Decision tables for mortality coding: methods and tools for the management and documentation of changes

Simone Navarra [1], Marisa Cappella [1], Lars Age Johansson [2], László Pelikan [3], Friedrich Heuser [4], Luisa Frova [1], Francesco Grippo [1]

Abstract

Decision Table Editor (DTE) is a web-based system developed by Istat in the framework of an international collaboration (Iris Institute). By means of this application, experts from different countries can collaborate on the coordinated and simultaneous maintenance and update of the decision tables used for the underlying cause-of-death selection. These tables provide criteria for the correct application of the selection rules of the International Classification of Diseases (ICD10), published by World Health Organization (WHO) and periodically updated. They derive from those originally developed by the US National Centre for Health Statistics (NCHS) for the ACME software and represent a major tool for enhancing the international comparability of mortality statistics. One of the major achievements of the DTE is the improvement of transparency and documentation of changes introduced in the tables which have a direct impact on mortality statistics.

Keywords: Mortality coding, ICD10 updates, decision tables, Iris.

1 Italian National Institute of Statistics - Istat.

2 Swedish National Board of Health and Welfare.

3 Hungarian Central Statistical Office - KSH.

4 Federal Institute for Drugs and Medical Devices - BfArM (formerly, German Institute for Medical Documentation and Information - DIMDI).

The views and opinions expressed are those of the authors and do not necessarily reflect the official policy or position of the Italian National Institute of Statistics – Istat.

1. Introduction

Decision Table Editor (DTE) is a web-based application for the maintenance of the decision tables used for the selection of the underlying cause of death (UC). It has been developed by the Italian National Institute of Statistics - Istat in the framework of the collaboration with the Iris Institute which emerged from an international cooperation for the deployment, maintenance and development of the Iris software, an electronic system for automated coding of causes of death. The Institute is hosted at DIMDI (German Institute for Medical Documentation and Information) and the current cooperating partners are statistical institutions from France, Germany, Hungary, Italy, Sweden and United States (Iris Institute website www.iris-institute.org). Istat officially joined the group by means of an agreement with the DIMDI signed in 2012. Nevertheless the collaboration of Istat with the other European partners for the development of Iris software had begun two years earlier.

The decision tables are central to the function of Iris. The tables are primarily used by Iris software but they also constitute a support for manual coding and represent the knowledge base for the consistent and harmonised application of the international rules for the selection of the UC according to the provisions of the International Classification of Diseases and Related Problems, tenth revision (ICD10), published and revised by the World Health Organization (WHO, 2010). These tables make it possible to apply these rules by computer programmes and by coders with limited medical experience.

The knowledge database was first developed by the NCHS (US National Center for Health Statistics) for the ACME system (ACME tables). Successively, since 2011 it has been maintained by the Iris Institute for the inclusion of the annual WHO official updates of the ICD. Hence, the tables used by Iris differ by some extent from the NCHS ones (CDC, NCHS, 2016) for the inclusion of updates since 2010 on.

DTE is also accessible to the general public for downloading the decision tables in pdf format at the web-address www.iristables.istat.it.

2. Selection of the underlying cause and harmonised statistics[5]

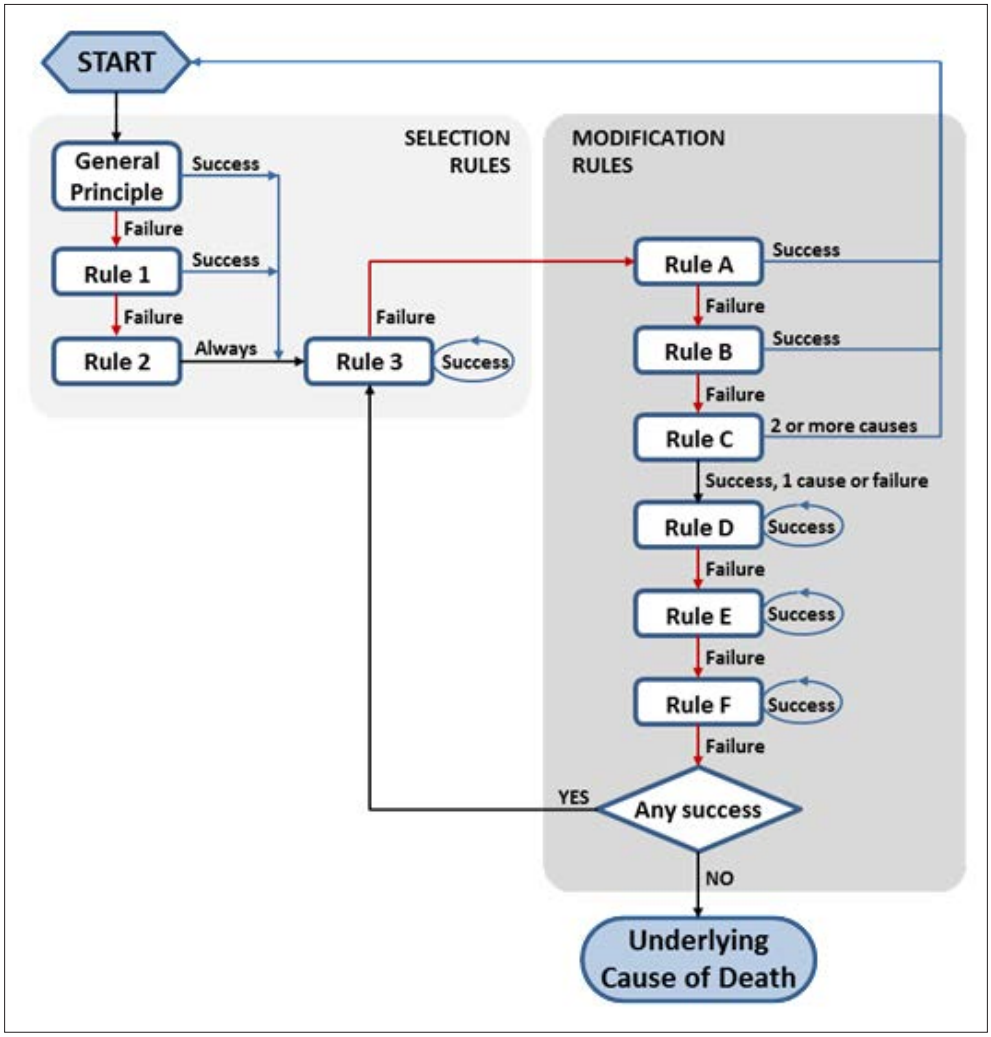
Comparison of mortality statistics are mostly based on the underlying cause of death. This is defined by the WHO (2010) as "(a) the disease or injury which initiated the train of morbid events leading directly to death, or (b) the circumstances of the accident or violence which produced the fatal injury". For each death, the UC is selected from an array of conditions reported by a physician on the death certificate through the application of the selection and modification rules contained in the appropriate revision of the ICD. Selection rules included in ICD are meant to be a systematic guidance for selecting the UC, thus ensuring comparability and uniformity in mortality statistics among different countries. Figure 1 represents a simplified schema of how the rules apply during the selection (WHO training tool). Although some details of the selection process are left out from this Figure, it clearly shows that the selection process can be seen as a complex algorithm with several decision nodes. The criteria for determining the success or failure of each node are described in specific instructions included in the volume 2 of the classification or by other provisions such as the inclusion/exclusion notes of the tabular list and the alphabetical index.

