



**MEETING OF HEADS OF WHO COLLABORATING CENTRES
FOR THE CLASSIFICATION OF DISEASES**

Brisbane, Queensland, Australia
14-19th October 2002

Title: Modelling ICD-10-AM

Authors: Kerry Innes¹, Patricia Saad², Rosemary Roberts³

¹ Associate Director

² Project Officer

³ Director

National Centre for Classification in Health, Sydney, Australia

Abstract: (no more than 200 words)

The emergence of electronic health records (EHRs) has focussed attention on the sensitivity and specificity of coded health data routinely collected. The debate continues about the need for precise and meaningful clinical information to support EHRs and communication between clinicians. It is characterised by arguments for the use of comprehensive concept-based terminologies, and against the suitability of classifications for these purposes. It is clear that classifications aggregate information for statistical and reporting purposes, and that this necessarily results in some loss of exactitude in clinical meaning. However, we argue that neither concept-based terminologies nor existing classifications in their present form can support the myriad demands for future use of coded health data.

Much emphasis is given to the differences between terminologies and classifications. Indeed, in terms of their distinct purposes, construction and ultimate uses, terminologies provide familiar, precise concepts and terms in a poly-hierarchical structure without the use of residual classes (other and unspecified). On the other hand, classifications, formulated to aggregate data, are characterised by mono-hierarchies and the use of residual classes. The result is that terminologies tend to be more specific, but less sensitive; classifications, on the other hand, are more sensitive but less specific.

However, this perspective disguises their shared similarities. Both terminologies and classifications deal with the same medical knowledge, they merely represent it differently. The concepts provided by terminologies are in fact present in classifications, though these are often not explicit or accessible. In the case of ICD-10-AM, extensive concepts and terms relevant for use in EHR and clinician communication are embedded in the Index (not provided directly in the Tabular classification).

In essence, terminologies and classifications are capable of expressing the same clinical knowledge. In order to maximise the benefits offered by terminologies (in their descriptiveness and exactitude) and classifications (with their ability to aggregate and report large volumes of data), it is necessary to integrate the two; modelling ICD-10-AM represents an early investigation of a methodology to achieve this end.

This document is not issued to the general public, and all rights are reserved by the World Health Organization (WHO). The document may not be reviewed, abstracted, quoted, reproduced or translated, in part or in whole, without the prior written permission of WHO. No part of this document may be stored in a retrieval system or transmitted in any form or by any means - electronic, mechanical or other - without the prior written permission of WHO.

The views expressed in documents by named authors are solely the responsibility of those authors.

Introduction

ICD-10 has a long history originating in the Bertillon Classification in 1893. Its evolution from a list of causes of death, to a classification incorporating morbid diseases (1948 – Sixth revision) to ICD-10 has dictated the format of the classification we see today and on which the disease section of ICD-10-AM is based. As well as classifying all known diseases, conditions and injuries, ICD-10 now has detailed sections relating to Symptoms, signs and abnormal clinical and laboratory findings, External causes of morbidity and mortality and Factors influencing health status and contact with health services. The disease chapters are organised in most cases according to body system or site. Aetiology is paramount for infections, neoplasms and congenital malformations. Status and time of life underpin chapters on obstetric and perinatal conditions. A separate section on morphology of neoplasms is extracted from ICD-O to use with ICD-10.

The emphasis in ICD is on classification or data output and reporting of diseases for a myriad of reasons. Its first volume is the ‘Tabular List’ which is the list of classes, including specific and non specific classes, organised according to the axes described above (site, cause, life stage or status). Access to the classification is via the alphabetic index which is where the specificity of terms and disease names is found.

Not only is the classification in the first volume, it was the first part of ICD-10 to be published by WHO (1992) followed by Volume 2 (Instruction manual) in 1993 and the index in 1994. Being a relative newcomer historically (1948), the index has never received the attention it deserves in its role of providing the link between terms used by clinicians in the patient record and classes in the tabular list. When the Unified Medical Language System incorporated ICD products into the metathesaurus, only the tabular list was included. This has meant that ICD-10 has an unwarranted reputation for unseemly bundling of ‘NOS’ and ‘NEC’ terms and rubrics such as *Other specified diseases of liver*. But all the specific terms are in the index, which because of its structure, complexity, and difficulty of manipulation in hard copy or word-processed format, has been largely ignored. The index of terms is the input device, the tabular list the output. The rules that bind the relationship between the index and tabular are the glue that holds the classification together and determines a ‘correct’ pathway between specific terms and bundled classes represented by code.

