulting from the linkage. If null the resulting code is the linked code.
IcdCodeAction	Text(1)		States if the IcdCode must be replaced or not 'N' the original code is kept 'Y' the original code is deleted (default)
LnkAction	Text (1)		States if the linkage process in the group must stop or not 'S' stop the linkage 'N' do not stop the linkage (default)
Likelihood	Text (1)		'C' : Certain 'P' : Probable
DateTo	Date - dd/mm/aaaa		Time until when the record has been used
DateFrom	Date - dd/mm/aaaa		Time from when the record has been used
Comment	Text (100)		Free text concerning the record

Comments

IcdCode is the basic code that will be changed or not depending on the conditions stated in the rest of the table. IcdCode has been given by the Index table and modified or not by the Code Modification table.

OrderNum is added to IcdCode to build a primary key and to order this subgroup of records.

LnkCodes1 and **LnkCodes2** give a set of ICD10 codes. Isolated codes, interval can be used. For instance:

I200	stands for the code I20.0
I20-I25	stands for all the codes from I20.0 to I259 included
I05-I09, I20-I25, I300	stands for I05.0 to I09.9 plus I20.0 to I259 plus I30.0

Set difference can also be expressed provided that parenthesis are used to avoid ambiguity:

(C0-C97) – (C76-C80) is equivalent to C00-C75, C81-C97

LnkRel1 and LnkRel2 state the relationship condition that must exist between IcdCode and the linked code. The following relationships are considered:

Causal relationship	
'CB' Caused by	The 2 codes are in part 1 of the death certificate and the linked code must be on an upper line than the IcdCode
'CL' line just above	The 2 codes are in part 1 of the death certificate and the linked code must be on the line just above the IcdCode
'CE' Explicit	The 2 codes are in part 1 of the death certificate on the same line and the linked code must be explicitly stated as cause by the IcdCode.

Non causal relationship:	
'NC'	The 2 codes are placed anywhere on the death certificate
'RB'	The linked code is on the same line immediately before the IcdCode
'RA'	on the same line immediately after
'RL'	The linked code is on the same line anywhere

ResultCode is the code that will replace or added to the IcdCode if the linkage conditions are satisfied.

IcdCodeAction indicates whether IcdCode is deleted or kept

LnkAction is a flag that indicates whether, when a linkage succeeds, the search for other linkage with the same code should continue within the subgroup or not.

4.5.9 Code Groups table

This table allows facilities to refer to group of ICD10 codes via symbols.

Attribute	Type (length)	Key/index	Comments
Name	Text (20)	Primary key	Symbolic name of the code group
OrderNum	Integer	Primary key	Sequence number
CodeGroupExp	Text (100)		Expression of the Icd10 codes belonging to the group
DateTo	Date - dd/mm/aaaa		Time until when the record has been used
DateFrom	Date - dd/mm/aaaa		Time from when the record has been used
Comment	Text (100)		Free text concerning the record

Comments

Name is a symbol including letters and/or figures without space.

CodeGroupExp is an expression of a set of ICD10 codes build in the same way that LnkCodes1 in the linkage table. symbols are excludes from this expression.

When a name is defined in the Code Group table, it can be used in the linkage table in the LnkCodes1 and LnkCodes2 expressions instead of the ICD codes.

4.5.10 Edit table

Attribute	Type (length)	Key/index	Comments
CodeInf	Text (5)	Primary key	Lower Icd10 code
CodeSup	Text (5)	Primary key	Higher Icd10 code
OrderNum	Integer	Primary key	Sequence number
Sex	Text (1)		Sex condition 1: code forbidden for male 2: code forbidden for female
AgeInf	Integer		Age condition, minimum age for which the code cannot be used
AgeSup	Integer		Age condition, maximum age for which the code cannot be used
AgeUnit	Text (1)		Y: years D: days
EditType	Text (1)		Type of edit I: impossible cause R: rare cause
CauseStatus	Text (1)		Status of causes concerned by the edit A: all causes U: underlying cause only O: causes other than the underlying cause
Message	Text (100)		Message to the operator
Likelihood	Text (1)		"C" : Certain "P" : Probable
DateTo	Date - dd/mm/aaaa		Time until when the record has been used
DateFrom	Date - dd/mm/aaaa		Time from when the record has been used
Comment	Text (100)		Free text concerning the record

Comments

The Edit table checks that the codes assigned are consistent with sex and age condition. Codes with usage restricted to one sex (e.g. malignant neoplasm of prostate) or to a precise age (e.g. antepartum haemorrhage) are listed with the conditions of restriction. "CodeInf" and "CodeSup" delimit the group of codes concerned by the record. The field "Sex", "AgeInf", "AgeSup" and "AgeUnit" specify the conditions restricting the use of the group of codes.

If these restrictions are not satisfied, the code is pointed out and the field "Message" is given to the operator.

5. Coding prototype

5.1 Objective

This prototype has been developed to test the text parsing function (see 4.1). This function is language dependent and involves a formal description of the tasks that must be applicable to different languages. In this perspective, a small prototype has been developed to evaluate the validity of the formalisation for both French and Swedish. This program is a tool used to help defining the specifications and is not a final product.

5.2 Realisation

The prototype was written with Visual C++ (Microsoft) as programming language and with the database management system Access 97 (Microsoft). The architecture of the program is close to the description given in chapters 3. Only the text parsing function has been implemented but the whole architecture has been considered for future development. The structure of the text parsing function is also very close to the description given in chapter 4.1.

Conventions

The following conventions are used for variables names

- Classes names always begin with a C (eg CAdcsView)
- Classes related to a database relation are suffixed with the "Set" mention (eg CStandardSet)
- Classed related to a dialog form resource are suffixed with "Dlg" (eg CTraceDlg)
- Global variables are written in capital letters (eg CERT)
- Local variables begins with a small letter (eg filterIn)

Classes architecture

This architecture is given in figure 5.2.1. The program is called Adcs and this name is given to the main class. The relationship between classes indicated by a line is a "has a" relationship. For instance a CTextParsing class object is declared in CAdcsView and a CStandardSet1 class object is declared in CTextParsing.

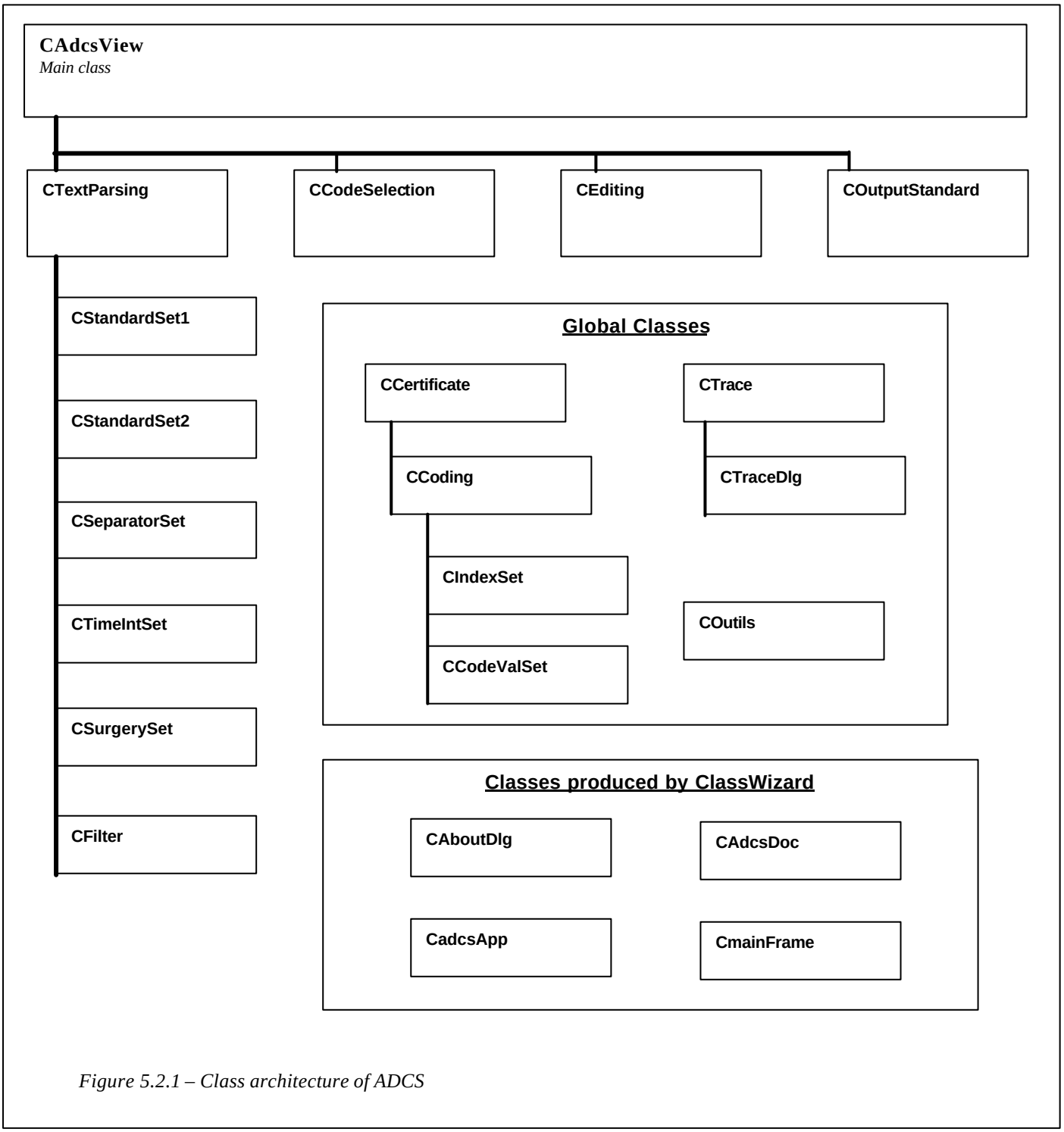


Figure 5.2.1 – Class architecture of ADCS

Class description

CAdcsView: main class generated by CWizard. This class manages the main form dialog resource (IDD_ADCS_FORM). All the button and the edit boxes declared in this main form are managed in CAdcsView.

CTextParsing: this class performs the text parsing function. The public function TextParsing is called in the CAdcsView class

CStandardSet1: describes the table Standardisation1 used in the syntax standardisation 1 phase

CStandardSet2: describes the table Standardisation2 used in the syntax standardisation 2 phase

CSeparatorSet: describes the table Separator used in the causal and non-causal separator standardisation phase

CTimeintSet: describes the table TimeInterval used in the time interval phase

CSurgerySet: describes the table Surgery used in the surgery indications standardisation phase

CFilter: this class manages all the processing of the filters of the Standardisation1 and 2 tables.

CCodeSelection: performs the Code selection function. Not implemented

CEditing: performs the Editing function. Not implemented

COutputStandard: performs the Output standardisation function. Not implemented

CCertificate: this class describes a certificate. All the data (diagnoses, Icd codes, time intervals...) and the functions necessary to process these data are included in this class.

CCoding: this class is used by CCertificate to perform the coding translation of text into Icd codes.

CindexSet: describes the Index table. Used by CCoding

CCodeValSet: describes theCode Validity table

CTrace: this class stores and processes the indications given by the different functions that standardise and code the diagnoses. These indications can be listed by the CTraceDlg class.

CTraceDlg: this dialog class lists on the screen the traces left by the functions.

COutils: this class includes functions of general interest.

Interface

The interface of the program is shown in fig 5.2.2. The individual data are limited to the sex, date of birth and date of death variables. The death certificate is represented with 5 lines in part 1 and one line in part 2. Each edit box allows entering the text of diagnoses. The size of boxes is not limited.

