Reprinted from the **BRITISH MEDICAL JOURNAL**, *26th March 1988, 296, 903-905*

Contemporary Themes

Körner, nomenclature, and SNOMED

RICHARD EARLAM

The latest efficiency drive in the National Health Service is based on management budgeting for the individual consultant or his or her unit. This scheme, however, will be difficult to operate because satisfactory data about patients and their treatment are not available. Two parallel systems are based on the collection of demographic facts by ward clerks and the coding of disease and operations from discharge summaries by lay staff. Most information collected has not been assessed by anyone, certainly not by doctors. This lack of proper data prompted demands for a new information system and led to the government setting up a committee, chaired by Mrs Edith Körner, to make recommendations. The new Körner data proposed by that committee represent a great step forward because information will be standardised and for the first time doctors must take part in coding. It is essential for the information and coding to be accurate.

Doctors will need to understand, firstly, what is meant by the word "disease" and, secondly, how complex the subject is. One difficulty in introducing medical audit has been to get the medical profession to agree terminology, whether it be for a disease or an operation. The NHS needs a system, with the same basic data for both audit and management budgeting, that is practical, can be put on a computer, can develop as medicine changes, and provides data for research.

This article aims to combine the practical with the intellectual. The present system of coding is old fashioned, and though the new Körner data are a step forward they are unfortunately based on the International Classification of Diseases, the latest (ninth) edition of which was published in 1975. When doctors begin to use it they will realise that the limitations of this coding system are as much the result of its precomputer origins as the complicated use of the word disease. The intellectual problem is to understand that disease covers a multiplicity of usages that must be analysed to obtain meaningful coding. Finally, I describe advantages of another fully computerised system based on the Systematised nomenclature of medicine (SNOMED).[1]

The London Hospital, London E1 1BB
RICHARD EARLAM, MCHIR, FRCS, consultant surgeon

Körner data

In response to the royal commission's report on the NHS (Merrison report) in 1979, which emphasised that the information available to help decision makers left much to be desired, the Department of Health and Social Security set up the NHS/Department of Health and Social Security Steering Group on Health Services Information in February 1980 chaired by Mrs Edith Körner. Its terms of reference were (1) to agree, implement, and keep under review principles and procedures to guide the future development of health services information systems; (2) to identify and resolve issues on health services information requiring a coordinated approach; (3) to review existing health services information systems; and (4) to consider proposals for changes to, or developments in, health services information systems arising elsewhere and, if they seemed acceptable, to assess priorities for their development and implementation.

Seven reports were published between 1982 and 1984, their main purpose being to identify a minimum data set that would be used by all districts in a standardised format and also be capable of centralisation and comparison.[2] The steering group's main concern was with information for health services management. Previously data on patients and bed usage had been derived from the Hospital In-Patient Enquiry run by the Office of Population, Censuses, and Surveys; SH3 returns organised by the Department of Health and Social Security; and regional Hospital Activity Analysis systems. These provided demographic data on the patient, prepared by the ward clerk as SH3 returns, and clinical data and details of diagnosis and operation, coded from the discharge summary by Hospital Activity Analysis coding clerks. The Körner data will now bring together these different systems and base the information on consultant episodes while the patient is in hospital, each of which will be coded separately by the medical profession before transfer or discharge. Ward clerks will fill in the demographic data of hospital number, sex, home address, postcode, date of birth, marital state,

etc. For the first time doctors are part of the data collection system because they must provide up to six diagnoses (based on the ninth edition of the International Classification of Diseases) and state the operation, using the classification of operative procedures of the Office of Population, Censuses, and Surveys (a new edition will be published in 1988 to replace the 1975 edition). Another change is that data will be collected annually from 1 April rather than 1 January to coincide with the financial year. So the medical profession will be obliged to take an interest in coding and not leave this important subject to clerks or paramedical staff. Although Körner data are being collected for management and so are oriented that way, they can equally provide data for monitoring or audit by clinicians, provided that the diseases and operations are correctly coded.