ICD-10 was developed before electronic tools were commonplace. It supports a manual coding function, hard copy records, and a person making a translation between a term and a class. That person in Australia is a clinical coder, who also makes decisions on what is relevant to code from an episode of care. For use as part of ICD-10-AM, NCCCH has built on the conventions of ICD and created the Australian Coding Standards, a guide to the application of codes to make the process as objective and scientific as possible.

Also pertinent to this discussion is the maintenance of the classification to retain its currency and clinical credibility, and to keep parity with related classifications such as ICD-O. WHO has an ‘official’ updating process (Update Reference Committee) which publishes major changes every three years but even this is cumbersome and difficult to control on an international basis. It must, of course, be balanced against the stability of the classification and the ability to compare data over time. The ICD-10/ICD-O nexus is a good example of how two classifications within the WHO

'Family' can get out of step unless there is a managed process of update. ICD-10 as used in most countries contains morphology codes from ICD-O Second Edition (1990) while ICD-O Third Edition was published in 2000 and has been used for cancer registration in Australia since 2001. The ICD-O Third Edition codes are included in the Third Edition of ICD-10-AM but not in WHO ICD-10.

Modelling rationale

A conceptual model is now required to express the tangled web of connections between terms, rules and classes. Only through the development of such a model will those building and maintaining classifications really come to grips with all its interrelationships. Such an understanding is absolutely necessary to present the classification to its users within health services (input) and those using the coded data that results.

The involvement of NCCH in the WHO Update Reference Committee has brought home the need for international collaboration on the update process. This is becoming more complex as countries use the ICD codes as the basis for casemix grouping, and rules for allocation of codes need to be strengthened to make the coding process as scientific as possible. Clinical modifications of ICD-10 are being developed by many countries to replace ICD-9-CM. Apart from Australia's ICD-10-AM, first published in 1998 and now in its Third Edition, other modifications have been created as shown in Table 1.

Table 1. Clinical modifications of ICD-10 for morbidity coding

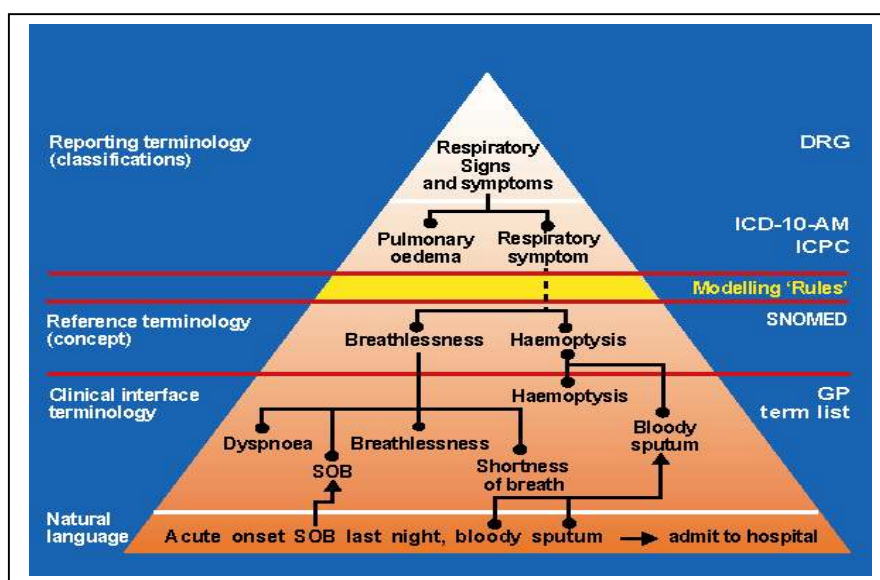
Country	Classification	First published	Most recent edition
Australia	ICD-10-AM	1998	Third, 2002
Canada	ICD-10-CA	2001	First, 2001
Germany	ICD-10 SGB-V	2000	Second, 2001
Thailand	ICD-10-TM	Under development	
USA	ICD-10-CM	Under development	

Managing these modifications so that comparability between concepts is maintained is becoming more and more difficult. Some years ago, NCCH suggested to WHO that a database of ICD-10 be developed so that countries creating clinical modifications could provide clinical extensions and share the rationale and results of these deliberations. Germany, the Nordic Countries and Australia now hold their modifications of ICD-10 as databases but there is still no formal international exchange through WHO or any other forum of an ICD-10 database or of the structure or content of these databases to inform national needs. In its license provisions for countries to use ICD-10, WHO makes certain conditions which attempt to limit the national clinical modifications to fifth digit specificity and to prevent change in concept and meaning of third and fourth level codes.