The "Code" button launches the text parsing step. The "Trace" button gives the detail of the processing performed. The "New" button resets the death certificate.

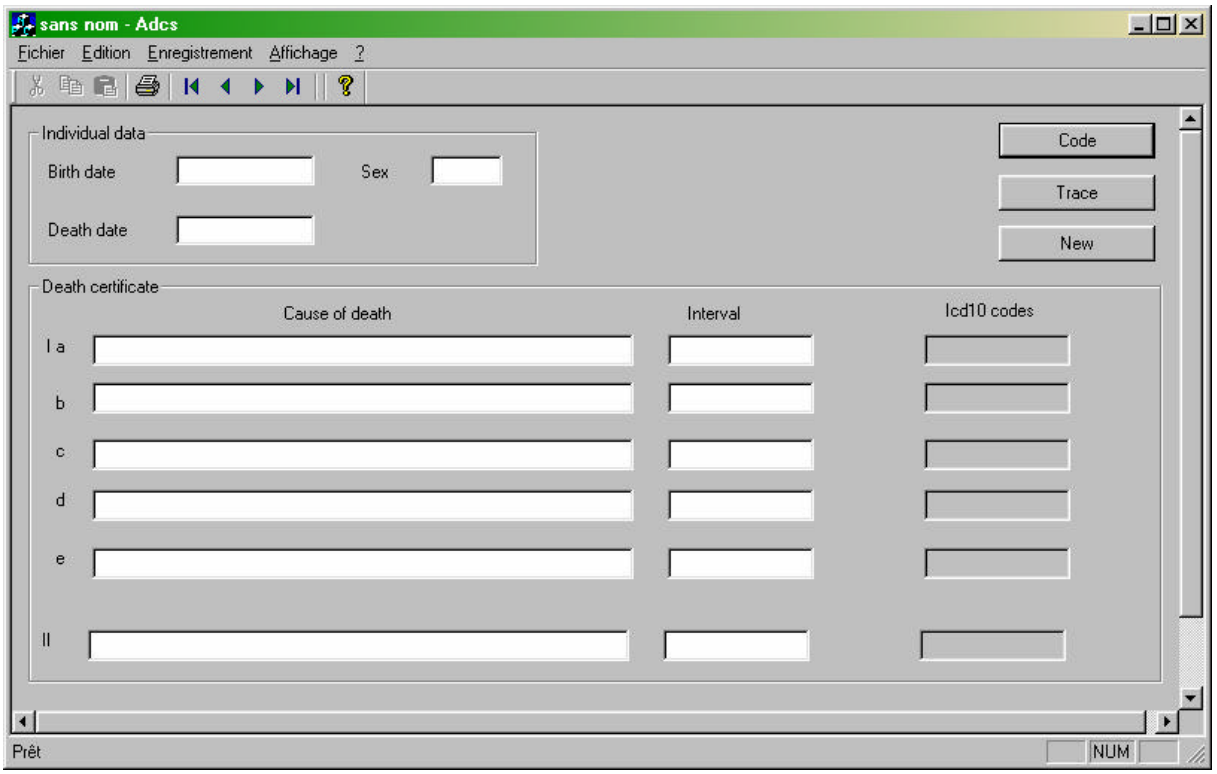


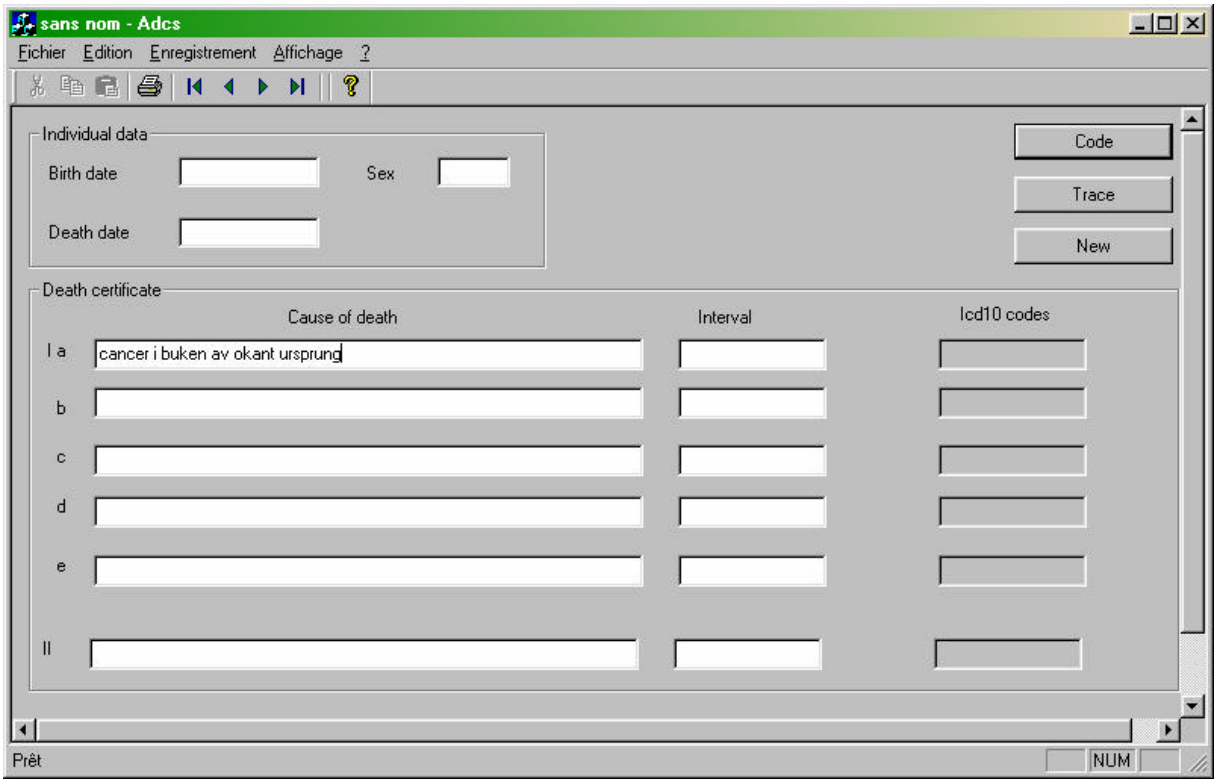
Fig 5.2.2 Adcs interface

5.3 Functioning

As specified in chapter 4.1, the aim of the text parsing function is to standardise the text of diagnoses in order to reduce the size of the index and to standardise additional information, namely separators, surgery and time indications. The algorithm is language independent and the language specific information is described in filters stored in a database. The following examples show the functioning of the text parsing function through French and Swedish examples.

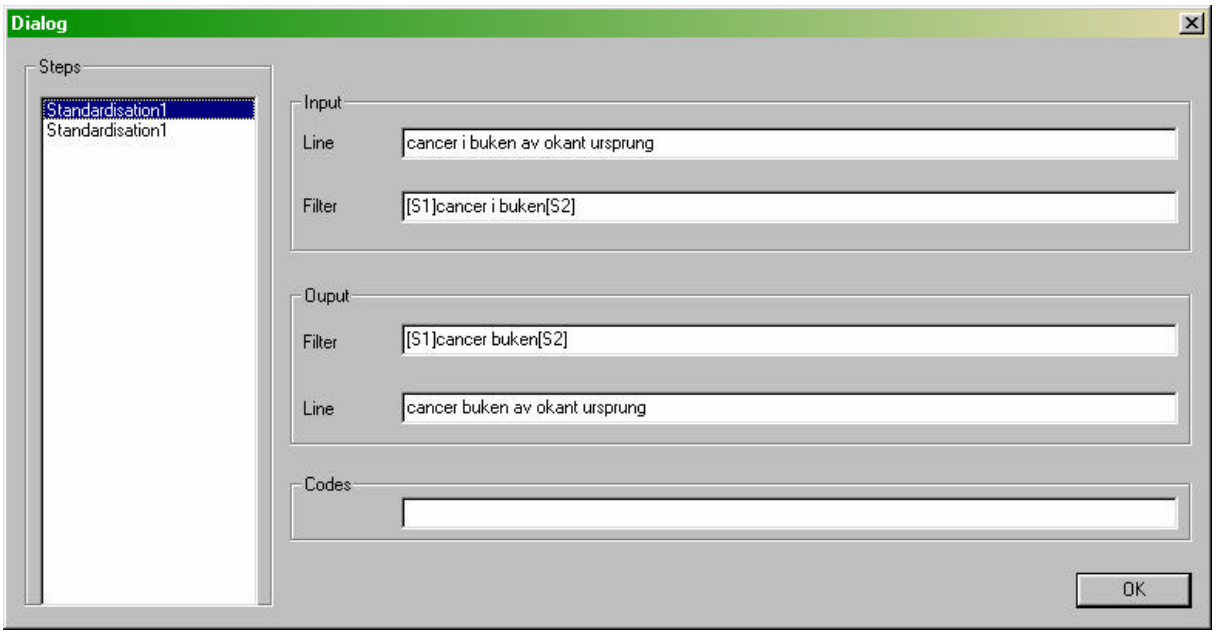
Example 1

The following example shows how the standardisation of diagnosis is done. The diagnosis "cancer i buken av okänt ursprung" mean "Malignant neoplasm of abdomen of unknown origin" and must be interpreted as "metastasis of abdomen":

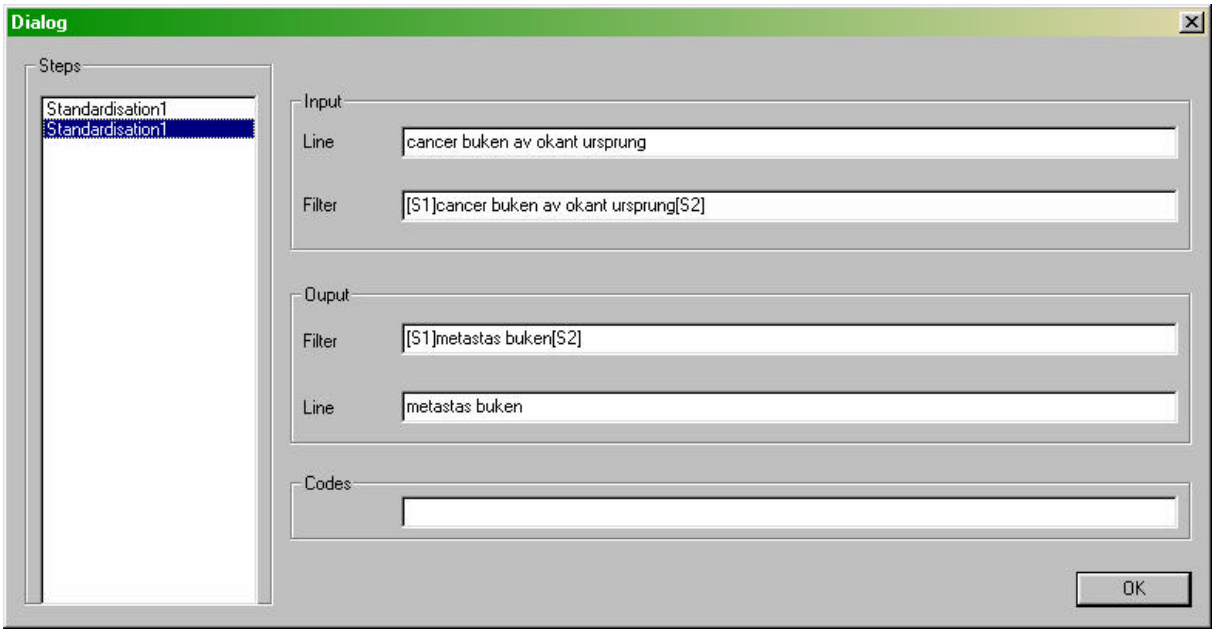


First, the Swedish expression "cancer i buken" is equivalent to "cancer buken". Since both expressions can be found in the death certificates, the expression "cancer buken" will be considered as the standard and included in the index. The expressions "cancer i buken" is then