Coding versus nomenclature

The present International Classification of Diseases (ICD) is a coding system established by the World Health Organisation. A single axis system, it has the limitation that each concept has to be coded with a separate number. It becomes crowded and cannot include all conditions, so this leads to the cramming of too many ideas into a single four figure code number, thus losing detail. Additionally, coding for Hospital Activity Analysis centrally is to only three figures, which implies that most detail is lost anyway. A double axis coding for tumours in the International Classification of Diseases using the suffix O (for oncology) has been developed by the World Health Organisation. Tumours have an anatomical as well as a pathological code, and this automatically reduces the total number of codes needed. There is no internationally agreed classification of operations; the United Kingdom has a classification of operations (Office of Population, Censuses, and Surveys) that has just been revised for use in 1988. The anatomical site has an alphabetical prefix, but essentially this is a single axis system with a separate code number for each operation.

Multiaxial systems usually use anatomy or topography for the first axis and disease or morphology for the second. For example, in arterial disease subdividing the coding for each particular artery would be avoided if one axis contained all the arteries and the other axis listed the possible diseases such as aneurysm, atheroma, embolus, thrombosis, or trauma. Coding implies the grouping of similar concepts; but in any coding system pigeonholes are for pigeons who have a home and then it is easy. In most instances, however, it is more like Trafalgar Square because the pigeons quite simply do not have a pigeonhole; they live on ledges without addresses. Nomenclature is the exact description of an individual, disease, or procedure.

The importance of this becomes apparent in computer terms. By using more digits the description becomes more specific until it becomes a nomenclature. Each of the extra digits can subdivide the previous number or numbers into 10 further subdivisions and so on. This is what is meant by a hierarchical system. The computer can code by using one or two digits, but in a hierarchical system the first digits are the main subdivisions or codes and the later ones are the specific nomenclature. Retrieval of the coded groups can then be presented as collections of specific individual descriptions or nomenclatures. With a computer multiple axes, with many digits for each axis, may be stored so that infinite numbers of combinations may be added later and the system can accommodate different degrees of detail. With computers a multiaxial hierarchical nomenclature system like SNOMED has much more flexibility than a coding system. We are now at the stage where a desktop personal computer can handle this system, and coding systems based on a punch card and knitting needle are obsolescent.

Use of the word "disease"

In discussing the merits of nomenclature versus coding it is essential to return to basic thoughts about disease and how the word is used in nosology—the science of classifying diseases. Professor J G Scadding approached the problem through his specialty of respiratory disease, pointing out that the word disease when applied to the following examples was based on different usage.[3-7]

(1) Bronchial carcinoma—based on morbid anatomy with two axes: (*i*) topographical and (*ii*) morphological.

(2) Tuberculosis—a term used rather inaccurately to describe abnormalities of any part of the body, but invariably requiring identification of the tubercle bacillus and described under three axes: (*i*) topographical; (*ii*) morphological; and (*iii*) aetiological (the tubercle bacillus).

(3) Bronchial asthma—a controversial term which some have used to describe an allergic condition characterised by wheezing, dyspnoea, orthopnoea, and cough but which Scadding suggests should be used for "a disease characterised by variable dyspnoea due to widespread narrowing of peripheral airways in the lungs and varying in severity over short periods of time either spontaneously or as a result of treatment." This cannot be properly described under the previous three axes; it requires more detail because there is disagreement, so its nomenclature might be (*i*) topographical—bronchi; (*ii*) morphological—narrowing of the bronchi temporarily not permanently, so unreliable; (*iii*) aetiological—inexact because allergies are not always found; (*iv*) symptomatic or functional—which would list dyspnoea and other symptoms; and (*v*) disease—using the word asthma for all to agree or disagree with.

Similarly, chronic bronchitis is a "disease" that can have different meanings. The usual clinical descriptive term now competes with the topographical-morphological term based on glands secreting excessive mucus.

These examples emphasise how a clinical diagnosis can be made. Historically the earliest method was pattern recognition of similar groups of symptoms and signs. With the development of morbid anatomy another basis for diagnosis developed. When further laboratory tests, especially microbiological ones, led to the identification of aetiological agents there was yet another method for diagnosis. "At whatever point the diagnostic process comes to a halt, its end-result is expressed in terms of diseases."[2]

It is essential in managing a patient to know that treatment or procedures can be started without a definite diagnosis; though an accurate diagnosis is not always possible, the patient must be treated. Doctors are then forced to cheat and with the present system choose an inaccurate code of classification compatible with the treatment just to complete the form. The classic example is the diagnosis on death certificates; this is notoriously inaccurate and unreliable and not always similar to necropsy findings. A doctor cannot write, "I do not know the exact diagnosis, I do not think it really matters either, but I did not think it was worth investigating an 80 year old when he was alive, nor do I think it worth doing a necropsy now." Such a conclusion is not necessarily medicine, but how should the diagnosis be coded?

