Medical Informatics Europe '97
C. Pappas et al. (Eds.)
IOS Press, 1997

Semi-Automatic Coding with ICPC: The Thesaurus, the Algorithm and the Dutch Subtitles

Rudolph S. Gebel, general practitioner
Staff member Dutch College of General Practitioners
Lomanlaan 103, P.O. Box 3231, 3502 GE Utrecht, The Netherlands

Abstract
In the ICPC Thesaurus Project, which ran from 1990 to 1992, the Dutch translation of the English version of the ICPC-components 1 and 7 was made available for automated coding by structuring and improving the thesaurus and by developing an algorithm for selecting possible ICPC-codes from a set of medical terms given as input to the program. The thesaurus and algorithm are available to the developers of GP information systems and are at present incorporated in all Dutch GP-systems. This paper brings you up to date with the semi-automatic coding system and the so called Dutch subtitles, an extension to the ICPC.

1. Introduction

In the Netherlands, the International Classification of Primary Care (ICPC) [1] is accepted as the standard for coding and classification in general practice.
The ICPC classification system consists of two axes. One axis consists of the different body systems, the chapters, including a general chapter and a chapter on social problems. Each chapter is represented by a letter (for example F for the eye). The other axis consists of seven different components of ICPC. In this paper we only deal with the first (complaints and symptoms) and seventh (diagnosis) component. Each component contains several titles, represented by two digits (for example, the diagnostic component runs from 70 through 99). Thus, each so called Short Title is represented by a letter and two numbers, for example D88 APPENDICITIS.
In 1988, the Dutch translation of the Short Titles and of the so called inclusion terms (a rudimentary thesaurus) was ready for use. [2]

2. The ICPC Thesaurusproject

At the beginning of the ICPC Thesaurusproject in 1989, the ICPC-thesaurus was barely structured, mainly due to the fact that the thesaurus consisted of nothing more than a translation of the inclusion terms. An example of the inclusion terms of the Short Title K80 ECTOPIC BEATS ALL TYPES is:

ICPC-code	inclusion term no	inclusion term
K80.	1	ECTOPIC RHYTHM
K80.	2	ECTOPIC HEART BEAT
K80.	4	PREMATURE BEAT
K80.	5	ECTOPIC HEART RHYTHM
K80.	8	PREMATURE HEART BEATS
K80.	9	NON-SPECIFIC PREMATURE BEATS
K80.	10	SUPRAVENTRICULAR PREMATURE BEATS
K80.	11	VENTRICULAR PREMATURE BEATS

Due to the structure of the inclusion terms, many general practitioners believed them to be a subclassification, which led to much confusion. As you can see from the example above, the different inclusion terms do not exclude each other, which is the basis of a good classification. So, the first task of the project was to structure the thesaurus in such a way, that it could be easily automated and maintained.
This was done by grouping terms into so called word clusters. Word clusters are groups of terms that all mean the same: synonyms, spelling variants, lay terminology etcetera. Each cluster is linked to one or more Short Titles. With each Short Title, the importance of each word cluster for that specific Short Title is expressed as either 'Essential' or 'Non-essential'. Only essential clusters are used to select candidate codes, non-essential clusters are only used to influence the order in which the candidate codes are displayed.
An example of a word cluster, translated into English, is VARICELLA, which contains two words: VARICELLA and CHICKENPOX. Another example is the cluster APPENDIX, which contains the words APPENDIX and APPENDICULAR. After a first testphase, the clusters were revised, using the comments we received from the GP's that participated in the testing, and adding terms from the ICD-10, using the existing conversion table between ICD-10 and ICPC. [3]
Secondly, we had to handle the fact that the ICPC was developed as an international classification for epidemiological research, and not as a coding tool for Dutch general practice. As a result, some diseases which are extremely rare in Dutch general practice, have a separate Short Title (like malaria, A73), whereas more common diseases, like bursitis, have their place in the so called ragbags, i.e. L99. This formed a barrier for many GP's to use the ICPC. To meet the demands, we developed the so called Dutch subtitles. [4]
Subtitles form an extension to a Short Title and contain frequently assessed diagnoses in Dutch general practice. For example, the Short Title 'F99 Other diseases eye/adnexa' is divided into the following subtitles:

F99.01	Ectropion/entropion/blepharochalasis eyelid
F99.02	Dry eyes
F99.03	Pterygium
F99.04	Scleritis/episcleritis
F99.05	Occlusion retinal artery or vein
F99.06	Diplopia
F99.07	Anopsia all forms
F99.08	Color blindness
F99.99	Other diseases eye/adnexa

In developing the subtitles, we used data about prevalence and incidence of the different Short Titles derived from the Transition Project [5], and the conversion table between ICPC and ICD-10, as mentioned before.
Through the subtitles, more specific coding is possible which makes it more satisfactory to the Dutch GP, because he is now able to use the ICPC-codes as a tool for preventive activities, for prescribing medication etcetera. Epidemiological research at the level of the Short Titles is still possible by discarding the last two digits. In 1997, a new edition of the subtitles is presented.
Over 300 Short Titles from a total of 691 have two or more subtitles:

Tabel 1: Number of subtitles per Short Title

no. of subtitles per Short Title	no. of Short Titles	percentage of Short Titles
0	380	55,0
2	118	17,0
3	99	14,3
4	41	5,9
5	27	3,9
6	10	1,5
7	5	0,7
8	2	0,3
9	7	1,0
10	1	0,2
11	1	0,2

3. The semi-automatic coding program

The semi-automatic coding program works as follows:
During the consultation, the GP enters the symptom, complaint or diagnosis he wishes to encode in the journal of the GP information system he uses. The word(s) used will point to the corresponding word clusters. As explained earlier, these clusters are coupled with one or more Short Titles. The Short Titles are represented in order of likelihood. The GP chooses a Short Title. He can then choose one of the subtitles that belong to the Short Title (when available). The symptom, complaint or diagnosis is then encoded with the correct ICPC-code and stored into the database.
For example, the doctor types:

HEADACHE

The system presents the doctor with the following options:

N01	HEADACHE [EX. N02,N89,R09]
N02	TENSION HEADACHE
N89	MIGRAINE
N90	CLUSTER HEADACHE
R09	SYMPTOMS/COMPLAINTS SINUS [INCL.PAIN]

At first glance, the presentation of the last possibility, R09, seems illogical. We have chosen to display the exclusion codes as well, as presented at code N01 in this case. That way, the doctor can make a responsible choice.
After choosing N89 MIGRAINE, the following subtitles are presented:

N89.01	CLASSIC MIGRAINE
N89.02	MIGRAINE WITH NEUROLOGICAL SYMPTOMS
N89.99	OTHER/NON SPECIFIC MIGRAINE

As you can see, in this case the last code is a so called ragbag. We have given the ragbag-subtitles the extension 99, so that it is possible to add other subtitles without having to change the numbers.
The doctor decides that the right code is N89.01. Now he can make a choice. By toggling with the TAB-switch, he can either store his own text (N89.01 HEADACHE) or he can store the text of the subtitle (N89.01 CLASSIC MIGRAINE).
In retrospect, it would have been better if he had typed 'MIGRAINE' instead of the less specific term 'HEADACHE'. First of all, the computer can't interpret the fact that the doctor means 'migraine' if he types 'headache'. Secondly, the outcome of the search with 'MIGRAINE' would have been limited to one Short Title, i.e. N89.

4. Further developments

Since 1995, a helpdesk is available at the Dutch College of General Practitioners to help solve coding problems encountered by the GP. The comments we get are used to improve the thesaurus and the algorithm. Most comments are concerned with not finding the correct code despite correct input.
The old ICHPPC-2-Defined criteria [6] will be replaced by the international ICPC inclusion criteria. As soon as the WONCA Classification Committee approves of these criteria, Dutch inclusion criteria based on the international standard will be developed to ensure better classification.
In 1997, a stand-alone version of the semi-automatic coding program will be developed, mainly for educational and scientific purposes.
Finally, a layman 'translation' of the Short Titles and subtitles will be developed as part of the promotional research of the author. One of the tasks that will be performed is a total review of all word clusters.

5. Conclusions

Through the semi-automatic coding system, the ICPC has become easier to use. This is demonstrated by the fact that the use of ICPC has significantly increased since the introduction of the coding system (from 33% of all general practitioners in 1993 [7] to over 60% in 1997 (preliminary results of the NUT-III-project)). The Reference Model 1995 (the blue print of all GP information systems in Holland) can cause a dramatical increase of the use of ICPC in the electronic patient record increases dramatically because of the introduction of the episode oriented registration. The ICPC will be an anchorpoint for episode oriented registration and be the link to prescribing medication, related consultations, other information sources, printing patient information leaflets etcetera.
Maintaining the ICPC and the coding system is a continuing process. The introduction of ICPC-2 forms a great challenge in the near future.

Acknowledgements

The author wishes to thank Dorry Cox, Henk Westerhof and Han van Overbeeke of the Dutch College of General Practitioners for their comment on the manuscript.

References

[1] Lamberts H, Woods M, eds. International Classification of Primary Care. Oxford, Oxford University Press 1987

[2] Kanter JS de, Lamberts H, Mulder JD. ICPC International Classification of Primary Care, Thesaurus, Dutch translation. Part I Systematic list. Part II Alfabetic list. Leiden, January 1989

[3] Lamberts H, Wood M, Hofmans-Okkes IM (ed.). The International Classification of Primary Care in the European Community with a multi-language layer. New York, Oxford University Press 1993

[4] Boersma JJ, Gebel RS, Lamberts H. ICPC with Dutch subtitles. Utrecht, Dutch College of General Practitioners 1995

[5] Lamberts H. In het huis van de huisarts. Verslag van het Transitieproject. Lelystad, Meditekst 1994

[6] WONCA. ICHPPC-2 Defined. Oxford, Oxford University Press, 3rd edition 1983

[7] Westerhof HP, Van Overbeeke JJ (red.). Report of the NUT-II-project, part II. Utrecht, Dutch College of General Practitioners, Utrecht, 1994