Another important element relates to procedure classification development. All the countries listed in Table 1 have created procedure classifications to accompany their country's modification of ICD-10. Many countries use WHO ICD-10 diseases with ICD-9-CM procedures, some with ICPM translated or updated from its 1978 version.

This emphasis on the ICD and procedure classifications is vital in our understanding of casemix classifications. At present, we are dependent on mapping between classifications to express morbidity data from any source in a range of casemix classifications. The mapping, of course, is always imperfect. If ICD classifications were modelled, the relationships between terms and classes would be more explicit, and inform processes such as mapping, new classification development and creation of specialised subsets of classifications, not to mention extraction of clinical terms from electronic health records for multiple uses. Modelling will support this translation of terms used by health professionals in describing patient condition and intervention into some sort of communication and reporting mechanism that is commonly understood. Figure 1 shows a diagrammatic relationship between clinical terms and casemix classes and helps to define what is meant by a terminology compared with a classification.

Figure 1. Relationship between terminologies and classifications



Terminologies

Australia is still considering its position on the use of reference terminologies such as SNOMED CT. NCCH has always considered that such a terminology would supplement rather than replace a classification, and that the connection between a terminology and a classification would have to be explicit. NCCH was part of the alpha trials of SNOMED CT, and in this capacity undertook a mapping between SNOMED CT alpha trial domains and ICD-10-AM. Similar mappings have been undertaken by UK NHS Information Authority between CTV3 and ICD-10 and by the College of American Pathologists between SNOMED CT and ICD-9-CM and SNOMED CT and ICD-10.

While the decision on terminologies is being considered by the Australian government, NCCH thought it would be useful to develop a model of terms within ICD-10-AM. This model could inform either a mapping between the Australian classification and an imported terminology (such as SNOMED CT) or the development of an Australian Clinical Vocabulary.

The NCCH modelling pilot study

Using Cyber+LE software (because it was available to us), we attempted to model two chapters of the ICD-10-AM classification, re-engineering it into a poly-hierarchical concept-based terminology. We chose two chapters, Diseases of the Eye (Volume 1, Chapter 7) and Procedures on Breast (Volume 3, Chapter 17), as test cases for this process.

Theoretically, a comprehensive modelling exercise would involve four processes:

1. Normalisation of index terms
2. Dissection of index terms
3. Construction of computer readable classification rules and conventions
4. Construction of clinically accurate hierarchical relationships

1. Normalisation of index terms

This process ‘unpacks’ the index, giving expression to modifier terms and non-essential modifier terms held within the disease and procedure index. Terms are then linguistically normalised so that noun, verb, adverb, and adjective phrasing, tense, capitalisation, spacing, and hyphenation are standardised.

2. Semantic parsing

Dissecting the index terms/phrases into the individual semantic components is known as semantic parsing. This process allows identification of domains such as procedure method, body site, infectious organism etc. providing a different hierarchical view of the concept. That is, terms such as conjunctivitis can be attributed to the anatomical domain of eye, as well as the disease domain. It serves to overcome problems raised by different methods of pre- and post-co-ordination of term phrases; for example the laparoscopic cholecystectomy can be parsed into meaningful components of Approach=laparoscopic and Procedure = cholecystectomy.

3. Creation of relationships to reflect clinical reality (rather than relationships that currently serve statistical aggregate)

The previous two steps disassembled the classification from its current form, and readied the content for this next step. Here the unpacked classification content is re-assembled in the form of a concept model, with relationships defined to reflect clinical reality. Creating these concept relationships provides pathways to terms and concepts by defining parent, child and sibling relationships.

It also provides greater access to terms by re-combining meaningful term constructs not visible through classification conventions, for example diabetic neuropathy. This process also allows synonymy to be defined.

4. Computer readable classification rules and conventions

Presently, classification rules and conventions are embedded within the marginalia, formats and guidelines in printed forms of the classification. No mechanisms exist that will allow computers to interpret the classification contents reliably. For example, inclusion and exclusion notes, ‘code also’, ‘see also’, and ‘see’ directives within the classification are represented textually, not in relational forms. This means that computers are unable to understand their intended meaning and to apply this interpretation to term, concept or code usage. Few existing terminology or

classification development environments provide functionality that automates this process. Building computer readable classification rules and conventions is likely to require a ‘work around’, a creative solution, or manual processing. The feasibility study using Cyber+LE demonstrated that a solution is possible, however, better approaches may be available.