The procedures for selecting the UC imply two main steps. In the first, the selection is finalised to identify the antecedent originating cause which is the starting point of the sequence of events leading to death. This step primarily involves the application of General Principle or Rule 1 or 2. For the application of these rules, the sequence reported by the physician on the death certificate must be examined in order to evaluate its correctness. The classification provides instructions on sequences to be accepted and those to be rejected. After one of these rules, Rule 3 is applied, in order to evaluate if the cause identified in the previous steps can be considered an obvious consequence of another condition reported. Also in this case the ICD provides instructions for detecting obvious consequences. In the second step of the coding process, a modification of the selected cause is performed. This step is finalised to select a more informative condition if the first selected is an ill-defined (Rule A) or trivial affection (Rule B); to combine information reported

5 This description and furher parts of this paper refer to the 2010 edition of the ICD10. Although in 2016 the rule application schema and the name of the rules were deeply revised, the content of this paper remains still valid and applicable to the new framework of the ICD10 rules.

in different parts of the certificate (Rule C, linkage); to select a more specific condition (Rule D). This modification allows selecting a more informative condition for public health purposes.

Figure 1 - Coding rules and coding algorithm (a)



(a) This Figure is an adaptation of the flow chart included in the WHO training tool "ICD10 Interactive self-learning tool" available on the WHO website. It reflects ICD10 instructions until 2010. Although it leaves out some details of the selection process and does not contain special instructions such as surgery and procedures, it shows the complexity of the coding. For a complete and up-to-date information on this topic refer to ICD10 volume 2 and its updates on the following link http://www.who.int/classifications/icd/en/.

2.1 Automated coding and the Iris system

The international rules and instructions for the selection of the UC, leave space for interpretation, resulting in a certain degree of variability of the tabulated UC among coders (Harteloh et al., 2010) and, thereby, across countries. The interpretation derives from both the complexity of the algorithm and the criteria for decision making in each node. In order to face the problem, since the 1960s, US National Center for Health Statistics (NCHS) has been the major investor in the research and development of an automated mortality coding system and in 1968 developed the Mortality Medical Data System (MMDS) for the coding of both the UC and multiple causes on the death certificate (CDC website, about MMDS). MMDS consists of two main components: MICAR (Medical Information, Classification And Retrieval) and ACME (Automated Classification of Medical Entities). MICAR module assigns an ICD code to each condition reported generating the input for the ACME module which then, by using the set of logical decision tables, applies the international selection and modification rules, resulting in the selection of the tabulated UC.

A number of European countries implemented MMDS in the '90s of the last century. In some of them automated systems in languages other than English have been developed using the ACME decision table logic (Pavillon et al., 1999). France and Sweden in particular, started to cooperate on a common tool thanks to the experience of these countries in the use of automated coding. Successively, Germany joined the project and finally, in order to improve the international comparability of mortality statistics, Eurostat (the statistical office of the European Union) supported the development of Iris, a common, language-independent coding system that can be used for coding death certificates, written in any language, according to ICD coding rules and guidelines for the selection of the UC (Pavillon et al., 2007, Pavillon 2012). Version 4 of Iris uses MMDS components while version 5 contains a newly developed module, MUSE (Eckert, 2014).

3. Decision Tables for the selection of the underlying cause

The decision tables represent the knowledge base for the coding, both manual and automated, which allow taking decisions for every step of the coding algorithm represented in Figure 1. They are a formalisation of the instructions included in the volume 2 of the ICD10. This formalisation basically consists in the translation of the provisions of the Classification into relationships between pairs of ICD codes.

The tables were first developed by NCHS as part of ACME and are still released on official website as Part 2c of the Vital Statistics Instruction Manual series (CDC, NCHS, 2016). Nevertheless, when Iris was developed, some changes in the tables were needed in order to fit the specificity of the new software and also for including some official WHO updates. Despite these changes the Iris tables maintained the same structure as the NCHS ones. In Figure 2, an extract of Iris 2014 tables is shown (print version).

The Iris tables can be summarised as follows:

- valid codes table (corresponds to the NCHS tables A, B, C, G and H), includes the list of the ICD10 codes with the description of the properties of each code for mortality coding purposes. Certainly, not all the codes reported in the ICD are used for mortality coding and some of them are not used for the UC coding, but they can be used for multiple causes. Therefore, code validity, for both multiple and UC coding, is documented in the table as well as other flags informing on other characteristics such as: ill-defined condition activating rule A; trivial affection which activates Rule B; created code and, for these, the correspondence with the ICD10 valid codes used for data tabulation (NCHS Table G). Created codes are special codes not included in the ICD, used for capturing information contained in the diagnostic term, which is necessary during the coding process. In some cases, the regular ICD10 code is not sufficient for describing such detail indeed. As an example, the code A16.9 is used for coding both diagnostic expressions “tuberculosis” and “respiratory tuberculosis”. Nevertheless these two expressions can have a different behaviour during the UC selection. In order to distinguish between these two situations, the table includes the plain code A16.9 for coding “respiratory tuberculosis” while the

created code A16.90 is used for the term “tuberculosis” without other specifications.

- causal table (NCHS Table D), contains the accepted causal sequences and it is used for the application of General Principle, Rule 1 and Rule 2.
- modification table (NCHS Table E) lists modification relationships between codes. Various relationships can exist between two codes according to the reference rule. It represents the main guidance in application of Rule 3, and modification rules A, C and D.
  Both causal and modification tables contain ambivalent entries also indicated as “maybe” relationships. The maybes are generated by the fact that the ICD codes are used for coding broad groups of specific conditions while causal and modification relationships might involve only subsets of these. In these cases the UC selection depends on the analysis of the text reported by the physician and must be manually revised according to the explanation reported in the text next to the relationship involved. In NCHS tables the maybe explanations are included in a separate Table F. The maybe explanations are provided only for the modification table. As discussed previously, the created codes are used as well in these situations, with the advantage of allowing these cases to be automated processed.