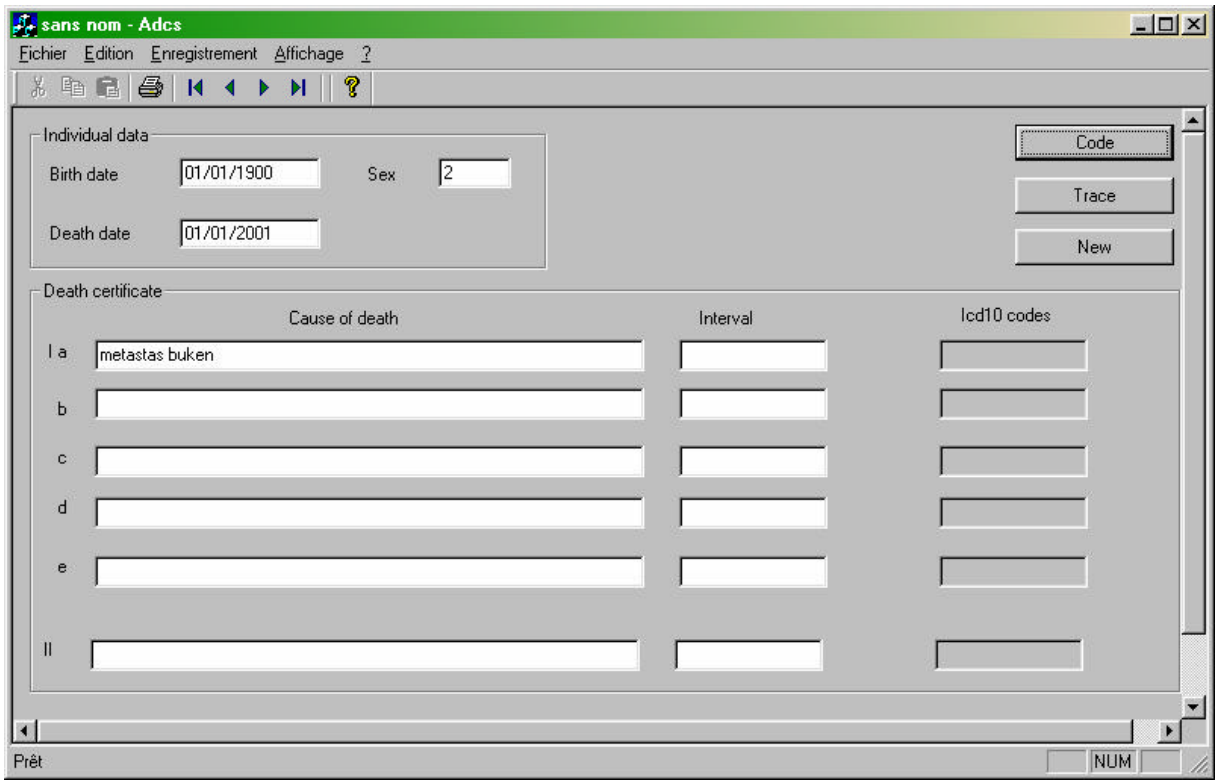
reduced in the first step. In the window below, the input line is modified into the output line according to the input and output filter (see 4.1.3):



In the second step the expression "cancer buken av okänt ursprung" is transformed in "metastas buken" using the same process but different filters.

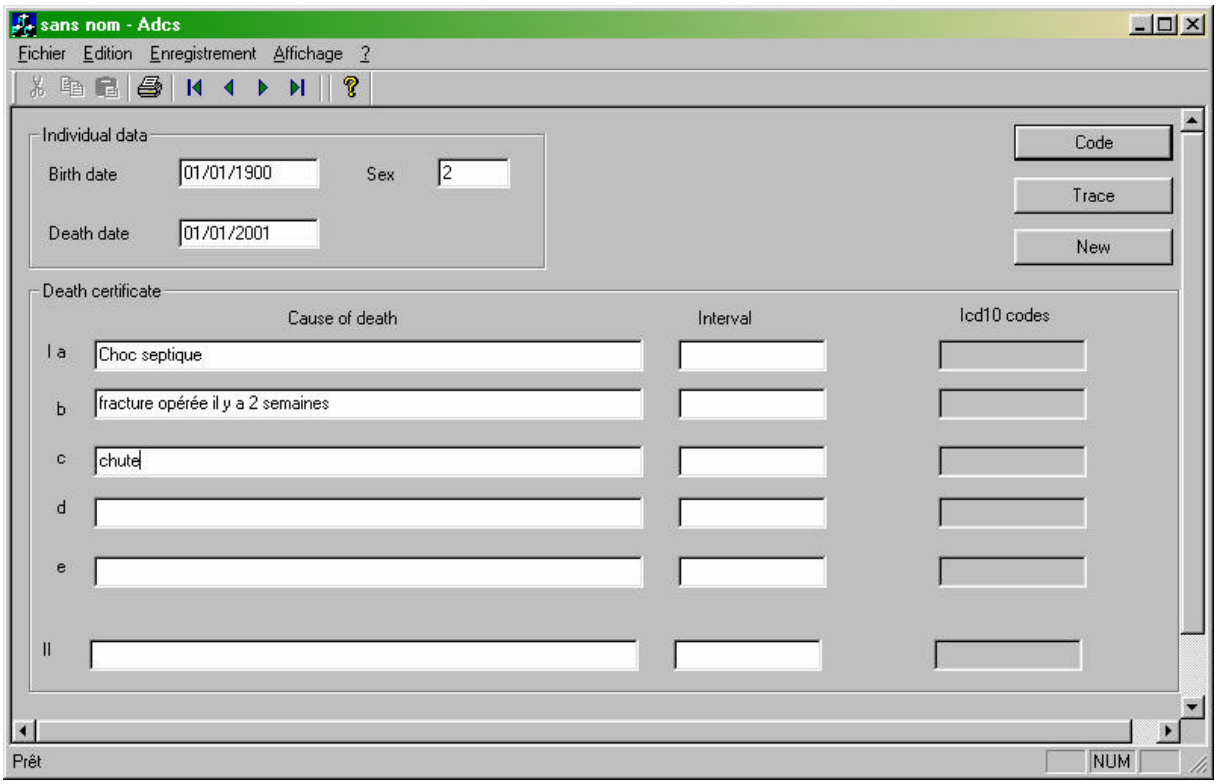


Finally, the standardised text replaces the input text:

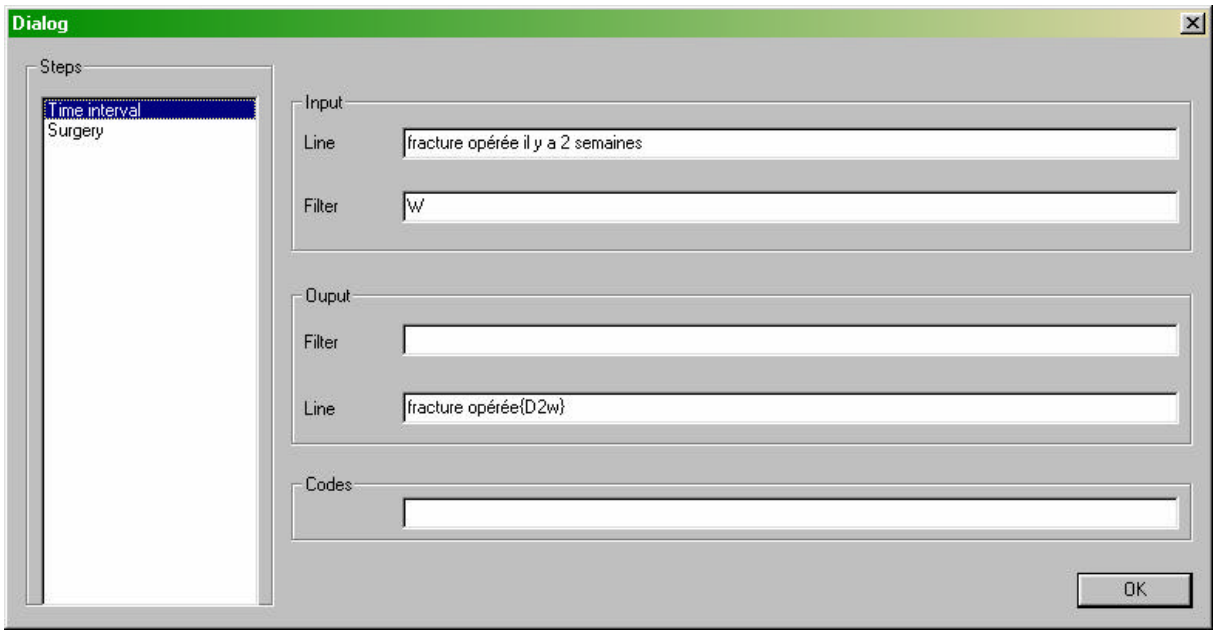


Example 2

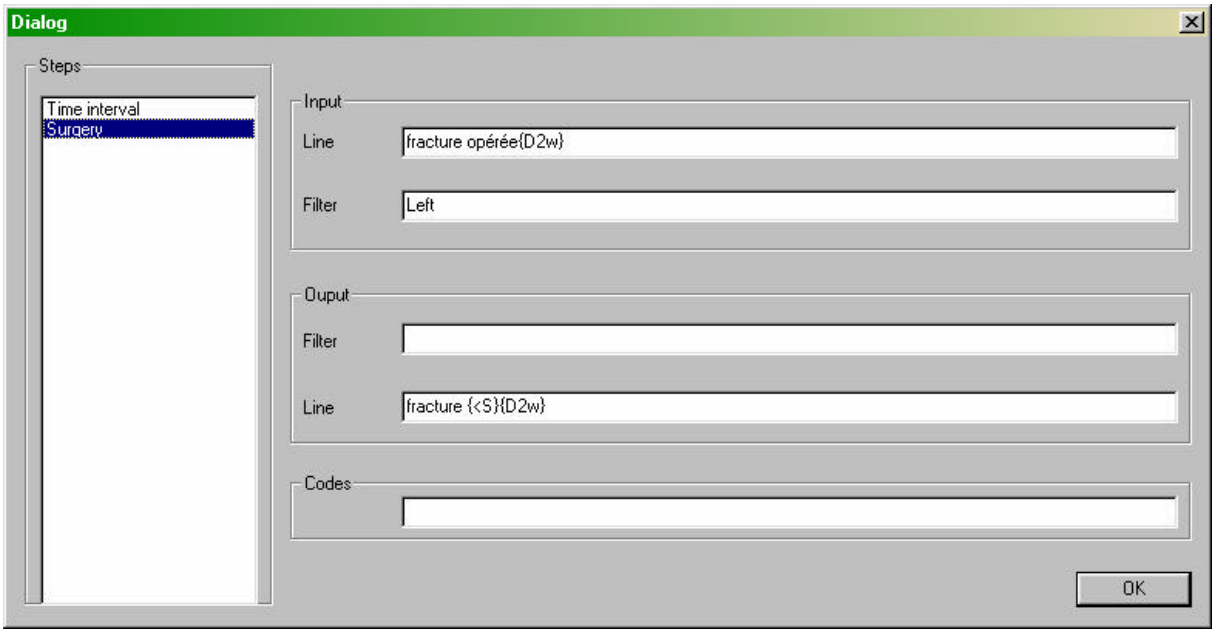
This example shows the text parsing function action on duration and surgery information. It is given in French. The diagnosis "Fracture opérée il y a 2 semaines" means "Fracture with surgery 2 weeks ago". This diagnosis includes both the mention of surgery ("opérée") and the mention of duration ("il y a 2 semaines").



