WHO-FIC NETWORK MEETING

Tokyo, Japan
16-22 October 2005

IRIS: A language-independent coding system based onthe NCHS system MMDS

Lars Age JOHANSSON (National Board of Health and Welfare – Sweden)
Gérard PAVILLON (National Institute for Health and Medical Research Inserm-CépiDc- France)

Abstract

Mortality data are widely used to describe and compare health characteristics. Mortality statistics are mainly based on the underlying cause of death. The selection of the underlying cause is directed by ICD rules and guidelines, and requires both extensive training and much experience. The US National Centre for Health Statistics introduced automated selection of the underlying cause in the late 1960ies. This system, the Mortality Medical Data System (MMDS), is now used by many other countries as well [ICE96][ICE99].

The expressions in the MMDS dictionary are linked to Entity Reference Numbers (ERN) and not to ICD codes. However, these ERNs cannot always be mapped to non-English expressions, which means that the MMDS term coding module cannot easily be adapted for other languages. This paper describes the development of IRIS, an add-on to MMDS, that allows non-English-spaeking countries to make use of the MMDS term coding module.

Abstract 1
Background 3
Objective 4
Method 4
Results 4
Perspectives (2005) 5
References 5

Background

Mortality data are widely used to describe and compare health characteristics between countries. The International Classification of diseases (ICD), published and updated by the World Health Organisation (WHO), is used by most countries to code causes of death [OMS92]. ICD is now more than one century old and the last revision of this classification (ICD10), reflecting the increasing complexity of medicine, is so complex that only classification experts can use it properly.

Mortality statistics are mainly based on the concept of underlying cause of death, defined by WHO as the "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". The selection of the underlying cause of death is based on the diagnoses reported by the certifier, usually a physician. The selection is directed by ICD rules and guidelines, and made by professional medical coders (nosologists). This task requires both extensive training and much experience.

In this context, quality and comparability of data are essential [Eurostat01b]. Comparisons between medical coders show important variations due to interpretation differences or errors. Software for coding and selection have now been available for some time. The US National Centre for Health Statistics introduced automated selection of the underlying cause in the late 1960ies. Further modules and functions have been added, and now the NCHS system, called the Mortality Medical Data System (MMDS), is used by many other countries [ICE96][ICE99].

The MMDS is divided into three functional parts: coding of medical causes of death reported by the physician (MICAR), selection of the underlying cause of death (ACME) and compilation of multiple causes of death (TRANSAX). ACME is now considered the de facto standard for countries using automated coding systems. Briefly, ACME selects and modifies the ICD code of the underlying cause of death from the set of ICD codes representing the conditions mentioned on the death certificate. This function is independent of the language. MICAR, however, assigns the ICD codes to the medical terms reported by the physician. This is highly language dependent. Even if MICAR performs perfectly in English, many problems appear when it is applied to other languages. Consequently, while the MMDS has been fully applied by English-speaking countries (US, UK, Australia), only one part of it, namely ACME, is used by non-English-speaking countries (Sweden, Brazil, France....). Using ACME without MICAR has its problems, however, since ACME does not work properly unless the coding of the medical expressions follows a set of very specific instructions. MICAR applies these instructions automatically, but countries that do not use MICAR have to apply them manually, or include the functions of MICAR in their own coding software. It is generally agreed that a common automated coding system, with modules for both text coding and selection of the underlying cause, would be the best way to improve mortality data quality and comparability. However, the problem of language is still a major obstacle.

Objective

The problem of MICAR with non-English languages is closely linked to the English dictionary of diagnostic expressions (medical terms). Each diagnostic expression is linked to an Entity Reference Number (ERN). The expressions and the ERN, however, cannot always be mapped to similar non-English expressions. There might also be expressions in the non-English language that do not correspond to any expression in the English dictionary. This all corresponds to the well-known difficulties of translation from one language to another.

In order to cope with this problem, we decided to study the possibility to implement the coding function (MICAR) in a language-independent way. The first objective is to produce a program that can be used with different languages by just changing the dictionary of causes of death. The second aim is to stay fully compatible with MICAR.

Method

At first, we planned to rewrite MICAR, and when doing so separate the language and the algorithmic aspects [Eurostat01a]. It would then be easier to apply the program to different languages. This task proved to be very complex and therefore the compatibility between MICAR and the language independent program could not be guaranteed. Also, it would be difficult to synchronize the future development of two different programs.

The present project is based on a different idea. We assume that most ICD codes can be mapped to a single ERN. ICD codes are international and they are defined in different languages. Therefore it seems possible to enter MICAR via ICD codes. With this approach the language aspects are present only in the dictionary, which is, of course, different for each language. This dictionary maps diagnostic expressions to ICD codes, and the system then translates ICD codes to ERN. MICAR, embedded as a component in the system itself, is called to apply the instructions for ACME input coding, and to finally produce a set of ICD codes for ACME. Finally, the system calls ACME which selects the underlying cause.

This system would achieve the two objectives of language independency and compatibility. It has been tested on a set of death certificates already coded with ICD10.

Results

A software named Iris has been developed by France and Sweden. The translation tables from ICD to ERN have been finalized in collaboration with the NCHS team in charge of the MMDS software.

In order to test Iris, we used the Swedish text coding software, Mikado, as a reference. We retrieved 42 748 records with both the default ICD codes (ICD codes for each medical expression on the certificate, disregarding any other information on the death certificate), and the final underlying cause. We translated the ICD codes

into ERNs and then submitted the records to MICAR for further processing. MICAR rejected about 3018 records, in most cases because MICAR had encountered a mention of surgery (at present, MICAR cannot code surgery and such records are rejected for manual review). Of the remaining 39 530 records, the underlying causes produced by IRIS differed from those obtained by Mikado in only 3018 cases (5.9%). We are currently analysing the differences.

Perspectives (2005)

The results show that the Iris system works. Some problems remain but they can be solved in the next few months. Further testing will be done on files provided by Spain and Hungary. The language module specified in [Eurostat01b] will also be developed by the end of 2005, possibly in collaboration with Germany.

References

[OMS92] International Classification of Diseases and Related Health Problems. World Health Organization. Geneva 1992

[ICE96] International Collaborative Effort on Automating Mortality Statistics, Vol I. Washington, November 1996

[ICE99] International Collaborative Effort on Automating Mortality Statistics, Vol II. Washington, November 1996

[Eurostat01a] Production of Method and TOOLS for Improving Causes of Death Statistics at Codification Level. Final report. Eurostat 3/2001/E/N°14

[Eurostat01b] Comparability and Quality Improvement of European Causes of Death Statistics. Final report. Eurostat EDC DGV/F3 SOC 98 20108