In general, the causal and modification tables have a common structure and can be seen as a single component. Nevertheless, for practical reasons, they are generally presented as separate tables. Actually, the causal and modification tables are used in two separate moments of the coding process, first when applying the selection rules and second during the modification.

Figure 2 - Decision table structure

VALID CODES TABLE

Code	Ill-defined	Trivial	Created	Code conversion	Validity
A000	No	No	No		Valid for multiple and underlying
A001	No	No	No		Valid for multiple and underlying
…					
A169	No	No	No		Valid for multiple and underlying
A1690	No	No	Yes	A169	Valid for multiple cause only
…					
F03	No	No	No		Not to be used if underlying condition is known

CAUSAL TABLE

---E140-E141---			---E140-E141--- Continue		---E140-E141--- Continue
B252			K850	-K861	Y525
B263			K868	-K869	Y527
C250	-C259		M359		Y543
C788		M	O244		
D136	-D137		P350		---E142---
D350			Q871		
E050	-E69		Q900	-Q909	B252
…			…		…

MODIFICATION TABLE

---D739----

SMP	C261		
SMP	C788	M	Suba must be spleen
DS	C810-C969		
SMP	D139	M	Suba must be spleen
SMP	D377	M	Suba must be spleen
SMP	D730-D378		

Figure 2 shows the tables as they appear in the paper-based format, where the causal and modification tables are separate sets. On the other hand, Prospect 1 describes the variables of the tables as they were a single set and provides a short description of the variables included.

The causal and modification tables contain address and subaddress codes. The address is either a single 3-5 digit code or a span of codes enclosed in dashes (e.g. “---E142---” is a single code, “---E140-E141---” is an interval of codes). The subaddress is given under the address and may also consist of a single code or a span of codes. Note, for instance, that the span E050-E69 includes all the valid codes from the valid codes table from E050 to E69. In the modification table the following acronyms precede each subaddress indicating the relationship with the respective address and designating the applicable rule: DS, DSC, IDDC, SENMC, SENDC, LMP, LMC, LDP, LDC, SMP, SMC, SDC. In some cases an additional code is reported on the right of the subaddress (not shown in the Figure). This code, referred as recode,

identifies a code resulting from the combination of the tentative UC (address) and another code on the death certificate (subaddress). Table D contains just one type of relationship between address and subaddress so the acronym is not reported but it is understood as DUE. The symbol "M" is used in both table D and E to denote ambivalent (maybe) relationships. Reasons to these ambivalences are displayed next to the "M" and provide further guidance in the selection of the most appropriate UC. For some cases special attention is required when applying a rule. These entries are flagged with a symbol "#" (not shown in the Figure).

Prospect 1 - Variables of causal and modification tables and types of relationships (a)

Variable	Modality	Description	
Address	A000-Y98	Also referred as anchor code or simply code. It is the tentative UC resulting from the selection process. It can be represented as a single code or as a span of codes (address1-address2).	
Subaddress	A000-Y98	Also referred as subanchor code or subcode. It is another code present on the death certificate. It can be represented as a single code or as a span of codes (subaddress1-subaddress2).	
Relationship		Also referred as rule, is the type of relationship that links address and subaddress codes and indicate which ICD10 rule is applicable.	
	DUE	Due to	General Principle, Rule 1 and 2
	DS	Obvious consequence	Rule3
	DSC	Obvious consequence with combination	
	IDDC	Ill-defined, in due to position with combination	Rule A
	SENMC[b]	Senility, in mention position with combination	
	SENDC[b]	Senility, in due to position with combination	
	LMP	Linkage, in mention position with preference	Rule C
	LMC	Linkage, in mention position with combination	
	LDP	Linkage, in due to position with preference	
	LDC	Linkage, in due to position with combination	
	SMP	Specificity, in mention position with preference	Rule D
	SMC	Specificity, in mention position with combination	
	SDC	Specificity, in due to position with combination	
Recode	A000-Y98	Is the code resulting from a combination of the address and subaddress, when modification rules are applied for the relationships DSC, IDDC, SENMC (b), SENDC (b), LMC, LDC, SMC, SDC.	
Maybe flag	M	Indicates ambivalent relationships: entries with ambivalent relationships are flagged with the letter "M". Both causal and modification table contain ambivalent relationships, but explanation are provided only for those in modification tables.	
Maybe reason	Free text	Shows the reason for ambivalent relationship. Reading the reason, the coder can decide if the relationship expressed in the entry is applicable.	
Special note	#	For some cases special attention is required when applying a modification rule. These entries are flagged with a symbol "#". This field is also referred as "neocode".	

(a) As causal and modification table share the same structure, they can be considered as an unique body. The causal table contains a single relationship which is "DUE". All the other rules refer to the modification tables.
(b) From the 2016 edition of the tables the rules SENMC and SENDC have been deleted. For SENDC the IDDC rule has been used, the new rule IDMC has been created to substitute SENMC and for other uses as well.

The structure shown in Figure 2 refers to the compressed format where relationships between ICD codes are represented, when possible, as intervals of codes. This representation is necessary in order to make paper-based tables more readable to coders. In this compressed form the tables includes more than 94,000 rows (2014 version). However, the relationship between intervals of codes is a synthetic representation of all the relationships between single pairs of codes. When the relationships between intervals of codes are resolved into relationships between single pairs of codes, the number of relationships expressed in the table account for more than 31 million. The tables in which the intervals of codes are resolved are referred as normalised tables. Table 1 shows the comparison between the compressed and the normalised structure of the tables.