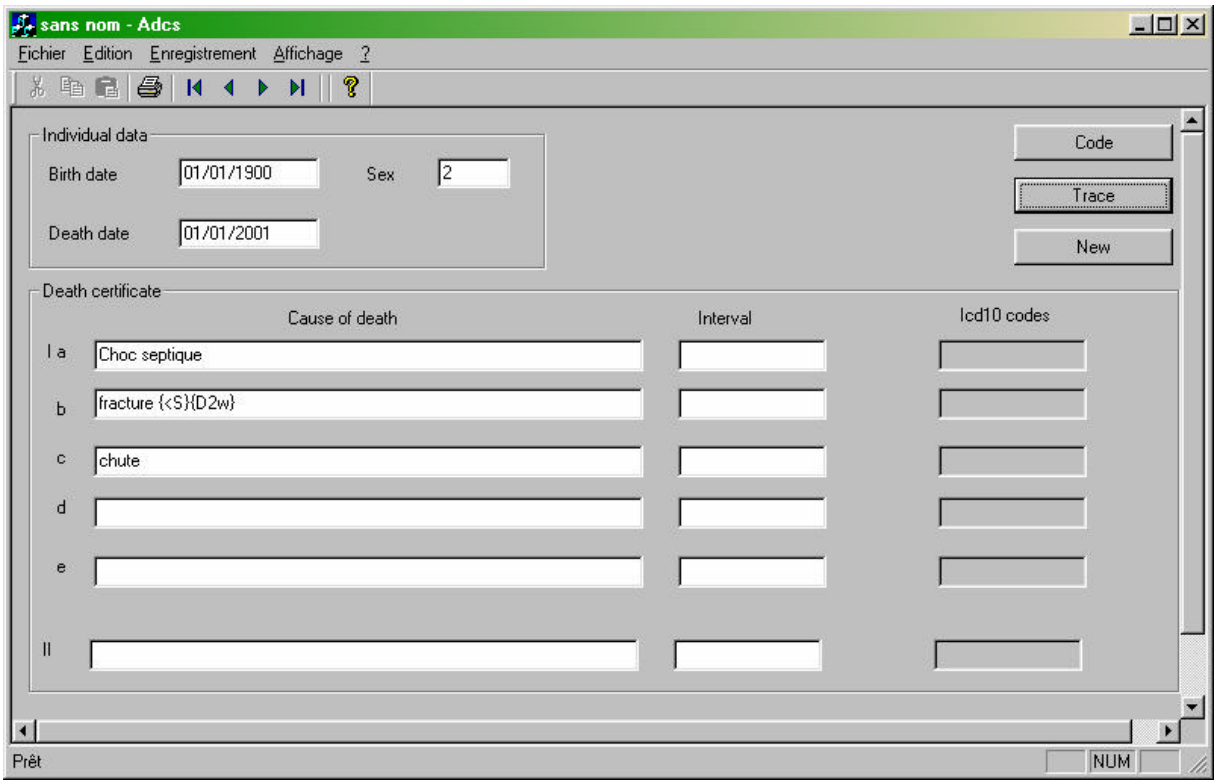
The first step is the standardisation of the mention of duration. The expression of the delay "il y a 2 semaines" is replaced by the standard tag [D2W] which means "duration = 2 weeks".



The second step standardises the mention of surgery. The standard tag [<S] that replaces the word "opération" means that there was surgery associated with the diagnosis placed on the left of the tag:



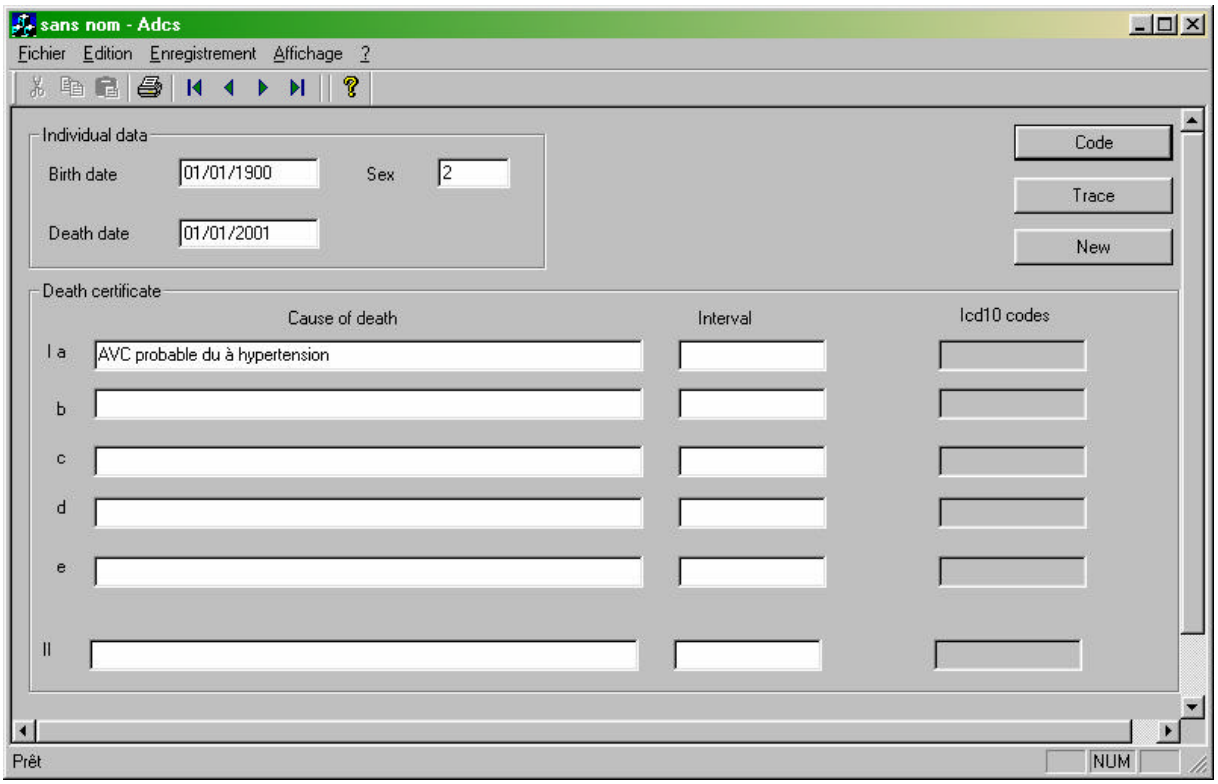
Finally, the standardised output text replaces the input text ("fracture opérée il y a 2 semaines"):



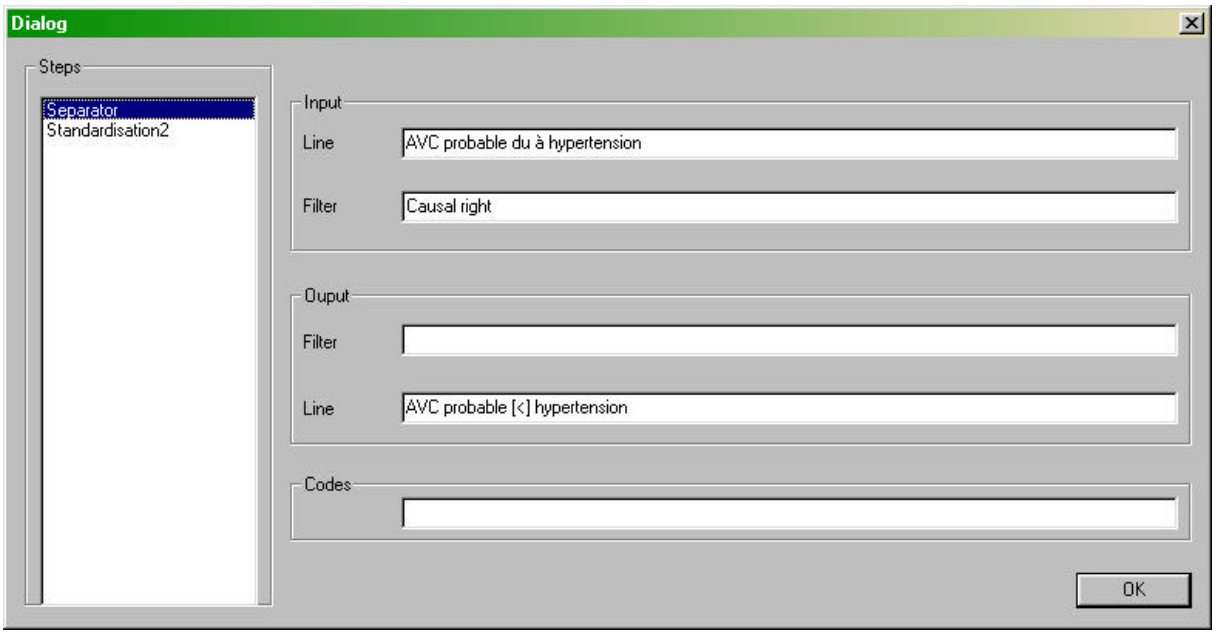
This standardised text is more manageable for automated codification than the "human" expressions because the syntax is formal, unique and unambiguous.

Example 3

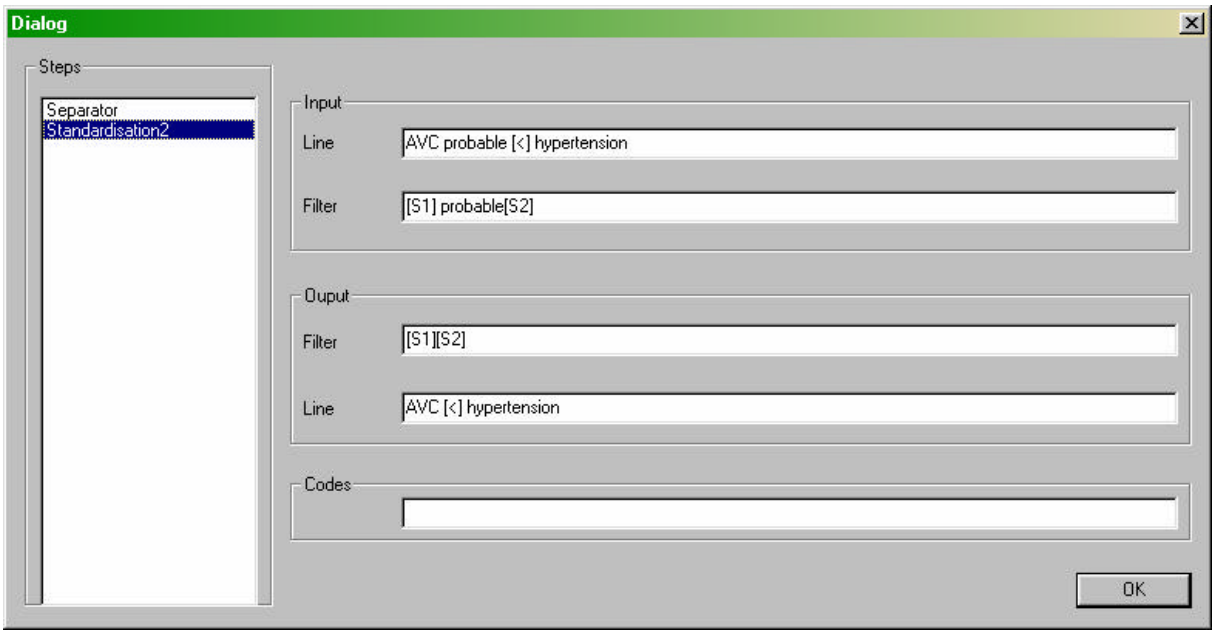
This last example shows the standardisation of causal separator. The French diagnosis reported in the windows below means "Probable cerebrovascular accident due to hypertension".



