NOW AND THEN

The language of rheumatology. II: Classification and grouping

Deborah Symmons

This is the second of two papers concerned with the language of rheumatology in an age of electronic data transfer. The first paper dealt with the definition and coding of clinical terms;¹ this one is concerned with the grouping of data. In particular, it focuses on the International Classification of Disease (ICD) and Healthcare Resource Groups (HRGs).

Classification

Nomenclature or terming is important to those who are interested primarily in individual events. Those, such as epidemiologists, who are interested in populations and in compiling statistics need a classification or taxonomy. A statistical classification of diseases must cover the entire range of conditions within a manageable number of categories. There is no single 'correct' way to classify a series of diseases, procedures, or investigations. William Farr, the first statistician to the Registrar General for England and Wales, wrote: 'classification is a method of generalisation. Several classifications may, therefore, be used with advantage; and the physician, the pathologist, or the jurist, each from his own point of view, may legitimately classify the diseases and the causes of death in the way that he thinks best adapted to facilitate his enquiries, and to yield general results.' The preferred system will thus depend on its use. An analogy can be drawn with maps: geographical data can be represented in either physical or political maps; the type of map required will depend on the intended application. Classifications may be:

- (1) Hierarchical, that is having the structure of a branching tree, each level giving more detail than the one above.
- (2) Partitioning, where the classes are mutually exclusive.
- (3) Clumping, where the classes can overlap.

The development of ICD10

THE HISTORY OF THE INTERNATIONAL CLASSIFICATION OF DISEASES (ICD)

William Farr was the first to suggest that there should be a uniform international classification of causes of death. In 1853, the First International Statistical Congress asked Farr to prepare 'une nomenclature uniforme des causes de décès applicable à tous les pays.' Farr suggested five groups of causes of death which are still found in the ICD today: epidemic diseases, constitutional or general diseases,

local diseases arranged according to anatomical site, developmental diseases, and diseases resulting from trauma. Farr's method of grouping diseases was accepted, but his list of 138 possible causes of death was not. In 1891, the International Statistical Institute asked Jacques Bertillon to prepare a new classification of causes of death.² This time the list was widely adopted.

The first International Conference for the revision of the Bertillon or International Classification of Causes of Death was held in August 1900.³ The sixth decennial revision conference, in 1948, was the first to include non-fatal conditions and the first to involve the World Health Organisation (WHO), which then assumed responsibility for the system. With the seventh revision (ICD7, 1955), the title 'International Classification of Diseases' first appeared.

Up to and including the eighth revision (1968), priority was given to the underlying cause of death, or to the chief anatomical site involved when the aetiology was not known. If a patient died from heart failure secondary to viral myocarditis, the cause of death would be coded as viral myocarditis. Clinical records, however, emphasise the principal condition treated (in this case heart failure).

ICD9 (1975) attempted to deal with this problem by introducing the dagger/asterisk (†, *) option. The dagger is the primary code, used to identify the underlying disease, especially for mortality statistics. The asterisk is a secondary code, used to indicate manifestations or complications. So, for example, psoriatic arthritis was coded as 696† 713.3*: 696 is the code for psoriasis, and 713.3 is the code for 'arthritis secondary to a dermatological condition'. Each revision has proved a greater task. Work on ICD10 began in September 1983. The 10th Revision Conference was held in September 1989, and greatly expanded the listings to include manifestations, syndromes, and newly recognised conditions. A new title, 'The International Statistical Classification of Diseases and Health Related Problems' was adopted.⁴ It was finally adopted in the UK on 1 April 1995.

The ICD has been criticised from several perspectives. Some people regard the absence of diagnostic criteria and of a multilingual glossary as a weakness. Long intervals between revisions are inappropriate; but changes are expensive and interrupt continuity in long term databases. Despite its shortcomings, the ICD

remains the international gold standard for disease classification. From now on, continuous maintenance of the classification will be attempted.