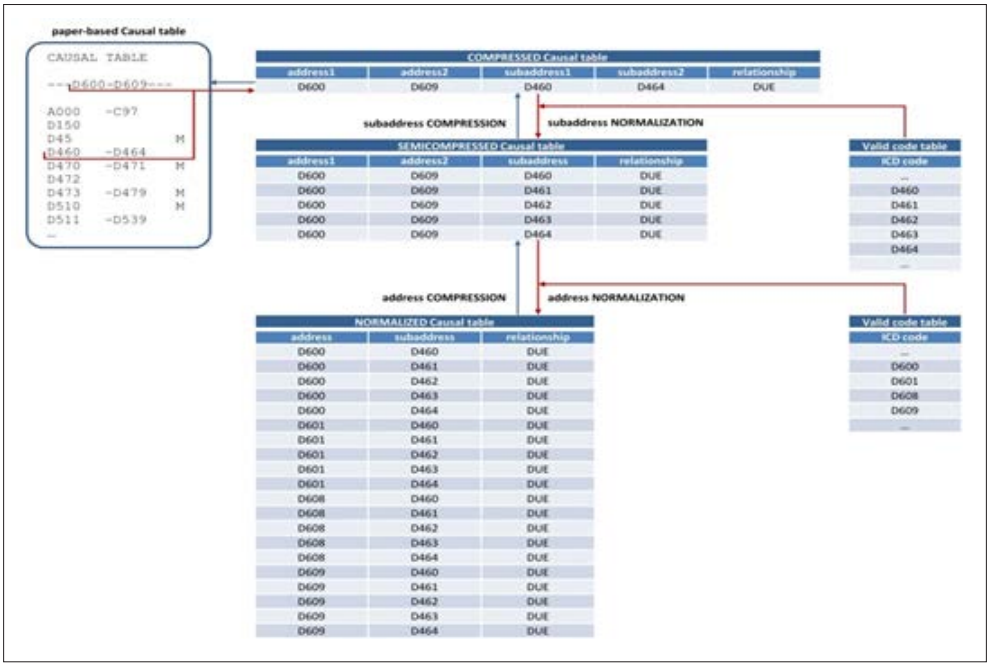
Table 1 - Compressed and normalised structure of the tables (2014 edition)

	Table D		Table E	
	Compressed	Normalised	Compressed	Normalised
Total rows	57,844	29,677,852	36,566	2,127,820
of which:				
rows with maybe	20,815	1,375,960	8,835	86,300
recode required (a)	-	-	14,496	98,756
other notes (a)	-	-	439	13,432

(a) Not applicable for causal table.

In Figure 3 an example of the normalisation procedure for a given row of the causal table is provided. To make this normalization, both address and subaddress intervals are resolved into single codes depending on the list of valid codes. The Figure shows how from a single row representing a “due to” relationship between two intervals of codes, 20 normalised rows are obtained: the product between 5 codes in the address interval (D600-D609) and 4 codes in the subaddress interval (D460-D464). Normalisation is a reversible process. Normalised tables can be compressed back to the non-normalised format through the compression procedure which is the inverse of normalisation.

Figure 3 - Normalisation and compression procedure



3.1 ICD updates and table editing

Maintenance of the decision tables is necessary for the up-to-date and the correct functioning of the Iris software. Actually, any change in the tables affects the result of Iris coding. Table maintenance consists in the annual revision in order to fulfill two needs:

- to correct errors such as incorrect or missing causal relationships or linkages;
- to apply the WHO official updates (WHO website, list of official updates). Actually, modifications in the ICD, its rules or their interpretations are implemented by editing the appropriate decision table.

It is convenient to remark that, although the editing of the tables is performed by cooperating partners of the Iris Institute, it strictly depends on decisions taken at international level and in particular it is performed, as much

as possible, according to the process of ICD updating. This process involves different organs within the net of the Collaboration Centers of the WHO for the Family of International Classifications (WHO-FIC). The official updates to the ICD10 are approved at an annual meeting by the Update and Reference Committee (URC) and published on the WHO website in the format shown in Figure 4. For the mortality application, a specific organ of the WHO-FIC called Mortality Reference Group (MRG) functions as a consulting body. The scope of the MRG is to improve the international comparability of mortality data by making decisions on coding issues, suggesting clarifications of coding instructions as well as other changes to the ICD10. This organ is also helped by a more practical group (Table Group) that recommends changes to the tables.

Figure 4 - WHO official ICD10 updates. Extract from the "Cumulative official updates to ICD10 of volume 2" available for download in pdf format on WHO official website

Instruction	Instruction manual entries	Source	Date approved	Major/ Minor update	Implementation date
Move location of sequelae of TB and add mention of chronic forms of hepatitis to section 4.2.2 of ICD-10 volume 2	**4.2.2 Accepted and rejected sequences for the selection of underlying cause of death for mortality statistics** … (a) Infectious diseases The following infectious diseases should not be accepted as due to any other disease or condition, except when reported as due to human immunodeficiency virus [HIV] disease, malignant impairing the immune system: • typhoid and paratyphoid fevers, other salmonella infections, shigellosis (A01-A03) • tuberculosis (A15-A19) • sequelae of tuberculosis (B90) The following infectious and parasitic diseases should not be accepted as due to any other disease or condition (not even HIV/AIDS, malignant neoplasms or immunosuppression): • cholera (A00) • botulism (A05.1) • plague, tularaemia, anthrax, brucellosis (A20-A23) • leptospirosis (A27) • tetanus, diphtheria, whooping cough, scarlet fever, meningococcal disease (A33-A39) • diseases due to Chlamydia psittaci (A70) • rickettsioses (A75-A79) • acute poliomyelitis (A80) • Creutzfeldt-Jakob disease (A81.0) …	MRG 1798	October 2011	Minor	January 2013

4. Decision Table Editor

4.1 Objectives

Since 2011 Iris Institute has updated and maintained decision tables taking into account the annual provisions of the WHO, even if not all the updates have been fully implemented. Updating process originally adopted was based on a spreadsheet structure. The major limits of this kind of tool were, first of all, a limited possibility to trace and retrieve changes introduced in the tables, especially for documenting the rationale of the changes. Second, it implied a significant manual intervention, increasing the chance of error. Certainly, the complexity of the tables shown above makes the table editing not a trivial task. For instance, the compressed format of tables D and E complicates data manipulation because, generally, the updating requires the disentanglement of many code intervals. Moreover the high interrelation existing among the relationships included in the tables implies that changes in one relationship can have impact on many others. Third, it did not allow the simultaneous work of different experts: updates could happen only in series but not in parallel. For all these reasons, it was essential to develop a reliable system for the annual table updates, as little dependent on direct manual intervention as possible.

To respond to the need of a continuous table updating, the Italian National Institute of Statistics - Istat, in the framework of the agreement with Iris Institute, developed the Decision Table Editor (DTE) web application. DTE is an online work platform conceived to allow international coding experts to cooperate in maintenance, production and distribution of the decision tables. DTE is therefore designed as a work and production environment rather than a mere instrument for table consultation, although data retrieval features are available for internal users.