First, the causal separator "du à" is replaced by the standard tag [<]:



In a second step, as specified in the ICD10, the word "probable" is deleted since it does not influence the codification:



Again, the standardised expression will be easier to manage in the other steps of codification.

Part 2 Knowledge base on bridge coding studies

1. Introduction

Bridge coding studies mainly aim at assessing changes in coding methods and/or classification. At national level, these studies provide indications on the shifts induced by the coding modifications. Bridge coding are used to tune the new coding methods and to compute accurate rates to correct time trend studies. At international level, bridge coding studies could be an invaluable tool to evaluate the comparability of data. However the results of these studies are not always available and even when papers are published or disseminated, results are difficult to compare due to differences in the method used and the presentation of results.

In the context of ICD10 implementation, it would be very important to centralize the results of bridge coding studies in a knowledge base. Moreover, in order to be able to compare results between countries, it is fundamental to present these results in a standard form.

In this report we focus on the ICD9-ICD10 bridge coding studies. The first objective is to inventory studies or plan for studies at international level and not only at European level. The result of this survey is presented in chapter 2. The second objective is to propose a plan to make these results available at European level for all the countries involved in ICD10 implementation. Chapter 3 presents this proposition.

2. Survey on Bridge coding studies

2.1 Questionnaire

The survey on the inventory of bridge coding studies started in October 2000. The survey takes in account all the countries that moved or plan to move from ICD9 to ICD10. The survey was not restricted to Europe for 2 reasons. Firstly the problems of classification change are the same for all the countries. Secondly, bridge coding studies are rare and it is essential to share the available knowledge. In addition, the final aim is to achieve international comparability for mortality data and in this context all international efforts must be taken into account.

The questionnaire presented in annex 1 was sent to all the countries having moved or planning to move from ICD9 to ICD10. The questionnaire first identifies the method used for the bridge coding studies. It also gives information on the manual or automated coding method used for the assignment of the multiple cause codes and for the selection of the underlying cause of death. Finally, the questionnaire asks for available data and papers on the study.

2.2 Results

The questionnaire was sent to 11 countries or statistical offices: Australia, Brazil, Canada, Catalonia, England and Wales, France, Italy, Japan, Scotland, Sweden and United States. 7 countries sent the questionnaire back: Australia, Brazil, Canada, France, Japan, Scotland and Sweden. Italy answered that they will sent the questionnaire when the date are available.

Table 3.2.1 synthesises the results on the bridge coding method and on the manual/automated coding system used. It appears that the situations and the objectives of the bridge coding studies are different depending on the country. Australia has undertaken 2 surveys, one to evaluate the changes induced by the adoption of an automated coding system in ICD9 (manual ICD9 Vs automated ICD9) and one to evaluate the changes due to the implementation of ICD10 (automated ICD9 Vs automated ICD10). Japan and Scotland evaluate the shifts due to the change of classification. Brazil and France evaluate at the same time the shifts due to the change in codification method (manual Vs automated) and to the change in classification (ICD9 Vs ICD10). In Canada, the procedures vary for the 10

provinces and the 3 territories. Thus the ICD9 or ICD10 coding can be manual or automated depending on the province or territory considered. Sweden evaluate the changes due to classification from ICD9 automated coding to ICD10 automated coding with a manual verification of the automated coding.

Country	Data year	Method	Icd9 assignment codes	Icd10 assignment codes	Icd9 underlying cause selection	Icd10 underlying cause selection
Australia	1997 1997-1998	Random sample Complete	automated Automated	 Automated	Manual and automated Automated	 Automated
Brazil	1994	Random sample 10 %	Manual	Automated	Manual	Automated
Canada	1999	Stratified Sample	Mixed	Mixed	Mixed	Mixed
France	1999	Random Sample 10 %	Manual	Automated	Manual	Automated
Japan	1994	Random sample	Automated	Automated	Automated	Automated
Scotland	1999	Complete	Automated	Automated	Automated	Automated
Sweden	1996	Stratified sample	Automated	Mixed	Automated	Mixed

Table 3.2.1 - Bridge coding method and coding system used

Table 3.2.2 describes the status of the bridge coding studies. Most of the papers or reports produced are based on preliminary studies. Brazil and France have published preliminary results [Santo97][Hatton96]. Japan has no data available but a paper describes very precisely the study [Oomi96]. Sweden have data and papers available [Johansson00]. Apart from this survey, United States and England published papers [Rooney97][Anderson01]. Italy and England are also working on reports.

Country	Data available	Paper or report available
Australia	Yes	
Brazil	2002	Yes
Canada	2001	
France	2001	Yes
Japan	No	Yes
Scotland	2001	
Sweden	Yes	Yes

Table 3.2.2 – Status of the bridge coding studies

3. Proposition of result dissemination

The objective of the result dissemination is to make available the results of bridge coding studies already done for all the European countries that are in the process of implementing ICD10 and/or implementing automated coding system. The knowledge base should be as complete as possible including an accurate presentation of results. It should also be easily accessible.

3.1 Presentation of result

A knowledge base on bridge coding studies should include 3 types of information. 1)A short abstract of the study with the objectives, the method and the data year involved. 2) A standard presentation of the results. 3) A bibliography with the reference of the papers and when possible the papers themselves.

The problem is to define a standard presentation for the results. Two aspects must be taken into account: the ICD short list and the indicators.

- It is necessary to use a short list because the ICD includes thousands of different codes and because most of the studies aim at comparing 2 different versions of the ICD. The short list must play 2 roles: it reduces the level of details of the classification by grouping diseases and it gives equivalence between 2 versions of the classification. The WHO "ABC" developed for time series study and the Eurostat short list (see Annex 2) could both be used.

- The indicators used for the presentation of the results must be easily read and widely accepted. They should also allow to get condensed tables. Most of the papers used comparability ratios [Anderson01][Johansson00]. For instance, if a study compares coding with ICD9 and ICD10, and that for a given disease the number of death is with ICD9 and d_2 with ICD10, the comparability ratio will be $(d_2 - d_1) / d_1$. If the value of the indicator is lower than 1, the number of diseases decreases with the new classification, if it is greater than 1 it increases. In addition the absolute numbers should be given in the table. Another indicator is the cross tabulation [Hatton96][Santo97]. It gives the repartition of deaths within a 2-dimensions table, each dimension corresponding to a classification or a

method. The cross tabulations are valuable since they give the exchanges between ICD groups. However, due to the size of the tables they can only be used at a very rough level or stored in a database with a query procedure for categories of particular interest.

3.2 Accessibility

The content of the knowledge base must be easily available to all the countries that need to query such data. In order to keep such a base accurate, it must also be updated with all the new studies that will be undertaken. The best way to manage these constraints is to make the knowledge base available on a web server.

The server should allow to query the data and to download papers and tables. It should also give easy contact and procedure to include new information.

Annex 1 - Questionnaire

Survey on bridge coding studies between IDC9 and ICD10

Eurostat project 9351005

October 2000

Please return this questionnaire to:

Gérard Pavillon
Sc8-Inserm
44, Chemin de ronde
78116 Le Vésinet Cedex
France

Name: ..... Date: .....

Address: .....
.....
.....
.....
.....

Tel: ..... Fax: ..... e-mail: .....

Method used

Complete data year ?
Sample ?

Year(s) of death .....

If a sample was used:

Sample size: .....

Random ? Stratified ?

If stratified variable(s) used for stratification:

.....
.....
.....

Survey on bridge coding studies between IDC9 and ICD10

Assignment of ICD multiple cause codes

-with ICD9:

Manual ? Automatic ? Mixed (both manual and automatic) ?

-with ICD10:

Manual ? Automatic ? Mixed (both manual and automatic) ?

Selection of the underlying cause

-with ICD9:

Manual ? Automatic ? Mixed (both manual and automatic) ?

-with ICD10:

Manual ? Automatic ? Mixed (both manual and automatic) ?

Can you send a file with the format described below?

Yes ? No ?

ICD9 code ; ICD10 code ; Number of death for this ICD codes combination

(records in fixed format or fields may be separated by semicolon or tabulation)

Example:

	I210	410	3
I213	410	1	
I219	410	655	
I229	410	4	

This means that ***3*** *deaths were coded "410" with ICD9and "I210" with ICD10*
1 *death was coded "410" with ICD9and "I213" with ICD10*
655 *deaths were coded "410" with ICD9and "I219" with ICD10*
4 *deaths were coded "410" with ICD9 and "I229" with ICD10*

Can you send us the final report or any other relevant documentation on your bridge coding study?

Yes ? No ?

If yes, please send us these documents on paper or in electronic format.

Annex 2 - European Short List

CAUSES OF DEATH, "EUROPEAN SHORTLIST"

Corrected list – October 2000

Nr	Disease or external cause	ICD-10 code	ICD-9 code	ICD-8 code
	All causes of death	**A00-Y89**	**001-E999**	**000-E999**
01	*Infectious and parasitic diseases*	*A00-B99*	*001-139*	*000-136*
02	Tuberculosis	A15-A19,B90	010-018,137	010-019
03	Meningococcal infection	A39	036	036
04	AIDS (HIV-disease)	B20-B24	042-044	-
05	Viral hepatitis	B15-B19	070	070
06	*Neoplasms*	*C00-D48*	*140-239*	*140-239*
07	Malignant neoplasms	C00-C97	140-208	140-209
08	of which Malignant neoplasm of lip, oral cavity, pharynx	C00-C14	140-149	140-149
09	of which Malignant neoplasm of oesophagus	C15	150	150
10	of which Malignant neoplasm of stomach	C16	151	151
11	of which Malignant neoplasm of colon	C18	153	153
12	of which Malignant neoplasm of rectum and anus	C19-C20-C21	154	154
13	of which Malignant neoplasm liver and the intrahepatic bile ducts	C22	155	155, 197.8
14	of which Malignant neoplasm of pancreas	C25	157	157
15	of which Malignant neoplasm of larynx and trachea/bronchus/lung	C32-C34	161-162	161-162
16	of which Malignant melanoma of skin	C43	172	172
17	of which Malignant neoplasm of breast	C50	174-175	174
18	of which Malignant neoplasm of cervix uteri	C53	180	180
19	of which Malignant neoplasm of other parts of uterus	C54-55	179,182	182
20	of which Malignant neoplasm of ovary	C56	183.0	183.0
21	of which Malignant neoplasm of prostate	C61	185	185
22	of which Malignant neoplasm of kidney	C64	189.0	189.0
23	of which Malignant neoplasm of bladder	C67	188	188
24	of which Malignant neoplasm of lymph./haematopoietic tissue	C81-C96	200-208	200-209
25	*Diseases of the blood(-forming organs), immunol.disorders*	*D50-D89*	*279-289*	*280-289*
26	*Endocrine, nutritional and metabolic diseases*	*E00-E90*	*240-278*	*240-279*
27	Diabetes mellitus	E10-E14	250	250
28	*Mental and behavioural disorders*	*F00-F99*	*290-319*	*290-315*
29	Alcohol abuse (including alcoholic psychosis)	F10	291,303	291,303
30	Drug dependence, toxicomania	F11-F16, F18-F19	304-305	304-305
31	*Diseases of the nervous system and the sense organs*	*G00-H95*	*320-389*	*320-389*
32	Meningitis (other than 03)	G00-G03	320-322	320