THE STRUCTURE OF THE ICD

The ICD is a single variable axis classification. The term 'single' refers to the fact that each disease has only one code, even if it can affect several systems. The term 'variable axis' implies that, rather than the ICD being one long list, diseases are grouped together in chapters based on anatomical site. All diseases have a three digit code. Some, but not all, have additional digits after the decimal point. The ICD is also hierarchical—the codes are built up in a step-wise fashion. Only very broad categorisation is indicated by the first character, whereas progressively increased specificity is conveyed at third, fourth, and fifth character levels. The value of a hierarchical structure is that it facilitates aggregation of data.

Certain conventions are followed in the assigning of codes. When a fourth digit exists, it should be possible to allocate to every patient with that particular disease a four digit code. For this reason .8 is used for 'other specified conditions' (sometimes designated NEC—not elsewhere classified) and .9 for 'unspecified' (sometimes designated NOS—not otherwise specified) (table 1).

INNOVATIONS INTRODUCED WITH ICD10

ICD10 represents the most major reworking of the system since 1948.⁵ The main innovation is the use of an alphanumeric coding system of one letter followed by two numbers at the three character level. As a result, the size of the coding frame has more than doubled, from 1000 (000–999) in ICD9 to 2600 (A00–Z99) in ICD10. Most chapters have been assigned a unique letter or group of letters, each able to provide 100 three character categories. All letters except U have been used. U, and some three character categories within chapters, have been left free for possible additions between revisions. ICD10 uses 2033 (78%) of the possible 2600 codes. The number of chapters has been increased from 17 to 21. The dagger and asterisk system has been clarified: in ICD9, coders could match any two codes together with a dagger and asterisk; in ICD10, each chapter has a limited number of asterisk categories. Great attention has been paid to making ICD10 'user friendly'. The layout is very clear, and the instructions for allocating codes have been clarified.

ICD10 has been published in three volumes. The first, which includes the tabular lists at the three and four digit levels for mortality and morbidity, definitions, and nomenclature

regulations, was published in January 1992. Volume 2, the instruction manual, was published in 1993, and Volume 3, the alphabetical index, was published early in 1995.

RHEUMATOLOGY IN ICD10

Diseases of the musculoskeletal system had 30 three character codes in ICD9; in ICD10 they have 100 codes (all of Chapter XIII-M). The connective tissue diseases no longer share a single three character code, and vasculitis has been transferred from the blood vessels chapter.

CODING ANATOMY IN ICD10

In the musculoskeletal chapter of ICD9, the fifth digit was reserved for optional identification of the anatomical site of involvement (table 2). Because there were only 10 codes available to describe all possible sites, this system was inherently unsatisfactory. The introduction to Chapter XIII in ICD10 states: 'As local extensions or specialty adaptations may vary in the number of characters used it is suggested that the supplementary site classification be placed in an identifiably separate position e.g. in an additional box.' This is a welcome development because:

- (1) It allows expansion of anatomical detail in a second digit of the 'anatomy box'.
- (2) It facilitates searching in a computerised database—whether for clinical, administrative, or audit purposes.
- (3) It leads to greater consistency with other chapters of ICD10.
- (4) It allows for expansion of diagnostic detail without making the codes unwieldy in length.

ICD10 R&O

A version of ICD9 for rheumatology and orthopaedics, called ICD R&O,⁶ was commissioned and published by the International League against Rheumatism (ILAR), but never achieved wide distribution or acceptance. It only included codes of direct relevance to rheumatology and orthopaedics. Anatomy was coded as an optional fifth digit (table 2), as a result of which expansions of ICD R&O had to be at the sixth digit and beyond. This made the codes very cumbersome, especially as the option to code anatomy was often not used. The International Conference for ICD10 recommended that WHO should act as a clearing house for specialty based adaptations. It was therefore in collaboration with WHO that ILAR established a committee to produce ICD10 R&O under the chairmanship of Professor William Felts from Washington, USA.

