The language of rheumatology. I: Nomenclature and coding

Deborah Symmons, Susan Sant

Effective communication is essential in an efficient health service. Increasingly, such communication involves the electronic transfer of data from one computer system to another. The Information Management Group (IMG) of the National Health Service (NHS) plans eventually to introduce an electronic patient record which will be accessible wherever the patient is treated. If this plan is to function effectively, then all patient based information collected within the NHS must be consistent, unambiguous, and exchangeable. Thus it has become important to consider very carefully the way in which medical terms are coded and then grouped.

This paper deals with the development of a national thesaurus of agreed clinical terms (the Clinical Terms Project) and the assignment of computer codes (Read codes) to these terms. A subsequent paper will discuss the need for systems of classification (such as International Classification of Disease (ICD)10) and grouping (such as healthcare resource groups) for statistical analysis, research, costing, and resource management.

NOMENCLATURE
Nomenclature (or terming) deals with the selection of names for specific items or concepts. A nomenclature is the equivalent of a thesaurus or catalogue. Nomenclatures do not, however, include definitions of terms. Nomenclatures need to be exhaustive. The terms can be presented in any order (alphabetical order is often chosen for convenience), and there is no particular emphasis on any one term. In a nomenclature any one item—for example a disease, drug, or procedure—is designated by a single term. In addition to this preferred term there may be permitted synonyms.

Since 1970, the Council for International Organisations of Medical Sciences has been preparing an International Nomenclature of Diseases that will complement the ICD. The main purpose of the International Nomenclature is to provide a single recommended name for every disease. A number of volumes have been completed and work is in progress on the musculoskeletal system.

CODING
Computers use numerical codes rather than free text. Coding is the process of allocating numerical codes to agreed medical terms. Coding systems should be designed to facilitate data entry and data retrieval.

The Clinical Terms Project
The Clinical Terms Project (CTP) was initiated by the IMG in 1991. The aim of the project was to assemble a thesaurus of medical terms that would be sufficiently comprehensive to cover anything that might be entered by a clinician into a patient's medical record. It was felt that the project should be led by clinicians, and so a working party was established that included representatives of the Conference of Royal Medical Colleges and their Faculties, the Joint Consultants' Committee, the General Medical Services Committee of the British Medical Association and the NHS Management Executive. The CTP was coordinated by the NHS Centre for Coding and Classification (NHS CCC) in Loughborough, and was launched as a two year initiative in April 1992. The resultant list of clinical terms, with their allocated Read codes (see below), will be used for recording patient encounters, care plans, treatment protocols, research, and audit. Collation of coded terms will aid epidemiological analysis, planning for health care needs, and the provision of central returns. There are equivalent projects involving nurses and the professions allied to medicine. More than 2000 doctors, nurses and allied professionals from all over the UK have been involved in the different Terms Projects. Integration of these projects is taking place from 1994 onwards.

The Work of the Rheumatology Speciality Working Group
The CTP was funded by a grant of £2.7 million. Approximately 40 Specialty Working Groups (SWGs) were established. The task of each SWG was to collect all acronyms and terms used in the medical record in their specialty. All synonyms were required to be identified and the preferred term chosen. SWGs liaised with one another to ensure that their lists were compatible and, between them, comprehensive. The SWGs were also asked to suggest appropriate hierarchies for terms. A hierarchy has a structure like a branching tree—each level giving more detail about the one above (table 1). Hierarchical structures have the advantage of facilitating aggregation of data.

The British Society for Rheumatology had set up an advisory group to look at Read codes
Table 1 A rheumatological hierarchy

<table>
<thead>
<tr>
<th>Level</th>
<th>Term</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Disorders of the musculoskeletal system</td>
</tr>
<tr>
<td>2</td>
<td>Inflammatory joint disease</td>
</tr>
<tr>
<td>3</td>
<td>Rheumatoid arthritis</td>
</tr>
<tr>
<td></td>
<td>Psoriatic arthritis</td>
</tr>
<tr>
<td></td>
<td>Spondyloarthrits</td>
</tr>
</tbody>
</table>

in 1990. This group became the Rheumatology SWG.1 The process of collection of terms was co-ordinated by the NHS CCC. The medical coding specialists from the CCC advised and guided the work of the SWGs, and regular meetings of the SWG leaders have ensured a consistency of approach. The lists from each SWG were submitted to the CCC by 1 April 1994. There was then a period of quality assurance, followed by an amalgamation of the different lists. Now that the lists have been amalgamated, there will be a period of two or three years during which the master list is piloted and refined. This is being done on a progressively larger scale as confidence in the thesaurus accumulates. Updates of the list are produced six monthly.

The rheumatology list covers musculoskeletal diagnoses (excluding the spine), rheumatic disease history, joint examination, drug monitoring, patient education, rheumatological investigations including synovial fluid analysis, and procedures such as joint injection. The Orthopaedic SWG was responsible for spinal disorders. The lists aim to cover terms used by the generalist in addition to the tertiary referral centre. Each member of the SWG was responsible for one of the above areas. With the help of the research worker, they extracted terms from medical records, scrutinised the contents and index pages of relevant text books, and consulted experts in the particular field. All terms found in commonly used classifications such as the ICD have been included. The work of the rheumatology SWG required wide consultation and in some areas, for example that of vasculitis, it proved impossible to reconcile all the views expressed. One aspect of drawing up a nomenclature became clear: it is impossible to meet every expectation and comply with every request!