33	*Diseases of the circulatory system*	*I00-I99*	*390-459*	*390-444.1* *444.3-458* *782.4*
34	Ischaemic heart diseases	I20-I25	410-414	410-414
35	Other heart diseases	I30-I33, I39-I52	420-423, 425-429	420-423, 425-429
36	Cerebrovascular diseases	I60-I69	430-438	430-438
37	*Diseases of the respiratory system*	*J00-J99*	*460-519*	*460-519*
38	Influenza	J10-J11	487	470-474
39	Pneumonia	J12-J18	480-486	480-486
40	Chronic lower respiratory diseases	J40-J47	490-494,496	491-493,518
41	of which asthma	J45-J46	493	493
42	*Diseases of the digestive system*	*K00-K93*	*520-579*	*520-577, 444.2*
43	Ulcer of stomach, duodenum and jejunum	K25-K28	531-534	531-534
44	Chronic liver disease	K70, K73-K74	571.0-571.9	571.0-571.9
45	*Diseases of the skin and subcutaneous tissue*	*L00-L99*	*680-709*	*680-709*
46	*Diseases of the musculoskeletal system/connective tissue*	*M00-M99*	*710-739*	*710-738*
47	Rheumatoid arthritis and osteoarthrosis	M05-M06, M15-M19	714-715	712-713
48	*Diseases of the genitourinary system*	*N00-N99*	*580-629*	*580-629, 792*
49	Diseases of kidney and ureter	N00-N29	580-594	580-594
50	*Complications of pregnancy, childbirth and puerperium*	*O00-O99*	*630-676*	*630-678*
51	*Certain conditions originating in the perinatal period*	*P00-P96*	*760-779*	*760-779*
52	*Congenital malformations and chromosomal abnormalities*	*Q00-Q99*	*740-759*	*740-759*
53	Congenital malformations of the nervous sytem	Q00-Q07	740-742	740-743
54	Congenital malformations of the circulatory system	Q20-Q28	745-747	746-747
55	*Symptoms, signs, abnormal findings, ill-defined causes*	*R00-R99*	*780-799*	*780-782.3* *782.5-791* *793-796*
56	Sudden infant death syndrome	R95	798.0	-
57	Unknown and unspecified causes	R96-R99	798.1-9,799 799.0,2-3,5-9	795-796
58	*External causes of injury and poisoning*	*V01-Y89*	*E800-E999*	*E800-E999*
59	Accidents	V01-X59	E800-E928	E800-E929, E940-E942
60	of which Transport accidents	V01-V99	E800-E848	E800-E845
61	of which Accidental falls	W00-W19	E880-E888	E880-E887
62	of which Accidental poisoning	X40-X49	E850-E869	E850-E877
63	Suicide and intentional self-harm	X60-X84	E950-E958	E950-E959
64	Homicide, assault	X85-Y09	E960-E968	E960-E969
65	Events of undetermined intent	Y10-Y34	E980-E988	E980-E989

Part 3 Test deck

1 Uses of test decks

In the context of automated coding, «test deck» signifies a set of death certificate data to be used for testing different aspects of ACS performance. Typically, a test deck record contains both input data, which will be fed into the ACS to be tested, and the expected result of the ACS processing. For example, a test deck for the selection of the underlying cause of death will contain both information on the conditions reported on the death certificate, and the underlying cause of death that the ACS should arrive at if the selection and modification procedures work properly.

Evaluating ACS selection of the underlying cause of death is perhaps the most common use of a test deck, but test decks can be developed for assessing other aspects of ACS performance as well. For example, test decks are very useful in monitoring changes in complex language standardisation procedures.

If the test deck has been compiled according to an international standard, the extent to which an ACS meets that standard can be measured. Again, this will most often apply to the selection of the underlying cause of death. For example, an international test deck can be used to compare ACSs used by different countries. In this case, the same test deck is processed by two or more different ACSs. Any differences in the selection of the underlying cause are analysed, both between the ACSs involved in the test, and between the ACSs and the expected outcome as defined in the test deck.

A test deck can also be used to assess continuity when ACS is introduced, an existing ACS is radically changed, or replaced by another ACS. In most countries, the introduction of automated coding resulted in more or less pronounced changes in mortality statistics. It is well known that the ICD instructions on selection and modification of the underlying cause of death leave much room for national and even individual interpretations, and it is only to be expected that the introduction of an ACS will bring about changes in the statistics. This is especially the case if the ACS was developed in another country. Similarly, a changeover from one ACS to another may also affect the resulting statistics, since the new system perhaps applies the ICD instructions differently from the old one. In both cases, the magnitude and

nature of the changes can be assessed by processing test deck data through both production systems. If the aim is to compare manual coding with ACS, the test deck data should be coded both manually and by the ACS to be introduced. Obviously, the manual coders should not have access to the «expected outcome» stored in the test deck records, but only to the input data. When comparing two ACSs, the same test deck should be processed by both systems, and the differences between the systems' output compared.

Sometimes even regular maintenance of an ACS may have unforseen effects on the selection of the underlying cause. Most users of ACSs have found it wise to test their systems on a regular basis to make sure that periodic maintenance has not introduced artefacts into the production of mortality statistics. A test deck, that is used repeatedly over an extended period of time, can be used to this purpose. Here, the analysis focuses on changes in differences between the ACS output and the expected outcome in the test deck records.

2 Types of test decks

The end product of automated coding, the underlying cause of death selected for statistical tabulation, is dependent on the combined result of several steps in the ACS processing. First, the ACS interprets the text of the individual diagnostic statements reported on the death certificate, ICD codes are then assigned to these statements, the ICD codes are arranged according to the position on the certificate of the corresponding medical expressions, and finally the ACS performs the selection and modification of the underlying cause of death. Therefore, test decks should cover ACS performance in all these steps. However, interpretation of diagnostic statements, and assignment of ICD codes to diagnostic statements, are obviously language dependent, which the remaining steps are not. Therefore, it seems necessary to compile at least two different test decks: one for interpretation and coding of the individual diagnostic statements reported on the death certificate, and one for the selection of the underlying cause.

3 Developing the test decks

3.1 Developing a test deck for coding of diagnostic statements

Developing an international test deck for interpretation and ICD coding of diagnostic statements obviously poses some problems, since an international test deck should not be language dependent. However, even if the actual text strings to be tested in different countries cannot be stated, it should be possible to develop international specifications of the contents. These specifications should describe in fairly fine detail the types of problems likely to be encountered and that a text coding module must be able to solve. For example, the text strings finally used in the test deck should test the ability of the text coding module to:

Match the diagnostic expression to be coded to a dictionary
Standardise the input text string, including:
- Removing strings that do not influence the coding
- Replacing strings (synonyms, alternative spellings)
- Separating composite diagnostic expressions into the constituent parts
- Handling of etiological, semi-etiological and connective separators, including their effects on code placement
- Handling mentions of surgery and other medical treatment: select proper code for the surgery/treatment, and for complications, if necessary
- Other modification of ICD codes: duration, age, sex, other medical conditions, trauma/natural, etc.
- Validating ICD codes: codes consistent with age and sex, other codes, queries/confirmation required, etc.
- Formatting the output string according to specified requirements, e.g., ACME's.

Some actual examples could also be included in the specifications, even if they cannot be used for other languages. They would still be useful examples of the kind of problems likely to be encountered, and would give people developing test decks for their own language an idea of what kind of text strings to include.

In developing such specifications, close attention should be paid to existing national coding manuals for the interpretation of death certificate entries. Since national coding manuals have

been developed in response to the difficulties and problems encountered by the coders in each country, they will provide much information on the types of text interpretation problems that the text coding module will have to cope with.

3.2 Developing a test deck for the selection of the underlying cause of death

In contrast, the test deck for the selection of the underlying cause of death can be fully international. It can be based on ICD codes, and will thus be language independent. The ICD codes, however, must be arranged according to a specified standard, e.g., ACME's.

Developing this test deck would involve:

- Deciding on the size of the test deck, and designing a sampling strategy

  Test decks used to develop or monitor existing ACSs have typically been quite large. For example, the «test deck» used by Sweden in developing the text coding module MIKADO consisted of over 100 000 records, which is more than the annual number of deaths in the country. The test decks to be used for evaluation of the NCHS ICD-10 coding system will contain 150,000 records for testing MICAR, the text coding module, and 250,000 records for testing ACME, which selects the underlying cause of death.

  Large, statistically representative test decks have several advantages. They provide a good picture of the resulting mortality statistics, and will thus give early warnings of artificial trends introduced by the ACS. Also, the outcome is not too sensitive to the exact sampling procedure applied, as long as some kind of random sampling is used. There are also disadvantages, however. For example, it is a very heavy task to manually assign the «expected underlying cause» needed for the comparison to a very large sample. Consequently, there is a risk that the expected underlying cause code is sometimes erroneous. If random, non-stratified sampling is used, there is also a risk that rare, but epidemiologically important cases will not be represented.

  Also, using a large test deck will perhaps return a considerable number of cases in which the expected underlying cause differs from the underlying cause selected by the ACS, especially in the early stages of ACS development. If the resources available for the

development project do not permit an in-depth analysis of all the differing cases, important deficiencies in the ACS system may simply drown in the great number of cases.

Thus, the ideal size of the test deck and the sampling method to be applied depends on the aim of the test: if the primary interest is to detect statistical artefacts, a large sample based on simple random sampling will do very well; but if the main aim is to discover technical difficulties or the effects on specific conditions, a carefully selected smaller sample covering known selection problems could be used.

In 1999, France, Sweden and the United States performed a comparison of their underlying cause coding. The comparison was based on a sample of 6922 Swedish certificates and showed an overall agreement of 85.5% (ICD-10, four-character level) between the three countries. Possibly this sample could serve as a starting point for developing an international test deck aimed at identifying technical difficulties. However, the deck should be expanded by cases discussed by the Mortality Forum, or by the Mortality Reference Group. Countries that have introduced automated coding should also be asked to submit further examples.

- Assigning multiple cause codes according to international specifications

Once the certificates to be included in the test deck have been selected, multiple cause codes, that will serve as the input to the underlying cause selection step, must be assigned. Currently, the American software ACME is regarded as the international standard for the selection of the underlying cause of death, and all automated systems for the production of mortality statistics make use of either ACME itself, or of the ACME «decision tables» that govern the selection of the underlying cause. Consequently, it is necessary to arrange the input codes according to the ACME specifications. This does not mean, however, that the NCHS multiple cause coding instructions must be followed in every detail. Several of the NCHS coding instructions are primarily on the interpretation of death certificate entries, and other countries may feel that alternative interpretations would be quite as adequate. Such cases should be discussed by an international reference group, preferably the MRG («Mortality Reference Group»). The MRG, which was set up by the WHO Collaborating Centres for the Classification of Diseases in 1997, is

responsible for clarifications and updates to the ICD-10 as concerns mortality. The group is composed of international experts in classification and use of mortality data, and is the international group best qualified to make authoritative decisions on interpretation of death certificate statements.

- Selecting an underlying cause code manually for each case

Each record should contain a reference underlying cause, in other words the «correct answer» to the case. If the aim of the test is to discover artificial trends in the national statistics, obviously these reference underlying causes should be selected according to the method previously used in producing the statistics, and to which the ACS is to be compared. However, if the aim is to study the ACS's problem solving capacity, or to make international comparisons of how well ACSs currently in use correspond to the ICD instructions, this reference code should be selected strictly according to the instructions in the ICD. Previous decisions of the MRG must also be applied. As noted above, countries differ in their interpretation of the ICD instructions, and the reference codes should therefore be selected by an international group of coding experts. Cases in which the experts cannot agree should be referred to an the MRG.