In summary, the objectives of the DTE are:

- to handle simultaneous and coordinated access to the tool of experts from different countries for updating decision tables;
- to document the annual updates;
- to check for duplications and inconsistencies;
- to avoid manual intervention on the tables;

- to produce the decision tables used by both Iris system and manual coders;
- to store, retrieve and browse annual versions of the tables.

4.2 System overview

The system is a Java web-based application which allows managing the updating process of the decision tables. In particular:

- editing:
  - valid ICD codes;
  - decision tables;
- validation and production of annual tables;
- management of primary tables;
- browsing and downloading.

The management of the system functionalities is performed by means of a very user-friendly interface. The database of the application has been designed in Oracle and comprises of two main data groups:

- the first group is the data storage of the historicised decision tables;
- the second group is designed for recording the changes required by annual updates, and can be considered a data flow recording.

The storage group contains tables for valid codes, decision tables (both causal and modification are stored in the same table) and maybe reasons. The information of the decision tables is kept in a normalised form, as described in Figure 3, i.e. the relationships kept in the tables refer to pairs of codes and not to intervals. This way of storing information, although highly memory consuming (more than 31 million rows are needed), facilitates the updating procedures and makes data retrieval more flexible. Moreover, each record contains fields for both start and expiration year as well as a reference to the reason for the change (Id of the update giving rise to the starting or expiration), making possible to store and retrieve all annual versions of the tables and the origin of change (historicisation).

The data flow group contains the information that the system uses for making the changes to the tables according to each implementation year.

4.3 Collaborative, coordinated and controlled workflow

The first problem encountered in designing DTE, was the need for a definition of a rigorous workflow for the table updating. In this paragraph the flowchart of the updating workflow is described.

Different profiles are designed for different tasks and, in order to trace all the activities performed on DTE, access to system requires username and password and implies a three-tiered permission architecture.

The three internal user profiles are: Administrator, Supervisor and Editor.

The implementation of annual WHO updates, as well as correction of errors, consists in the addition of new rows and modification or deletion of existing rows from the tables of the previous year. Nevertheless, with the DTE, these modifications are not directly performed on the tables but are inputted in a specific encoding panel and successively applied to the tables by the system itself. Every change in the tables is maintained in order to track and retrieve different annual versions. This updating process is designed for releasing and storing a single annual version of the tables in the database. Changes can be made several times in a given year but only one final edition is kept.

The complete workflow is represented in Figure 5 and it is described below.

Phase 1. Data input: update definition and check

Editors are involved in phase 1 of the process. Their main task is to insert data derived from the agreed updates to be implemented in the year. DTE system is designed to manage simultaneous access of different editors. Furthermore, when one or more editors work on data entry, changes to database are univocally identified allowing to trace the source of any modification and the operator who made it. The detailed steps of this phase are the following:

- *Updating of valid codes table.* Definition of expiring date for expired codes, addition of new codes, modification of attributes (trivial, ill-defined, etc.). Each modification in the table must be documented in an appropriate field with reporting the rationale and source.
- *Update definition.* For each given year, the list of updates impacting on causal and modification table is defined with the description of the rationale and source. This task is reported in a specific input panel (Figure 6, upper part of the update input panel).
- *Encoding.* For each update, editors enter in the lower part of the input panel (Figure 6) the rows of causal and modification tables which must be deleted, added or updated according to the instructions reported in the upper part of the panel. During data typing the system performs online check of the input.
- *Check "within".* After the encoding is completed, a check is run in order to identify possible inconsistencies among encoding rows referring to the same update. In order to carry out this quality control, the system performs normalisation of the encoding (Figure 3). From this point onward all check procedures operate on normalised tables. When this check does not find errors, the normalised encoding rows are stored in a table called update table. Updates will be applied to the historicised causal and modification tables in a later stage.

Phase 2. Table production

This phase is coordinated by the supervisor, who runs the check and updating procedures. In the updating procedures the changes described in the normalised update table are applied by the system to the historicised tables. Before changes become effective, test tables are produced.

- *Check "between".* Update table produced in the previous phase comprise updates entered by different editors. This implies that updates may be incompatible with each other. In order to identify these errors, the supervisor runs the encoding check "between" procedure. This could produce errors that must be revised manually by the supervisor through the correction panel #1. The procedure cannot proceed further until these errors have been corrected.

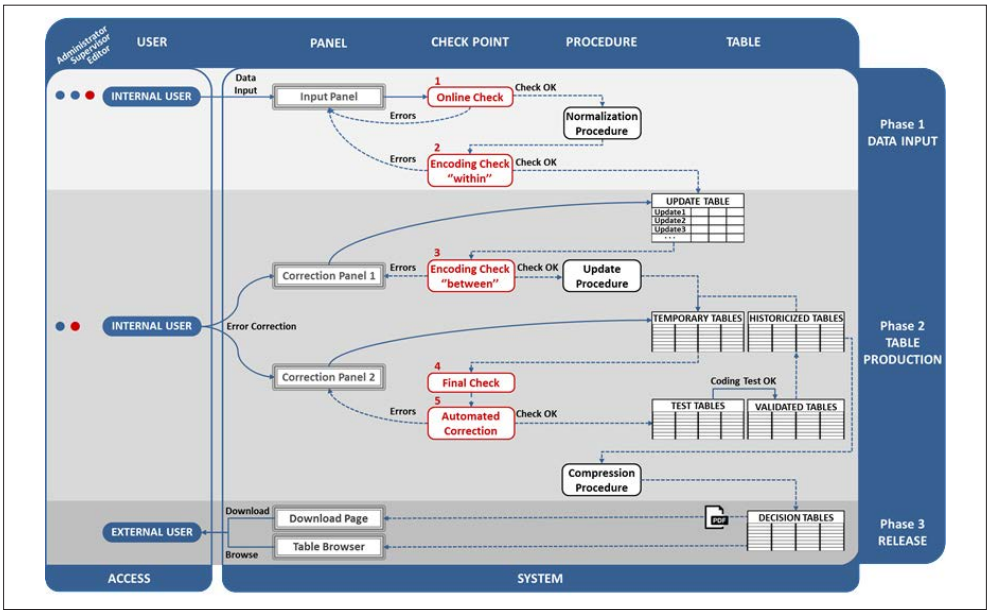
- *Update procedure.* This procedure compares the update table with the historicised tables, applies changes according to the actions specified by the editor and produces a temporary set of updated tables (temporary tables). These must undergo the validation steps described in the following bullets before making the changes effective and stable.
- *Final check and automated correction.* This procedure checks for inconsistencies in the temporary tables. Most errors require a deterministic correction and they are automatically corrected. Others are displayed and must be revised manually by the supervisor through the correction panel #2. The presence of errors stops further processing.
- *Coding test.* Once the updated tables are free of errors, the supervisor can download tables (test tables) for running coding tests that would show the impact of the updated tables on the data and identify possible errors occurred during the update;
- *Validation.* After analyzing the test results the supervisor validates the tables. The validation is a procedure that transfers the changes from the temporary validated tables to the historicised tables.