Read codes

BACKGROUND

About 10 years ago James Read, then a general practitioner in Loughborough, began to develop a coding system that would cover the entire medical record. His codes were divided into chapters (table 2). The Read thesaurus now comprises a list of preferred healthcare terms and their synonyms, eponyms, and acronyms. Each preferred term has a unique five character alpha numeric code. The synonyms and eponyms have the same code as the preferred term. Thus the preferred term lateral epicondylitis can also be found using the synonym tennis elbow. The alphanumeric characters used are 0–9, A–Z and a–z. The letters 'I' and 'O' are excluded in order to avoid confusion. There are 656 356 768 possible codes. The codes are hierarchical.

A major advantage of Read codes is that they map back to all currently used national classifications, including ICD, Office of Population Censuses and Surveys (OPCS), and the British National Formulary. They can also be adapted to map to other classifications such as ICD10, provided that all the terms used in the classification have been allocated Read codes.

COMPUTERISED CODES IN THE NHS WORKING PARTY

A working party set up by the Joint Computing Group of the Royal College of General Practitioners and the General Medical Services Committee of the British Medical Association in 1988 recommended that the Read Clinical Classification should become the standard method of recording medical data in general practice, and that the implications of this should be considered throughout the NHS. In order to facilitate the implementation of the working party's report, the Read codes were purchased by the Secretary of State for Health in 1990 and became Crown Copyright.

THE STRUCTURE OF READ VERSION III

The original Read codes had two functions: they identified a concept (labelling function) and they defined the position of the concept within the classification system (structure function). This meant that each code could appear only once in the classification tree. With the introduction of ICD10, many clinical terms had a new position in the hierarchy without changing their meaning. This meant that the Read codes would have to be re-allocated, or they would lose their structure function. Re-allocation of codes could lead to confusion throughout the NHS. It also became clear during the CTP that, even within a single specialty, there was often more than one way to group and classify terms.

It was therefore decided that, in Read version III, the labelling function and the structure function of the codes should be separate properties. Thus the Read code will remain a unique identifier for a particular concept, but the codes can be grouped—by use of appropriate software—into whatever display tree may be required. Consequently, existing Read codes need not be changed, but simply added to. The separation of the labelling and structure function also means that diseases can appear at more than one position in the hierarchy.

The display hierarchy of core terms will consist of a series of 'parent' and 'child' links.
Each parent can have many children and each child can have many parents. For example in table 1, inflammatory joint disease (the parent) has three children, one of which is spondyloarthritis. Spondyloarthritis might also appear in a hierarchy based on back pain. Thus spondyloarthritis may have more than one parent.

One consequence of this decision is that each term must 'stand alone' as it may appear in many positions. In ICD, some diseases are grouped under the umbrella 'not otherwise specified.' Thus in ICD10, juvenile psoriatic arthritis has the code M08.8 'other juvenile arthritis.' There is a Read code for juvenile psoriatic arthritis, but not for 'other juvenile arthritis', as this is not a term used in the medical record. In the process of mapping Read codes to ICD10, juvenile psoriatic arthritis will map to M08 'juvenile arthritis.'

During the course of the CTP, specialties proposed hierarchies for use in Read version III. Up to eight levels (that is, eight axes of classification) were allowed. It was important not to mix concepts within a single axis. It was also desirable for the levels to follow a logical sequence, mirroring the thought processes of the clinician. In composing a hierarchy, the needs of data retrieval are as important as those of data entry. Those involved in data entry are concerned with accurate coding of the individual patient; the needs of those involved in data retrieval are more diverse. Often they are interested in identifying meaningful groups of patients whether for costing, health care planning, or clinical research. It is easier to retrieve data if only a single code has to be entered, rather than if a long sequence of codes is required.

The Use of Qualifiers

In developing the hierarchy, it became clear that some 'descriptors' were better expressed as qualifiers rather than being embodied within the hierarchy. ‘Core terms’ express the essential nature of the problem, diagnosis, and procedure, while optional ‘qualifiers’ add as much detail as the expert might want. The use of qualifiers helps to prevent terms from becoming too long. It also leads to economy in the size of the coding frame, as many qualifiers can be applied to multiple terms (for example: left, right; mild, moderate, severe). Qualifiers will be presented automatically when entering data. For example, there are many ways of classifying osteoarthrosis. It can be described as primary or secondary; nodal or non-nodal; erosive or non-erosive; and monoarticular, oligoarticular, or polyarticular. Any combination of these descriptors is possible. If these descriptors are used as qualifiers, then the clinician can select his choice. If they are used in a hierarchy, duplication is required at each level (table 3).

Application of Read Codes

There is some scepticism about the feasibility, or indeed desirability, of coding the entire medical record. However, the main concern of critics of Read codes is that the size and complexity of the thesaurus produced will render it unusable. No one complains that it is impossible to write a book containing all the words in the Oxford Dictionary. Authors would, however, complain if they looked up a word in the dictionary and could not find it. Similarly, it is not essential that computer applications should use the entire Read thesaurus. A computerised rheumatology drug monitoring programming will not need the Read codes for surgical operations on the breast. Nevertheless, all programmers of computer systems for use in the NHS should be able to find Read codes for the terms they wish to use.

The Rheumatology SWG comprises the following people: Peter Dawes, Peter Fowler, Gill Hoisey, Tom Kennedy, Philip Platt, Adam Young, Deborah Symmons, and Eddie Tunnicliffe. The research worker was Susan Sant.

The language of rheumatology. I: Nomenclature and coding.

D Symmons and S Sant

*Ann Rheum Dis* 1996 55: 4-6
doi: 10.1136/ard.55.1.4

Updated information and services can be found at:
[http://ard.bmj.com/content/55/1/4.citation](http://ard.bmj.com/content/55/1/4.citation)

**Email alerting service**

Receive free email alerts when new articles cite this article. Sign up in the box at the top right corner of the online article.

**Notes**