In this context, it should be noted that the test deck developed by the NCHS uses reference underlying cause codes derived from the MICAR and ACME decision tables. However, international users of the NCHS system have suggested that these tables need revision, at least in places. Accordingly, an international test deck should not use reference underlying causes derived exclusively from the ACME decision tables.

4. Prerequisites and suggestions

Before international test decks can be developed, a number of practical issues must be addressed. First, it is quite obvious that the quality and usefulness of the test decks are dependent on the resources available for developing them. Therefore, the work cannot be done if adequate resources are not provided. Second, the composition of the expert group responsible for the reference coding is of crucial importance. To guarantee the widest possible international acceptance of the reference codes in the test deck, the expert group responsible for the coding must be generally recognized as the international elite in this field. Preferably, the test decks should be developed within the WHO-MRG framework, since the MRG is the only group authorized by the WHO to make binding decisions concerning the interpretation of the ICD as concerns the classification of causes of death.

It should be emphasized that automated coding of mortality data is an international, and not only a European concern. In 1996, the National Center for Health Statistics, USA, set up an «International cooperative effort» (ICE) on automated coding. The ICE has brought together people from all over the world who work with automated coding, which makes it the natural focal point for international initiatives in this field. The ICE steering group meets annually and also functions as an international reference group for further development of both American and other countries' automated coding software. Some of the European members of the steering group are also members of the Eurostat Task Force for mortality statistics. Therefore, a logical next step towards developing international test decks for automated coding could be to bring the matter to the ICE steering group. Several steering group members also participate in the work of the Mortality Reference Group and the ICD Update Reference Committee, which should give test decks developed on the initiative of the ICE the greatest possible international authority. The wide network of the ICE also seems the ideal way to distribute test decks to countries and individuals interested in automated coding. The test decks should also be made available on the Internet, for example on the ACS website currently being set up by the Australian Bureau of Statistics, or on the website planned for European users.

Part 4 Workshop on Automated Coding Systems

1. Background

Automated Coding Systems (ACS) are recognised to be the best way to assure reliability, homogeneity and comparability of the coding process for coding of causes of death. ACS allow to reduce the biases between countries and to assure greater comparability on statistics on causes of death. Many Member States (MS) consider ACS also as the most effective way to implement ICD-10 coding.

Building on the extensive experience gathered in the US on using ACS, several initiatives have been launched at European level and followed by the Eurostat Task Force on Causes of Death statistics (TF/CoD) :

- in 1997-1998 a Eurostat study was carried out on ACS, which resulted in as set of recommendations and guidelines for ACS design and use in the EU;
- a first pilot exercise between 3 countries was carried out in 1999: the results of manual coding in ICD-10 by Sweden were compared with the ACME coding in the US and the automated STYX coding system in France ;
- end 1999, Eurostat has launched a study to be carried out by INSERM and Statistics Sweden on preparing assistance for the implementation of AC in all EU MS and EEA/EFTA countries.

Since more and more MS are engaging in or planning to start with AC, the need for good collaboration at European level became apparent. This is why, in the framework of LEG Health, Eurostat and SC8-INSERM decided to organise this Workshop.

2. Objectives

The aim of the Workshop is to gather, on the basis of the experience of EU MS, US and other countries that are already using ACS, practical and reliable information and assistance which could be used by countries planning to use ACS. Another objective of this Workshop is to define a common approach for ACS at European level, and to define (a) most appropriate system(s).

The specific objectives of the Workshop are, that for the different steps in the procedure of using ACS (see below), MS :

- exchange their own experience gathered by using their software programme(s);
- indicate the key issue(s) to be considered for each of these steps and
- indicate possible ways on how to deal with these key issues, by focussing on how a common approach could be of benefit whilst taking into account the specific requirements of the country.

The steps in the procedure referred to above are :

1) data entry and language issues;
2) data throughput, i.e. how are the ICD coding rules applied in the processing of the data through ACS ;
3) evaluation and monitoring of the system used, i.e. with respect to cost/benefit, quality assurance.

3. Organisation

- The Workshop is structured along 7 sessions ; at the beginning of each of these sessions one or two participants (see agenda) will give a short introduction to start the debate; they will formulate the questions to be addressed by the working groups.
- All participants are expected to actively contribute by giving information on the experiences in their own region/country on the topics of the sessions. In order to facilitate the debate, most of the discussions, which are expected to be of a practical and technical nature, will be held in two smaller working groups.
- It should be noted that there will be no official papers but only working papers, nor will the presentations and contributions of participants have an official character.
- The draft agenda is not strict but there should be enough flexibility to adapt the time schedule according to emerging requirements.

4. Participants

15 participants were present in the workshop from 6 different countries. Marleen De Smedt and Jacques Bonte represented the European Commission. Donna Glenn from the NCHS, responsible for the development of the US automated coding system also participate to the Workshop.

Name	**Country**	**Organisation**
Mr Jean BOILEAU	France	CépiDc - INSERM
Mr Jacques BONTE		European commission - DG ESTAT-E3
Ms Silvia BRUZZONE	Italia	ISTAT
Ms Lois COOK	United Kingdom	ONS
Ms Marleen DE SMEDT		European commission - DG ESTAT-E3
Ms Donna GLENN	United States	NCHS - CDC
Mr Lars Age JOHANSSON	Sverige	Statistics Sweden
Mr Eric JOUGLA	France	CépiDc – INSERM
Mr Jan KARDAUN	Nederland	Statistics Netherlands
Ms Adele LOGAN	United Kingdom	Vital Events & NHS Branch
Mr Stephano MARCHETTI	Italia	ISTAT
Mr Gérard PAVILLON	France	CépiDc - INSERM
Mrs Gisèle RENAUD	France	CépiDc - INSERM
Ms Sue SMITH	United Kingdom	Office for National Statistics
Ms Lida SPRUIJT	Nederland	Statistics Netherlands

6. Agenda

For 3 days (26-28 March 2001), 7 sessions were organised in the office of CépiDc – Inserm in LeVésinet (France).

session 1: Actual situation of implementation and use of ACS in EU-MS and in the US
Participants are asked to give a short presentation about the situation in their country, focussing on :
- data entry : procedures, % of certificates coded by AC ;
- data throughput, i.e. software functions ;
- effects of the introduction of ACS on the data output (i.e. validation procedures in AC, effects on comparability over time) ;
- general evaluation of introduction of AC : cost/benefit, advantages and problems .
This first session should result in identifying the key issues to be addressed in the Workshop.

session 2: Special issues with respect to data entry
Experiences with incomplete certificates, changes in the certificate introduced as a result of using AC and inititiatives/plans for electronic certificates linking to AC

session 3: Data entry and language issues
Experiences with thesaurus prepared : number and types of diagnostic terms selected
Experiences with translations : i.e. which type of entries to be translated : complete diagnostic statements, synonyms, abbreviations -
Specifications of language-independent coding systems – intro by G.Pavillon ;

session 4: Data throughput
Multiple cause coding as entry in ACS and decision tables used for determining the underlying cause and/or for multiple cause statistics , possibilities for common coding system specifications - intro by L.Johansson ;

session 5: Assessment of the ACS

Performance assessment and quality control of the ACS, percentage of coding errors, effects on the output, exercises on re-coding (manual versus AC), bridge-coding knowledge base ; international comparability, test-deck for CoD coding software

session 6: Special topics
Natural versus unnatural death, stillbirths – perinatal deaths – old age

session 7: Blueprint for implementation of ACS
Key issues

7. Report

The report is based on the summary of each session.

Session 1

At first participants gave a short presentation about the experiences in their country with introducing AC either on a permanent basis, either on a trial basis. Some of the point raised in the discussions were:

- in most countries between 15 and 20 % of the forms are rejected at the level at data entry;
- there are different types of bridge-coding; either manual to manual from different classifications (ICD-9 versus ICD-10) either from automated coding (AC) to AC, either from manual coding to AC;
- most countries have different procedures of handling two streams of data: administrative/demographic data at the one side and 'medical' data at the other side;
- most countries emphasised the advantages of using AC for improving data quality; benefits in times of use of resources is less apparent.

Participants agreed that the key issues, that came out of the presentations, to be dealt with are in fact already introduced in the agenda:

1. borderlines and completeness (session 2 and 3);
2. special certificates and issues, including in addition: AIDS (session 6);
3. bridge-coding (session 5);
4. quality of coding procedures, including acceptable' error-rates for different aims (statistical versus research/ individual interests/data linkage) (session 4 and 5)
5. costs and benefits (could be treated in session 7).

Session 2

Introductions were made by G.Pavillon (on incomplete DC and electronic certification) and by J.Kardaun on scanning and Optical Character Recognition (OCR) and voice capture.

Results from the discussions of these two topics in the working groups are as follows:

- automated administration and demographic data systems can contribute substantially to an efficient AC system;
- AC could help on querying in two ways: 1) in selecting incomplete forms to be queried and 2) AC eases the administrative follow-up of the querying procedure;
- not all forms, liable for querying, can (not enough resources) or should be queried (some obvious underlying causes (UC) should not be queried anymore); as a consequence there is a wish expressed by the participants to establish 'common priority lists' in the EU for querying;
- when thinking about using electronic forms, on should consider this has implications:

1) on *legal aspects* (which may require some time to be resolved) such as acceptance of electronic signature, the procedure for dealing with the DC is often regulated by legal act;

2) on *technical aspects:* secure to be established, need for identification check of 'permitted' certifiers/physicians

3) all *interested parties* have to gain from it: physicians/ coroner/ registrar/ funeral director/ statistician; one weak actor in the chain can reduce overall benefits

4) it is not appropriate to use '*interactive'* forms can for filling in causes of death (CoD).

Session 3

Language independent aspects

The language independant part comes after the coding, i.e. mainly concerns teh selection of the underlying cause and multiple cause coding. Several elements of information are to be considered to multiple cause code correctly the conditions reported on the death certificate:

ICD codes of each separate conditions
relations between conditions
some classes of death
time indications
surgery indications
sex
age

Apart from sex and age, these pieces of information are strongly dependent on the language used but, if they can be translated in the same formal representation, the rest of the coding can be performed independantly of the language. These remaining steps are:

modification of codes according to sex and age and some other information
linkage between conditions
edits

The common implementation of these function would guarantee more compatibility between different systems. In addition, the input to the multiple cause codes in this formal representation should be stored for more sophisticated analyses since it contains more information than the only ICD codes.

As an introduction, L Johansson reported on experiences from Statistics Sweden's work with the Mikado dictionary. The main points were:

- Developing a dictionary is a major effort, much of the resources needed to build an ACS go into building and maintaining the dictionary

- Terminology on death certificates is not the same as in most readily available electronic lists (medical dictionaries, ICD Indexes)
- Consequently, a coding dictionary should be based on terms actually found on death certificates
- Start with the most frequent terms (a few very frequent terms will cover a large percentage of the terms to be coded)
- Check that epidemiologically important terms are handled correctly in the dictionary
- Decide early in the process on how to handle code modifications (generally, ACME requires that terms are coded without reference to other information on the certificate, but there are a number of important exceptions)
- Efficient language standardisation is crucial (most terms occur only once)
- Devise a way of keeping track of synonyms (a change to a term should also apply to all synonymous terms)

At present, the Mikado dictionary for ICD-10 has about 7000 expressions, and returns a match for about 90% of the terms on the death certificates.

Silvia Bruzzone presented the Italian work of translating the Micar dictionary. The entire dictionary (about 187 000 terms) has been translated into Italian. For ICD-9, the expressions on the Italian death certificate were processed to conform to English syntax. About 80% of the expressions found on the certificates are coded automatically, but external causes and AIDS/HIV cases are coded manually. For ICD-10, Istat will use a front-end software based on the Canadian general coding software ACTR.