For the purposes of our feasibility study, using Cyber+LE, we did not undertake semantic parsing to the extent necessary for a comprehensive model. Nonetheless, the pilot was valuable in informing future options for the building of domains within the model.

Figure 2 shows a small section of the ICD-10-AM Tabular relevant to code H46 *Optic neuritis*. Figure 2a shows how this concept is represented in the Index. Figure 3 shows the concept model of these ICD-10-AM entities, after modelling procedures of normalisation, relationship definition and convention specification have been deployed through Cyber+LE.

Note that relationships are specified in the model example: ‘optic neuritis’ is a child of ‘disorders of the optic nerve’ and also a child of ‘inflammation of eye’. Synonyms are fully expressed. Example: ‘optic nerve neuropathy’. Note also that computable rules are represented (for example, the ‘excludes’ note at the code H46 for ‘neuromyelitis optica’ is reflected as a lateral relation).

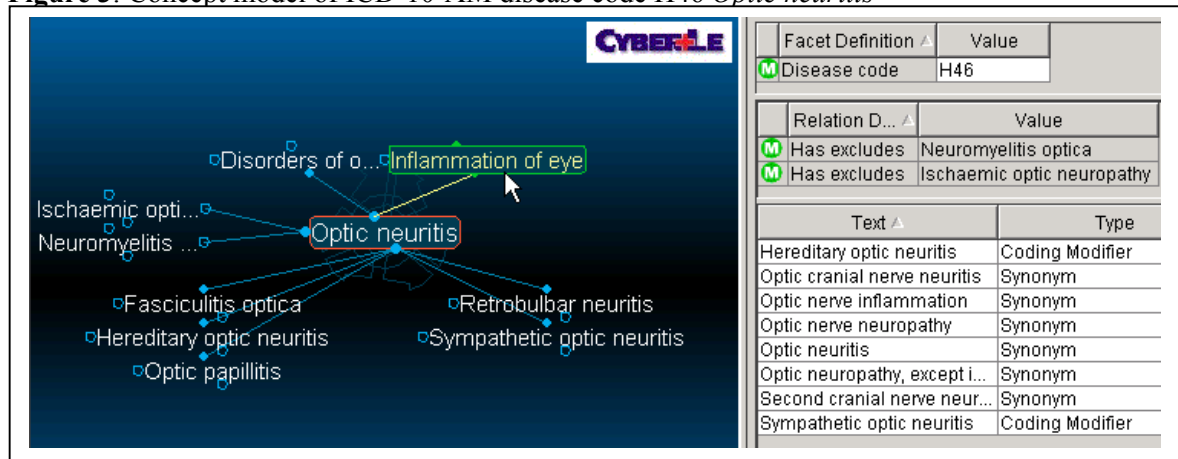
DISORDERS OF OPTIC NERVE AND VISUAL PATHWAYS (H46–H48)	
H46	Optic neuritis
	Optic:
	• neuropathy, except ischaemic
	• papillitis
	Retrobulbar neuritis NOS
	<i>Excludes:</i> ischaemic optic neuropathy (H47.0)
	neuromyelitis optica [Devic] (G36.0)

Figure 2: Extract from ICD-10-AM tabular list

Fasciculitis optica H46
Inflammation, inflamed, inflammatory (with exudation)
- optic nerve H46
Neuritis
- cranial nerve
- - second or optic H46
- optic (hereditary) (sympathetic) H46
- retrobulbar H46
Neuropathy, neuropathic G62.9
- optic (nerve) H46
Papillitis H46
- optic H46

Figure 2a: ICD-10-AM index entries for code H46

Figure 3: Concept model of ICD-10-AM disease code H46 *Optic neuritis*



The concept model brings together the terminology in the Index and the Tabular and is re-engineered to reflect a clinically relevant hierarchy. Terms (synonyms, abbreviations, grammatical variants etc) can be added to build upon the depth of the terminology and child/sibling relations can be modified to ensure clinical appropriateness.

Working with procedures on the breast, the same methodology was used to model surgical concepts. Figures 4 and 4a represent the existing delineation of concepts as they are represented with the tabular and index of the ICD-10-AM classification (respectively). Figure 5 shows the resultant concept model displayed by Cyber+LE.