Phase 3. Release: browse and download

After validation, the tables become available for downloading and browsing. Finally, external users can download the decision tables in pdf format from the "Download" section of the site.

The table update activities are coordinated by a supervisor who controls the transition to the next steps of the process. In particular, the supervisor defines the timing for the termination of the encoding and the starting time for the table updating procedures. Morover, the supervisor can unlock, if necessary, the activities of editors on already implemented updates and re-run updating procedures for a given year.

Figure 5 - DTE workflow overview (a)



(a) Access to the editing part of the system is limited to internal users. Spots in the "ACCESS" area represent user profiles allowed in the different phases; red spots indicate user profiles mainly involved in the specific phases. Manual and automated procedures are represented by solid and dotted arrows respectively. Check points (in red) are ordered by occurrence.

4.4 Table editing

Editors enter updates by translating text instructions into relationships between ICD codes. To do this, DTE provides an input panel where editors can document single updates by specifying a unique identifying name, textual recommendation, source and implementation date. The input panel is also equipped with an encoding panel where editors can specify the relationships between codes to be added, deleted or modified in the tables for the implementation year. Therefore, besides the transformation of text into relationships between codes, the editor must specify which actions should be performed for each specified row, namely addition, deletion or modifying.

In Figure 6 the general structure of the input panel is shown and a practical example of manual encoding is also provided. Referring to Figure 4, WHO recommends the implementation in 2013 of an update to the volume 2 of the ICD10. The instruction is:

"The following infectious diseases should not be accepted as due to any other disease or condition, except when reported as due to human immunodeficiency virus [HIV] disease, malignant impairing the immune system:

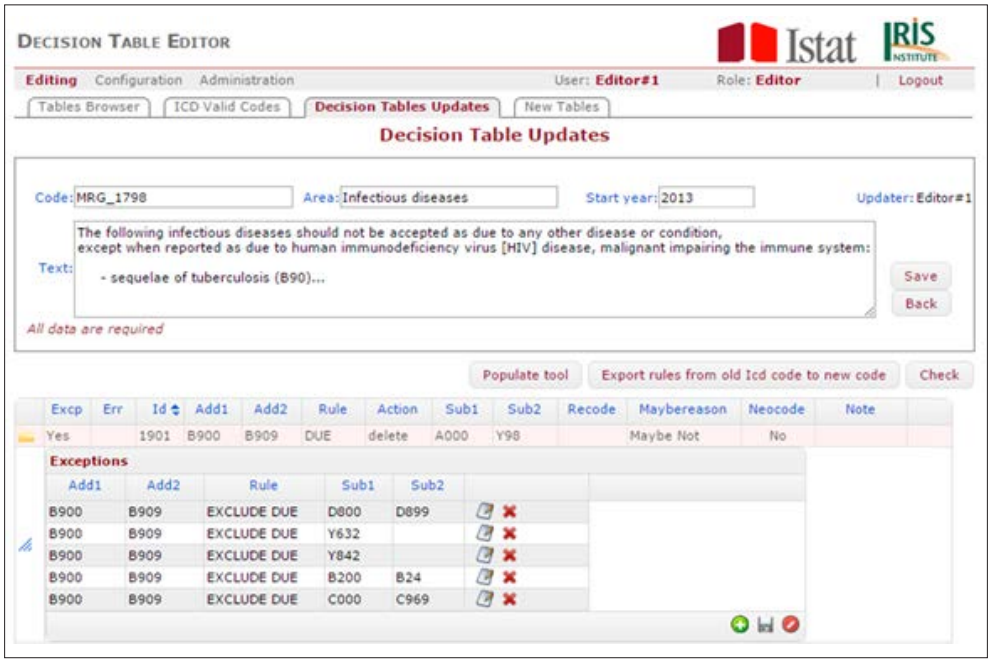
• sequelae of tuberculosis (B90)…"

This instruction includes a statement and the related exception. The editor has to manually encode both of them row by row.

The encoding of the statement *"sequelae of tuberculosis should not be accepted as due to any other disease or condition…"* implies a "delete" action as the address *"should not be accepted as due to"* the subaddress so the relationship must be deleted from the tables. In case of affirmative statement (e.g. *"can be due to"*), action field would be set to "add".

The exception to the previous statement *"…except when reported as due to human immunodeficiency virus [HIV] disease, malignant impairing the immune system"* is encoded as well and rule is automatically set according to the rule entered in the related encoding.

Figure 6 - Update input panel



4.5 Table browser and encoding features

Table browser

To allow retrieval of table data, the DTE system is equipped with a table browser utility. The table browser allows retrieving data from historicised tables which are stored in a normalised structure and returns data in a compressed form. The search can be performed with very flexible criteria such as: year of edition, codesets[6], type of relationship, maybes or recodes. The upper part of the table browser panel allows specifying all the criteria for the search.