Table 2 Site codes used in ICD9 and ICD10

0	Multiple sites
1	Shoulder region
2	Upper arm
3	Forearm
4	Hand
5	Pelvic region and thigh
6	Lower leg
7	Ankle and foot
8	Other (eg head, neck, vertebral column)
9	Site unspecified

Table 1 Use of .8 and .9 at the fourth digit in ICD10

M00	Pyogenic arthritis
M00.0	Staphylococcal arthritis
M00.1	Pneumococcal arthritis
M00.8	Arthritis caused by other specified bacteria
M00.9	Pyogenic arthritis, bacterium not specified

Table 3 Example from ICD10 R&O of the code for a child with pauciarticular juvenile arthritis of the knee with eye involvement

Level of specificity	Character	ICD rubric	Information content
Lowest	1	Mxx.xx	Diseases of the musculoskeletal system and connective tissue
	2	M0x.xx	Arthropathies and related disorders
Intermediate	3	M08.xx	Juvenile arthritis
	4	M08.4x	Pauciarticular juvenile arthritis
Highest	5	M08.40	Pauciarticular juvenile arthritis with eye involvement
Anatomical site		M08.40 [67]	Knee joint

GROUND RULES FOR ICD10 R&O

The remit for ICD10 R&O was to expand the existing ICD10 codes. No three or four character codes could be altered. Thus:

(1) All diagnostic extensions had to be at the fifth digit or beyond.

(2) All anatomical extensions had to be at the second digit of the anatomical code box or beyond.

(3) No new asterisk codes could be created. Dagger categories could be expanded if required.

(4) In keeping with the general layout of ICD10, the convention for the fifth character was:

.x8 = due to a specified disease or condition.

.x9 = not due to a specified disease or condition.

Work on ICD10 R&O was completed in December 1992. Publication has been delayed until the index of ICD10 was finalised. ICD10 R&O will include all codes from ICD10 commonly used by rheumatologists and orthopaedic surgeons, not just those in the musculoskeletal chapters. Expansions have been made at the fifth digit level (and in a few cases the sixth digit). Table 3 shows an example.

The anatomical site code is recorded in a separate box of two digits—the first digit being the official WHO code (table 2) and the second the ICD10 R&O extension.

Grouping

HEALTHCARE RESOURCE GROUPS (HRG)

Casemix groups are tools for data analysis. Coded patient data are used to create groups of patients who have certain similarities. Groups may be devised comprising patients whose treatment has similar costs (isoresource groups), or with similar expected outcomes (isoprognosis groups).

The first example of such groupings were the Diagnosis Related Groups (DRGs) developed in the USA. These have been criticised for being too clinically and resource diverse. In the United Kingdom, a set of Healthcare Resource Groups (HRGs) for inpatient admissions has

been developed by the National Casemix Office in collaboration with clinicians. HRGs are devised using a number of pieces of information already collected in minimum datasets. They include ICD diagnosis codes, Office of Population Censuses and Surveys codes for procedures, age, and comorbidity. The system should allow patients to be allocated to groups that are clinically similar and use similar amounts of resources. However, the only indicator of resource used in the minimum dataset is duration of hospital stay.⁷

Version 1 of the HRGs was largely based on DRGs. Version 2 has recently been released. Version 2 was developed in consultation with a clinical working group leader for each specialty. Both have also been evaluated statistically by the National Casemix Office using the national dataset (about 9 million episodes for 1991/2) to ensure that they are homogeneous, both clinically and in terms of duration of stay. HRG 2 was widely distributed for consultation in November 1994. In addition, an independent assessor for the specialty, nominated by the appropriate Royal College, provided quality assurance. Details of the final groupings were released in August 1994. It is already accepted that there will be a need for an HRG 3, and this is scheduled for May 1997; it will use further refinements based on the work of the Clinical Terms Project and Read coding.

HRGs IN RHEUMATOLOGY

Rheumatology HRGs are to be found in Section H: the Musculoskeletal System. This section also includes HRGs for trauma and elective orthopaedic surgery. The majority of rheumatological inpatient admissions will be coded under eight groups (table 4). These groups performed well in describing duration of stay. Rheumatologists will probably be more concerned, however, about the next step—the development of outpatient groupings—than with inpatient HRGs.