Gérard Pavillon described the project in process dealing with "specifications of language independent coding systems". INSERM, in cooperation with Statistics Sweden, is currently developing specifications of language independent modules for language standardisation, code modification, and code verification (edits). All language dependent parts (medical terms, words and strings to delete or replace, separators...) are placed in tables that can easily be edited, while the program logic itself is not dependent on the language. In the prototype phase, the software is tested on French and Swedish certificates. Since there are many differences in language structure between French and Swedish, this will give a good indication whether it is possible to develop and use language independent software.

Language-dependent aspects

Edits (validation of input data) should be considered while building the dictionary. Language- or country-dependent aspects include:

- Social context of death certification (different social pressure on the certifier means that you need different warning systems in different countries)
- Possibilities to train people who report data (different needs for checks and edits)
- Layout and contents of the death certificate (if the sequence in Part I is bottom to top or from top to bottom, the fraction of handwritten certificates, typical errors, text editing - if any - before the text reaches the statistical office)
- Terminology and spelling (the same abbreviation may mean different things in different countries, the same term may have different meanings/uses, the meaning of a term may shift at different points of time)
- Syntax and grammar, hyphenated words, punctuation (lexically corresponding words may have different implications in different languages, punctuation marks are used differently, differencies in word formation - compounds or prepositions)

Session 4

Multiple cause coding (MC).

Uses of MC: the underlying cause (UC) alone is not always sufficient. When the death is due to several independant causes, more information must be obtained by the way of MC analyses. This situation is becoming more and more common. In particular:

- analysis of causes of death for elderly
- analysis of chronic conditions (e.g. diabetes)
- risk factors studies (alcohol, tobacco)

Sometimes the use of MC is also important to add more detail on the underlying cause:

- multiple injuries in violent deaths
- polydrug deaths

MC are needed to select all deaths with a specific disease (e.g. tuberculosis).

Some quality checks can be done only with the help of MC; the ratio between the number of mentions of a given disease and the number of mentions of the same disease as the UC can be used to highlight changes changes in coding and/or certification.

The only guide available in the ICD for MC coding is the index, which helps to combine conditions. Also some notes in volume 1 can be used. The TRANSAX system from NCHS is the only software available to perform this function. The experience of 2 countries Sweden and Italy suggests that some revision of TRANSAX is needed. Some measures were proposed to evaluate the performances of the system: comparison between TRANSAX output and manual MC coding; comparison between MICAR and TRANSAX output.

Session 5

Changes in trends and international comparisons

The need to prepare and inform the users of causes of death statistics to changes in trends was stressed. These changes are due to the ICD10 implementation and also to the implementation or the new versions of ACS. Several indicators can be used to measure these changes at differents level of detail. If the more general level should be widely published, the very detail level should be made available on request.

Bridge coding exercise are encouraged. The comparison of the result of these bridgecoding analyses at the international level will provide additional indications on the modification observed.

The reason for the differences highlighted by the indicators should be explained, in particular the changes due to classification or due to ACS should be identified. In particular multiple causes analyses can be very useful (e.g. changes between Viral hepatitis and cirrhosis in Italy).

Quality control

Quality control is needed:

- as continuing recoding of small samples of production certificates;
- for maintaining of a minimal level of manual coding expertise;
- after introduction of minor changes in the system;
- after introduction of major changes such as AC or when going from ICD-9 to ICD-10;
- for international purposes.

Various test decks are needed:

- for AC: text is used here for the test; this is a language-dependent exercise but nevertheless common specifications could be developed for a test deck;
- for MC: here a common test deck at EU level is possible; an important part of information could already be provided by MS already doing MC.

It was recommended that, when going to ICD-9 to ICD-10, countries should best aim for a 1-year 100 % bridge coding. Even without introduction of specific changes, it was recommended to do a bridgecoding exercise at regular intervals, i.e. every 5 years.

Session 6

Special Topics

Some types of death are difficult or impossible to code automatically, among these are perinatal and violent deaths.

Perinatal deaths. The countries using the death certificate form of WHO for perinatal death cannot select the underlying cause because this certificate does not provide the sequence of morbid events leading to death. This is the case for England and Wales. In France this certificate is used, however a sequence has introduced.
In all the case these deaths should be manually controlled because they represent a important public health issue.

Violent deaths. The circumstances of these deaths are often given in free text and therefore are difficult to code manually. This information is highly language-dependant. This leads to code manually these deaths. The US automatic system provides questions to help refining the coding. This avoid personal differences between coders.

Others causes of death need special attention: drugs related death, surgery, prothesis... However the problems are rather due to insufficient information at certification level. In this case the ACS can help querying the certifier.

Session 7

Key issue 1 - Data flow

- **Decentralised - centralised**

Is this different from asking if we are to have central European coding or coding in each country?

A decentralised organisation is more difficult to handle:

- version control
- technical support (software not always user-friendly: home-made front end systems...)
- comparable coding of rejects?
- regional groups too small to ensure efficiency?

However:

- current structures must be respected (some "regions" are bigger than many countries)
- quality assurance essential (as it is for comparisons between countries)

- **Data flow - single, double, triple**

ACS is NOT a complete mortality data production system. While the coding modules can (and should be common), the data management system must be different between countries, because we the data flow is different.

However: the ACS should include a description of the ACS input (standardised interface), for example:

- diagnostic terms
- surgery (if possible)
- sex and age

If you implement ACS, you will probably have to develop a new IT system. Pay attention to how to handle the queries.

- **Small countries**

Small countries may not afford to implement ACS. They could use:

- excess capacity in other countries
- European/regional coding centres

But: there might be legal obstacles

ACS might improve classification quality in countries with a small number of deaths, since their coders do not get enough experience of manual selection of the UC.

To consider:

- number of deaths
- number of people working with mortality statistics
- informatic skills (necessary knowledge to install and run the system)
- ACS provides extra detail on causes of death: how much do you need it, and how can you exploit it?

Key issue 2

- **Common EU dictionary**

To build a dictionary, a reference ICD is needed in order to assign the codes to the expressions. A national version or a close version.

At this point there are 2 possibilities to build the dictionary:

-start from MICAR dictionnary

-start from actual expression on death certificates

Countries with close languages should try to share experiences in building their dictionaries

Dictionaries should be designed according to the requirements of the software used.

- **Subdictionnaries?**

  Yes : violent deaths, perinatal causes, cancer by gender

- **Common language**

  Use of dict. as metadata

- **Formal/legal obstacles**

  WHO copyright: ask for sublicence to use for official statistical purposes

Key issue 3 - Quality control

- **ACME as reference - clearing house**

There is already a procedure in place: According to a recommendation of the ICE on automated coding, issues concerning the decision tables are referred to the Mortality Forum, and from there (if no consensus) to the Mortality Reference Group. The NCHS will implement changes decided by the MF or MRG.

- **Bridge coding**

It is absolutely necessary to bridge code when you implement automated coding. Suggestion: use the last year with manually coded data, recode it by ACS.

Future bridge codings:
- new revisions of the ICD
- other substantial changes to the system

The net effect of the changes is the only interest of most epidemiologists.
The influence of each separate change (new decision tables, new ICD instructions, changes in death certificate form) must be analysed - the statistical office will need to know this.

With continuous updates to the ICD-10, bridge codings will be needed at regular intervals.

- **Quality checks**

Parallel automated-manual coding:
- to maintain knowledge of selection of the UC
- to monitor the ACS

Reject coding is not enough to maintain knowledge in UC selection, because of the bias (may not be possible to code a random sample in every country).
ICE recommendation: code 300-400 certificates manually a week (also not possible everywhere)
ACS will NOT improve data quality, only consistency. You need other data quality measures, these should be in place (or at least designed) before you implement automated coding.

Consider implementing audit trails to assist in identification when a record has been manipulated.

Key issue 4

- **Transax and multiple pathologies**

TRANSAX tables available in Sept 2001

Studies?

Depends on availability of ICD10 data. Follow implementation of ICD10 coding and ACS (2003?)

Key issue 5

- **Speed MS**

For those who have ACS. We need some kind of network, partially electronic, supplemented by face-to-face meetings.

For those who are interested. Each country must decide on the participants, we would like see both decision makers and technical people. Make an inventory: which countries? Who is to take care of them? Try to attract the "right" kinds of people.

For those who are not converted. They might be easier to persuade if we can show an open communication with the NCHS. We realize the NCHS has not the resources for this at present. What could we do to support the NCHS? How to use the Australian bulletin board to best effect?

- **How to learn from each other**

Updated databases on

- ACS
- bridge coding data

Personal contacts and working visits

- **Info session**

Ask the initial expectations and concerns of the countries interested by info on ACS content:

- should cover issues treated in the sessions, extended demonstrations in small groups
- share the topics of the program between the experienced countries
- participants / countries interested

- **Time Plan**

  ICE plenary meeting in 2002 to be considered

  Info session date end of 2001?

  Implementation target date 2005

Bibliography

[Anderson01] Robert N. Anderson et al, Comparability of Cause of Death between ICD-9 and ICD-10: Preliminary Estimates, National Vital Statistics Reports, Vol 49, Number 2, May 18, 2001

[Eurostat99] G. Pavillon, M. Coleman, L.A. Johansson, E. Jougla, J. Kardaun. Final Report on Automated Coding in Members States. Eurostat Working Papers - Population and Social conditions, 3/1999/E/n°1, Doc. OS/E3/98/COD/3 en, 1999

[Eurostat01] Final Report on Comparability and Quality Improvement of European Causes of death Statistics. Eurostat, *Forthcoming*

[Hatton96] Hatton et al, Bridge Coding between ICD9 and ICD10 on 5,179 cases, World Health Organisation, Meeting of Heads of WHO Collaborating Centres for the Classification of Dieseases, WHO/HST/ICD/C/96.30, 1996

[ICD92] International Classification of Diseases and Related Health Problems, Tenth Revision, World Health Organisation, Geneva, 1992

[ICE99] Proceedings of the International Collaborative Effort on Automatic Mortality Statistics, Volume 1, Kimberley Peters, Editor, 1999

[Johansson00] Lars A. Johansson, Mortality Bidge Coding ICD9/ICD10: Preliminary Results from a Statistics Sweden Study, World Health Organisation, Meeting of Heads of WHO Collaborating Centres for the Classification of Dieseases, WHO/HST/ICD/C/00.66, 2000

[Jougla98] Eric Jougla et al, International Comparability of Causes of Death Data Inside the European Community, World Health Organisation, Meeting of Heads of WHO Collaborating Centres for the Classification of Diseases, WHO/HST/ICD/C/99.44, 1998

[NCHS01] Instruction for Classifying Multiple Causes of Death, Instruction manual Part 2b, Centers for Disease Control and Prevention / National Center for Health Statistics, 2001

[Oomi96] Kenichi Oomi et al, Report on Change in Vital Statistics of Japan, World Health Organisation, Meeting of Heads of WHO Collaborating Centres for the Classification of Dieseases, WHO/HST/ICD/C/96.36, 1996

[Rooney97] Cleo Rooney, ONS Bridge Coding Study of the Change from ICD9 to ICD10 for Mortality Data in England and Wales, World Health Organisation, Meeting of Heads of WHO Collaborating Centres for the Classification of Dieseases, WHO/HST/ICD/C/97.50, 1997

[Santo97] Augusto H. Santo, Bidge Coding between ICD9 and ICD10: underlying and Multiple causes of death, Preliminary Report, World Health Organisation, Meeting of Heads of WHO Collaborating Centres for the Classification of Dieseases, WHO/HST/ICD/C/97.61, 1997