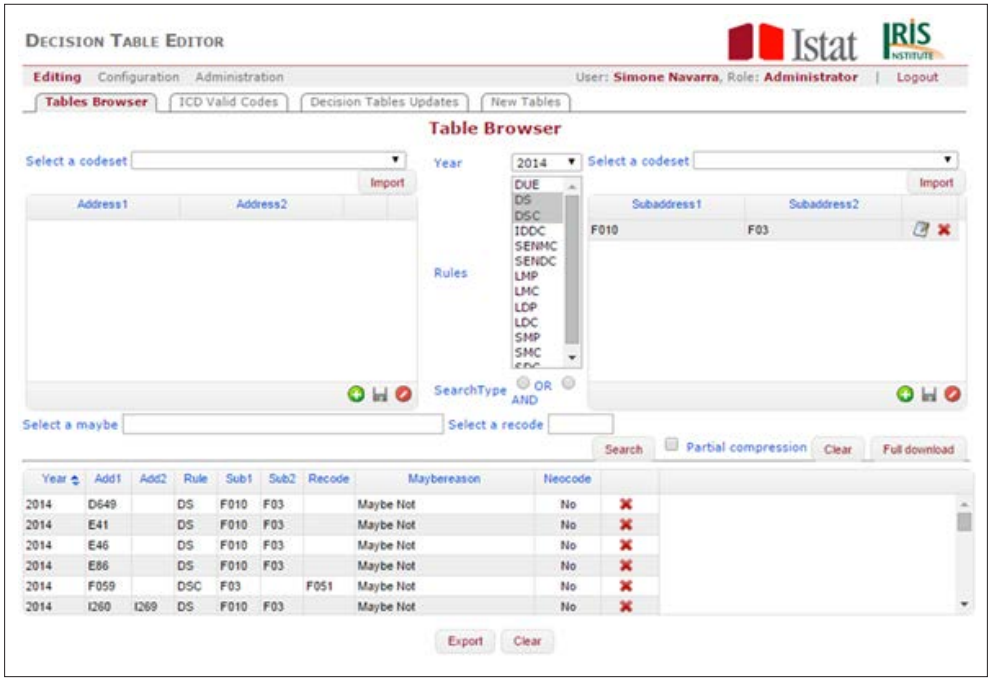
The search results are restituted in the bottom part of the panel in different formats also specified in the criteria panel:

- partial compression (only subaddress is compressed into intervals);
- double compression (both subaddress and address are compressed);
- exported in csv format.

A screenshot of table browser is presented in Figure 7.

6 A codeset is a collection of non-consecutive ICD10 codes which refer to a specific broad group. For example the codeset “dementia” groups codes from different ICD chapters such as F01-F09, G30.

Figure 7 - Table browser (a)



(a) The Figure shows a search on 2014 tables of all conditions that can be considered obvious consequence (rules DS and DSC) of dementia F01-F03. Only a part of the results retrieved are shown in the Figure.

Populate tool

Manual encoding is the simplest way to enter encoding rows. Nevertheless it does not take into account the information of the relationships contained in the actual tables (for instance, whether or not relationships specified in the updates already exist in the tables). In this sense it is a blind update. To avoid this problem a tool is designed for retrieving and modifying existing rows from the tables. As a support to manual encoding, the input panel provides the editors with a populate tool. This instrument is especially useful when large sets of relationships need to be handled at the same time avoiding time-consuming manual data entry.

For example, it may be necessary to modify (according to the WHO update) all the relationships involving a given code or pair of codes. The populate tool allows searching for all these relationships in the existing tables

(the last updated version) and uses them to populate the encoding panel where they can be manually edited. In other cases it may be required to link groups of address and subaddress codes through a given relationship in all possible combinations. The populate tool allows to calculate all these combinations sparing the user the effort to type them one by one in the encoding panel.

The populate tool shares most of the functions with table browser but differs from it in the following features:

- table reference year cannot be selected but it refers to the last available;
- search results are exported to the encoding panel.

An additional tool provided, called rule export, is used when it is necessary to create all the relationship for a code, for instance when an update creates a new code. In these cases, by means of the rule export tool, it is possible to attribute to the new code all the relationships of another code (both in address and subaddress). These are successively exported in the encoding panel for revision.

4.6 Quality control and validation

During the process, many check points have been designed in order to ensure the quality of the information produced. When checks are run, the errors discovered are distinguished into:

- *hard errors.* They must necessarily be corrected by the operator otherwise the process cannot go to the successive step;
- *soft errors.* They are displayed to the operator but they can be either corrected or accepted;
- *automatically corrected errors.* They are not displayed and they are automatically corrected because the correction is univocal.

Online check

The online check verifies the formal correctness of each row entered in the panel, independently from other rows. It is performed during data input by procedures embedded in the encoding panel. The following aspects of data consistency are checked:

- *code validity:* according to the year of implementation of the update;
- *consistency of the spans.* When a span is reported, the second code in the span must be a successor of the previous;
- *applicability.* Some modalities of the relationship variable (see Prospect 1) can be applied to a restricted set of address codes:
  - IDDC can be used only if the address contains exclusively ill-defined codes, whose list is specified in the valid codes table;
  - SENDC and SENMC can be used only if the address contains exclusively senility codes, whose list is specified in the valid codes table[7];
  - LMC, LMP, LDC, LDP cannot be used for ill-defined and senility codes;
- *recode.* The recode must be specified only for relationships requiring combination (DSC, IDDC, SENMC, SENDC, LMC, LDC, SMC).

When an error occurs during data input, prompt messages are triggered.

Encoding check "within" and "between"

The encoding check is a procedure for verifying the consistency of each encoding row inputted in the database with the others. Therefore it takes into account the overall data input, not the single row. The "within" check is run by the editors and examines the consistency of rows referring to a single encoding panel. The "between" check is run by the supervisor and verifies the consistency of the overall encoding for a given year. Inconsistencies checked in these steps are:

- *contradictory actions (hard error).* Two or more rows contain the same address, subaddress and relationship but action is opposite (i.e. add and delete the same relationship);
- *duplication.* If there is a duplicated combination of address, subaddress and relationship the presence of a maybe reason or not defines whether

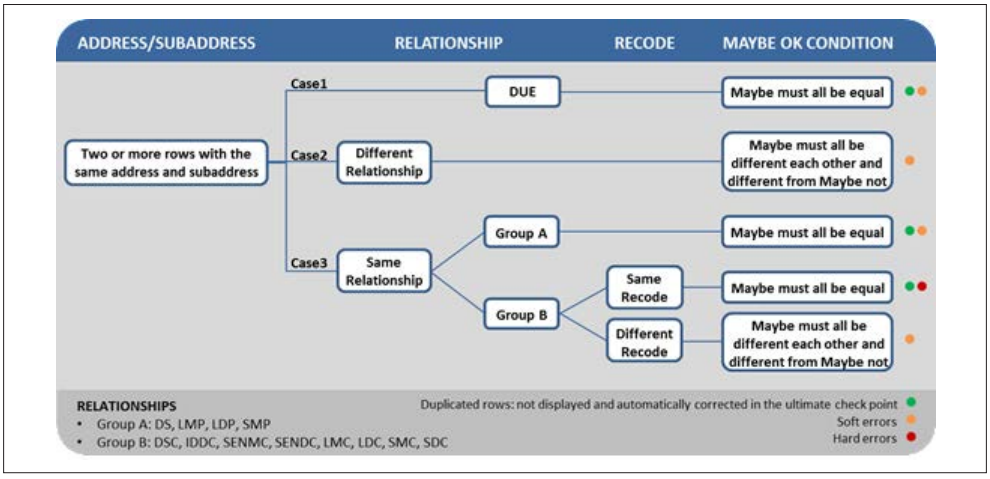
7 From the 2016 edition of the tables the rules SENMC and SENDC have been deleted. For SENDC the IDDC rule has been used, the new rule IDMC has been created to substitute SENMC and for other uses as well.

the error is a soft error, a hard error or an automatically corrected duplication (Figure 8). Rows containing the same address, subaddress, relationship, maybe and recode are considered duplications. They are not shown as errors because they will be automatically corrected in the ultimate check point;

- *maybe reason specification (soft error).* Two or more rows contain different modification relationships and no maybes are specified (more details in Figure 8).