Table 4 Rheumatological Healthcare Resource Groups, Version 2

Group	Description
h19/20	Non-inflammatory back and joint problems (age over 74 or with comorbidity)
	(age under 75 without comorbidity)
h28/29	Acute and chronic inflammatory joint disease or connective tissue disease (age over 64 or with comorbidity)
	(age under 65 without comorbidity)
h30/31	Acute and chronic infection of bones or joints (age over 59 or with comorbidity)
	(age under 60 without comorbidity)
h36/37	Soft tissue disorders. Signs and symptoms of musculoskeletal system/connective tissue (age over 69 or with comorbidity)
	(age under 70 without comorbidity)

THE USE OF HRGs

HRGs have been developed to help in the process of costing and contracting. In 1994–5, all hospitals were required to cost two out of three named specialties (orthopaedics, ophthalmology, gynaecology) using HRGs.⁸ By 1997–8, all specialties will be costed as part of the contracting process. This will enable purchasers to compare prices on a more equitable basis than at present. Although cost information will be available by HRG, it will not be necessary to contract at HRG level.

HRGs have been developed in England. In Scotland, the costing project is considering the introduction of HRGs. In Wales, preliminary experiments are continuing with DRGs and a decision on whether to move to HRGs will be taken in 1995. In Northern Ireland, no decision has yet been made about casemix groups.

HEALTH BENEFIT GROUPINGS

There is anxiety about groupings that are based on duration of stay—and using a methodology that excludes outliers. Thus there is no allowance in the system for patients with exceptionally prolonged stays in hospital. The aim is to move towards groupings that require similar care and have similar outcomes. It is believed that these would then facilitate audit and the assessment of quality.

Discussion

These papers focus on several related projects. The ICD classification is essentially 'flat' and designed for use in a paper format. As it is designed for use in all countries of the world and for all health personnel, this is entirely appropriate. However, the sophistication offered by computer software does make it possible to develop classification schemes that are not flat, but multidimensional. The Read codes are such a scheme. It would be impossible to produce the Read codes in a usable paper form.

There is another important distinction between ICD10 and the Read codes. ICD is intended to be applied by coders, and they are required to code at a particular level (either three, four, or five digits). If a particular disease has a four digit expansion, then it must be possible to allocate to every patient with that disease a four digit code. Hence the use of the categories NEC and NOS. In contrast, Read codes aim to code directly from the terms in the clinical record. Clinicians do not use terms such as 'other inflammatory joint disease' and 'osteoarthritis—not otherwise specified'. There is thus no requirement for such terms in the Read hierarchy.

It remains to be seen whether the additional detail supplied in ICD10 R&O will ever be used. Most physicians have little interest in coding diagnostic data, and the task is left to coders. Unless those demanding the data require the degree of detail now encompassed in ICD10 R&O, the tool may be used only by the enthusiast.

While Read codes contain tens of thousands of terms, and ICD10 contains thousands of terms, HRGs include only a few hundred. HRGs provide the broader view needed for costing and contracting. Casemix groups are tools for data analysis—and of no value in describing the individual patient.

- 1 Symmons D, Sant S. The language of rheumatology. I: Nomenclature and coding. *Ann Rheum Dis* 1996; 55: 4–6.
- 2 Bertillon J. Classification of the cause of death. Transactions of the 15th International Congress on Hygiene and Demography, 1912; 52–5.
- 3 Israel R. The history of the International Classification of Diseases. *Health Trends* 1990; 22: 43–4.
- 4 *International Statistical Classification of Diseases and Related Health Problems*, 10th revision, vol 1. Geneva: World Health Organisation, 1992.
- 5 Ashley J. The International Classification of Diseases: the structure and content of the Tenth Revision. *Health Trends* 1990; 22: 135–7.
- 6 *ICD R&O—An application of the International Classification of Disease to Rheumatology and Orthopaedics*. New York: International League against Rheumatism, 1985.
- 7 Buckland R W. Healthcare resource groups. *BMJ* 1994; 308: 1056–7.
- 8 NHS Executive EL(93)92. *Comparative data for the contracting process—acute services*. Leeds: NHS Management Executive, 17 December 1993.