"The concept of a disease is thus an abstraction from the reality of phenomena observed in patients, useful because it permits of thinking, speaking and writing in generalisations."[8] The methodology of using diagnostically related groups cannot deal with such problems and certainly does not solve them intellectually, whereas SNOMED does because its system has recognised how doctors make a diagnosis of disease or not, as the case may be, despite investigation, treatment, and operation. Surgery is similarly inexact because appendicectomy does not imply a confirmed diagnosis of acute appendicitis.

SNOMED

The systematised nomenclature of medicine (SNOMED) was developed from the pathological system (SNOP) for morbid anatomy, which has two axes) (*i*) topography or anatomy, and (*ii*) pathology or morphology. SNOMED (figure) is more complicated and requires six axes: topography (T), morphology (M), etiology (spelt according to American style) (E), function (F), disease (D), and procedure (P).

The essential points about SNOMED and SNOP are that they are

multiaxial; hierarchical; nomenclatures (they describe each individual in full detail); capable of coding, by using less than the full numbers of digits; and for some reason unknown and untested in the United Kingdom.

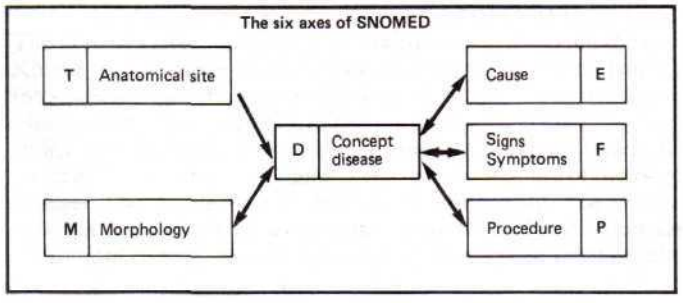
There is no need to have actually played on a one armed bandit to know the principle of a fruit machine with four windows. This is what multiaxial means. The fact that it is difficult to win shows that the method is ideal for a unique description as the combinations are almost infinite. Each of the windows in SNOMED is alphanumerical, which means that the numbers 0-9 plus X and Y add up to 12—the duodecimal system—and as many additional digits can be added as required.

Topography and morphology in SNOP are the basis of morbid anatomy and are compatible with the World Health Organisation's classification of tumours (ICD-O). In SNOMED aetiology is the result of laboratory investigations, function is the axis for signs and symptoms, and disease is for the diagnosis in common usage. Is it fortuitous that this system is similar to Professor Scadding's concept of disease and its method of diagnosis—or perhaps great minds just think alike? Fairly simple notions like tumours, trauma, and inflammation are best coded by the T and M axes. Traditionally there is no need to code signs and symptoms separately, but for completeness or for research into pattern recognition they can be listed under F. The E axis often helps but may confuse. For a long time aetiologies, particularly agents of infections, were seen as absolute causes of clearly defined disease, but the originators of SNOMED well appreciated that this cause and effect relation was not unique. D for Disease is the diagnosis under which a discharge diagnosis is usually made, but it includes all usages of the term.

The sixth axis, Procedure, contains a list of administrative, diagnostic, therapeutic, and preventive procedures, but for surgical procedures it is pure genius. In the old days cynical doctors described surgery as doing only two things properly: letting out pus and relieving obstruction. SNOMED does better for the modern barber surgeon and with a single digit, under mutually exclusive headings, forces the surgeon to decide which of many alternative procedures he has done: 0=incision; 1=excision; 2=injection, implantation; 3=endoscopy—that is, look at it but do nothing; 4=repair, transplant; 5=destruction—for example, burn it, freeze it, poison it, kick it; 6=closure; 7 and 8=manipulation.

When the object of the procedure has already been described in a histopathological report by topography and morphology, nothing could be more simple than to add a third axis P.