If errors are present, details are displayed in a separate window. Further, inconsistent rows are highlighted in the encoding panel so that the editors can correct them. The encoding check (within) procedure can be repeated several times until all inconsistencies are removed. The encoding of checked updates is closed and the updates are directed to the supervisor (Phase 2) so that editors can no longer modify data unless the supervisor considers necessary to unlock and return them to Phase 1.

Figure 8 - Maybe reason specification check



Final check and automated correction

Changes introduced by the update procedure may be a source of new inconsistencies between rows in the updated tables. Therefore, a set of checks is performed for the following aspects:

- *applicability (hard error).* Described above;
- *symmetry.* This refers to the presence of two rows containing the same relationship where the address and subaddress are interchanged.

  Example:

row #	address	subaddress	relationship
1	A	B	DUE
2	B	A	DUE

  Relationships can be divided into symmetric (DUE and LMC; can display symmetry), and non-symmetric (all the others):

  - IDDC, SENMC, SENDC, LDP, LDC, SMP, SMC and SDC relationships must not display symmetry (hard error);
  - Presence of symmetry for DS, LMP, IDDC and SMC relationships (soft error);
  - LMC relationship must display symmetry (missing rows are automatically inserted);
- *modification relationship (soft error).* A pair of address and subaddress cannot have more than one type of modification relationship (all relationships except for DUE are modification relationships);
- *duplication.* Described above;
- *maybe reason specification (soft error).* Described above.

If errors are present, details are displayed in a separate window and errors are manually revised by the supervisor.

In the very last automated correction the following aspects are checked and errors automatically corrected:

- *duplication (hard error).* Simple combinations of address, subaddress and relationship as well as duplications deriving from previous checks are deleted (only one row is kept);
- *reflexive due to relationship (hard error).* Every code must be linked to itself by DUE relationship. Missing rows with DUE relationship are automatically added;

- *implicit due to relationship (hard error).* A pair of address and subaddress linked by DS or DSC relationship must be linked by DUE relationship as well. Missing rows with DUE relationship are automatically added.

5. Conclusions and future steps

The current version of DTE includes decision tables for the selection of the UC. However, an additional set of tables is designed for a preliminary step of the coding.

The UC selection is only a part of the overall coding process indeed, and in a previous step an ICD code is assigned to all the conditions reported on the death certificate. This task, referred as multiple cause coding, is critical for the successive step of the selection. During the multiple cause coding other information should be taken into account because conditions can get different ICD codes according to variables such as:

- age and gender of decedent;
- interval between onset of diseases and death, when reported;
- manner of death;
- presence and positioning of other conditions on the certificate;
- pregnancy status.

In analogy with the UC selection tables, a set of multiple cause coding tables exist and have been developed as documentation of Iris.

The future development of DTE envisages the inclusion of these tables in order to provide a tool for their systematic management.

This is a step toward the standardisation of multiple cause rules which will result also in better multiple cause data, that will be available for innovative research purposes.

Acknowledgments

DTE has been developed by the Italian National Institute of Statistics - Istat as one of the activities for the collaboration within the Iris Institute, established through an agreement with the German Institute of Medical Documentation and Information (DIMDI), and signed by the two institutions in 2012.

Besides the authors, the development of DTE has been possible thanks to the advice of the other members of the core group of the Iris Institute, namely Robert Anderson, Isabelle Bonellie, Olaf Eckert, Xavier Lavin, Gérard Pavillon and Stefanie Weber.

The software has been developed for Istat by the contractor Top Network S.p.A. In particular, the activities were performed as follows: Angela Ciocci, coordination of the development activities; Claudio Carotenuto, development of the interface; Sara Fiorini, development of database and procedures.

References

Centre for Disease Control and Prevention – CDC, National Center for Health Statistics – NCHS. *About the Mortality Medical Data System*. http://www.cdc.gov/nchs/nvss/mmds/about_mmds.htm.

Centre for Disease Control and Prevention – CDC, National Centre for Health Statistics – NCHS. National Vital Statistics System. 2016. I*nstruction manuals. ICD-10 ACME Decision Tables for Classifying Underlying Causes of Death*.

Eckert, O. 2014. "Improvement of Mortality Statistics by MUSE". Presentation at the *European Conference on Quality in Official Statistics – Q2014*. Vienna, 2-5 June 2014.

Harteloh, P., K. de Bruin, and J. Kardaun. 2010. "The reliability of cause-of-death coding in The Netherlands". *European Journal of Epidemiology*, Volume 25: 531-538.

Iris Institute. *Official website*. www.iris-institute.org.

Pavillon, G., M. Coleman, L.A. Johansson, E. Jougla, and J. Kardaun J. 1998. "Final report on automated coding in member states". Eurostat Working Papers, Population and social conditions. Luxembourg: Office for Official Publications of the European Communities.

Pavillon, G., L.A. Johansson, D. Glenn, S. Weber, B. Witting, and S. Notzon. 2007. "Iris: A Language Independent Coding System For Mortality Data". In WHO – Family of International Classifications Network – FIC. *Annual Meeting.* Trieste, Italy, 28 October – 3 November 2007.

Pavillon, G., and L.A. Johansson. 2012. "The Iris International Coding System Of Causes Of Death". In WHO – Family of International Classifications Network – FIC. *Annual Meeting*. Brasilia, Brazil, 13-19 October 2012.

World Health Organization - WHO. *ICD10 Interactive Self-Learning Tool*. http://apps.who.int/classifications/apps/icd/icd10training/

World Health Organization - WHO. *International Statistical Classification of Diseases and Related Health Problems. 10th Revision, 2010 edition*. Geneva, Switzerland: WHO.

World Health Organization - WHO. *List of Official ICD-10 Updates*. https://www.who.int/standards/classifications/classification-of-diseases/list-of-official-icd-10-updates.