1744	Local excision of breast
	Lumpectomy Tylectomy
	<i>Code also when performed:</i>
	• excision of axillary lymph nodes (30332-00, 30332-01, 30333-00 [808])
	• localisation of lesion of breast (30361-00 [1740])
30342-00	Local excision of lesion of breast Excisional biopsy of breast

Figure 4: Extract from ICD-10-AM tabular list

Biopsy
- breast (aspiration) (fine needle)
- - excisional 30342-00 [1744]
- - local 30342-00 [1744]
Excision — see also <i>Removal</i>
- breast — see also <i>Mastectomy</i>
- - local 30342-00 [1744]
- cyst — see also <i>Excision, lesion</i>
- - breast 30342-00 [1744]
- fibroadenoma
- - breast 30342-00 [1744]
- lesion — see also <i>Excision, tumour and Excision, cyst and Excision, polyp and Excision, by histology</i>
- - breast 30342-00 [1744]
- tumour
- - breast 30342-00 [1744]
- - fibroadenoma
- - - breast 30342-00 [1744]
Lumpectomy (breast) 30342-00 [1744]
Tylectomy (breast) 30342-00 [1744]

Figure 4a: ICD-10-AM index entries for 30342-00

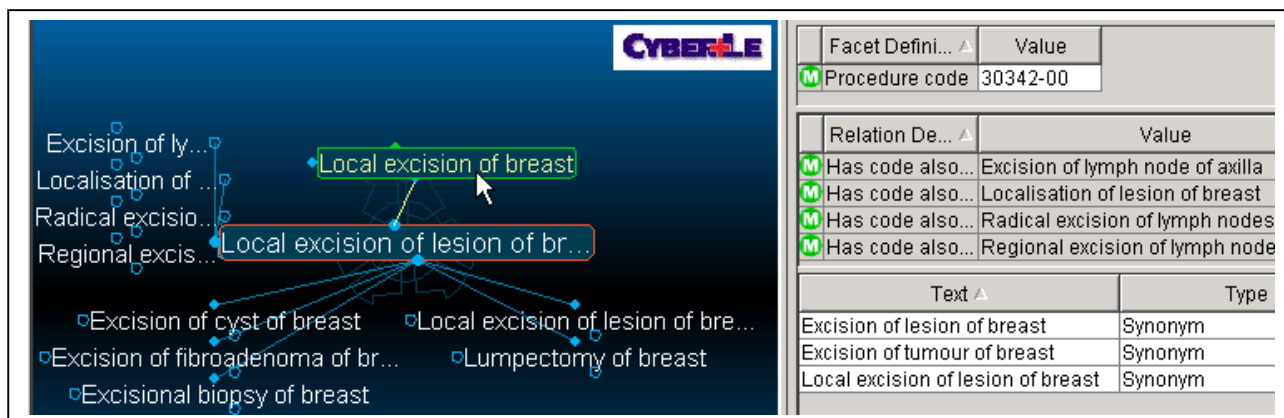


Figure 5: Concept model of ICD-10-AM procedure code 30342-00 *Local excision of lesion of breast*

Discussion

The modelling project has led us to a deeper understanding of the ICD-10-AM classification and exposed the evolutionary process it has followed in its hard copy format to deal with multiple hierarchies and guiding coders to follow reliable pathways through use of rules and conventions. We wish to continue the process for the whole ICD-10-AM classification. As well, it would be extremely useful to reach international agreement on both a model and a database for ICD-10 itself, so that

international and specialty variants can be developed from this base. It would also help inform translations of ICD-10.

One of the most beneficial outcomes of modelling will be the provision of classification rules and coding conventions in a computable readable format. This has never been done before, and is necessary to fully understand the statistical aggregation of clinical concepts to classification rubrics and to demonstrate logical and consistent pathways from terms to codes. That is, to provide a knowledge base of relevant and reliable classification rules and coding conventions.

Conclusion

Over a century of work has resulted in the WHO ICD-10 classification of diseases. ICD-10 is no longer a list but a complex web of classes, clinical terms and instructions which guide coders in relating patient characteristics to the classification rubrics. This evolution provides a solid foundation for application of electronic modelling tools to represent the classification, its structure and content, so it can be understood, held and presented to users in a format that is both efficient and credible in capture, communication and use of data. These developments are essential to ensure the WHO ICD is relevant in this electronic age where the use of classifications of different aggregations governs so much that happens in health.