Section X of the sixth axis lists procedures in radiography, radiotherapy, nuclear medicine, and ultrasonography. Section Y contains nursing, home care, and disability evaluation procedures. All of these contain comprehensive lists that will make the administrator weak with envy at the potential choice for accounting but eventually bankrupt. So section Y will enable the ancillary procedures that accompany any surgical operation to be assessed for audit.



The six axes of SNOMED.

Conclusion

One of the most famous scientific classifications was by Carl von Linne, Linnaeus in Latin. His original plant coding used the topography of number of pistils and stamens. But in 1750 he progressed to a binomial nomenclature for animals and plants based on genus and species—for example, *Homo sapiens* and his best friend, *Canis familiaris*—rather like human beings use surname and forename to identify themselves. The old system of identifying cars with three letters and three numbers is also a double axis system. Coding in the International Classification of Diseases and the new operation procedures list of the Office of Population, Censuses, and Surveys are single axis systems. Multiaxial systems are more flexible and can easily be managed by computer. When they are hierarchical not only may they be completely specific for the individual—that is, a nomenclature—but they can also be coded to any degree of complexity depending on the number of digits. SNOMED is such a system.

Another advantage of SNOMED over the present coding system is that it allows for different uses of the word disease: (*a*) topographical and morphological description; (*b*) a pattern of symptoms and signs; (*c*) due to an aetiological agent; or (*d*) in its agreed common usage when there is vagueness and disagreement. In both hospital and general practice diagnoses frequently cannot be made so the individual symptoms can be listed separately under axis F.

Examples from upper gastrointestinal disease will show the flexibility of SNOMED. There are three common anatomopathological diagnoses (hiatus hernia, duodenal ulcer, and gall stones) and in simplified terms there may be three surgical procedures used to treat them (repair of haitus hernia, vagotomy and pyloroplasty, and cholecystectomy) but they are carried out only in a minority of cases. So it is essential for audit of medical practice to add subdivisions under axis F to these diagnoses and include symptoms and complications. Numerically most of these diagnostic anatomopathological end points are actually asymptomatic. SNOMED can deal with these problems but our present system of coding cannot. The reasons why SNOMED can cope with this complexity are that it is multiaxial and hierarchical and based on the same methodology with which doctors diagnose disease and use the word.

Coding requires automation for accuracy, and computers must be used to help the medical profession.[8-10] Desktop word processors can produce documents without spelling mistakes. It is difficult to produce a summary of a hospital admission without mentioning the diagnosis, and this word or combination of words can be searched for to enable automatic coding to be done. Experimental systems have already been set up for SNOP and SNOMED in America. Ward discharge summaries, pathology reports, and descriptions of operations could also be encoded automatically.

References

1 College of American Pathologists. *SNOMED. Coding manual; alphabetical index; numerical index.* 2nd ed. Skokie, Illinois: College of American Pathologists, 1979.
2 Department of Health and Social Security. *Health services development. Reports of the steering group on health services information.* Implementation programme. London: DHSS, 1984. (HC(84)10.)
3 Scadding JG. Principles of definition in medicine with special reference to chronic bronchitis and emphysema. *Lancet* 1959;i:323-5.
4 Scadding JG. Meaning of diagnostic terms in broncho-pulmonary disease. *Br Med J* 1963;ii:1425-30.
5 Scadding JG. Diagnosis: the clinician and the computer. *Lancet* 1967;ii:877-82.
6 Campbell EJM, Scadding JG, Roberts RS. The concept of disease. *Br Med J* 1979;ii:757-62.
7 Côté RA, Robboy S. Progress in medical information management systematized nomenclature of medicine (SNOMED) *JAMA* 1980;**243**:756-62.
8 Milan J, Coles T. Automatic ICD-9 codes. *British Journal of Health Care Computing* 1986;**3**:25-8.
9 Rothwell DJ, Hause LL. SNOMED and microcomputers in anatomic pathology. *Med Inf (Lond)* 1983;**8**:23-31.
10 Cameron M, Prior MC, Miao S, Barnett GO. Computer assisted pathology encoding and reporting system (CAPER). *Am J Clin Pathol* 1977;**68**:715-20.

(Accepted 3 November 1987)

Printed in Great Britain by Bourne Offset Limited, Iver, Bucks.