

Medical Subject Headings in MEDLARS

BY WINIFRED SEWELL, *Deputy Chief*

*Bibliographic Services Division
National Library of Medicine
Bethesda, Maryland*

ABSTRACT

Use of an identical subject heading list for both machine searching and publication creates problems. Necessary compromises include precoordination of some headings for publication and use of some coordinates in the computer system that are not appropriate for a printed index. Determination of the headings under which a citation will appear in *Index Medicus* and those which will be used for it only in the computer requires anticipation of the way the published index will be used for search. Methods of selecting new terms are discussed and opinions solicited. The computer will be used in several ways to update the system and provide for annual publication of the subject heading list.

THE first official list of subject headings published by the National Library of Medicine appeared in 1954 under the title *Subject Heading Authority List*. It was based on the internal authority list that had been used for publication of *Current List of Medical Literature* which in turn had incorporated headings from the Library's *Index-Catalogue* and from the 1940 *Quarterly Cumulative Index Medicus Subject Headings*. With the inception of *Index Medicus* in 1960, a new and thoroughly revised *Medical Subject Headings* appeared. It was this list which served as the basis for the *Medical Subject Headings*, 2nd edition, that was distributed to all subscribers as Part 2 of the January 1963 *Index Medicus*.

Though the number of subject headings in the second edition was a third greater than the number in the first edition, we followed the basic principles of assigning subject headings in medicine as set forth in the first edition. We are convinced of the value of using an identical authority list for the indexing of periodicals and the cataloging of books, and we regard subject headings as directional signals or vectors which, with other headings, serve to locate the essence of a particular paper or book in the universe of medical information. Rarely will a single subject heading encompass the total content of a citation.

The advent of MEDLARS added two criteria to those used for earlier medical subject heading lists. By providing for much greater coverage and deeper indexing, it thus increased the need for specificity in descriptors. In addition, it became possible not only to search for a single heading, such as ANTIBIOTICS, but also to include, in the search for that

concept, all the specific terms that are comprehended in the meaning of the larger term, in this case, for example, PENICILLIN, TETRACYCLINE, etc. This capability necessitated a delineation of all hierarchical relationships in the system, or at least all that would be useful for search purposes.

The first step in the revision of *Medical Subject Headings* thus became the sorting of all the terms in that list into broad categories. About three-fourths of all the words could be divided relatively easily into four broad groups according to the kind of word involved: Anatomical Terms; Organisms; Diseases; and Chemicals and Drugs. The remainder had to be sorted more arbitrarily. Here various kinds of words were grouped together into categories more closely resembling a classification system.

When the categories in the subject heading list had been identified, and the terms within them further related to one another, the groups were compared with various authoritative classifications and nomenclature lists. A few of the outstanding ones are listed in the bibliography at the beginning of *Medical Subject Headings*, 2nd edition, but other, more limited lists were used to solve individual problems. At the same time, suggestions from within and outside the Library were considered to see whether they brought to light important gaps in our coverage of a subject. Finally, the cross-references which in the first edition referred from a specific term not used to a more general term occurring in the list, were carefully examined to see whether the more specific term should also occur in the list.

Combined with this approach was one based entirely on the amount of material that appeared under each heading in *Cumulated Index Medicus* for 1960 or 1961. It should be pointed out that in a system which was set up exclusively to put references into a computer and to recall them in response to a specific search, the fact that a certain concept attracted a great number of references would create no particular problem. Since a given descriptor is usually coordinated with several others in a search, the fact that one heading attracted 1,000 citations would not matter if another with which it was coordinated had been applied to only 500, and both to only 50 articles. But in a printed index, 1,000 citations on a single subject present a problem no matter how helpfully they may be arranged. The most time-consuming single effort that went into the revision of *Medical Subject Headings* was to determine useful and meaningful subdivisions of the citations that clumped in groups of 100 or more under certain headings in the 1960 or 1961 *Cumulated Index Medicus*.

In this endeavor we were frequently able to take from group terms more specific entities. There are more drug names, more virus names, more specific anatomic parts, and many others. In general, we believe the *Index*

Medicus user will be pleased with these new terms even though he does have to pay the price of searching under forty-six virus headings now to find all viruses, whereas formerly he could have located them all under six. The tradeoff of being able to get material on the specific virus when he wants it justifies the increased effort required when he needs everything.

And, of course, a generic computer search can be used to find all virus articles related to some other concept. The more specific virus terms are obviously useful in a machine search as well as in the printed index, because if one wanted articles on the polyoma virus only, there would be no appropriate coordinate if we were still using the 1960 subject headings. One would have to get from the machine all the citations under the general heading VIRUSES, and look at them individually, hoping that the polyoma virus was mentioned in the title whenever it was discussed in an article.

The fact that many specific terms were taken from headings which attracted a large number of citations explains why we have some headings in a given hierarchy and omit others of equal rank. Actually, there will be found in *Medical Subject Headings* only two complete arrays of terms of equal rank—bacterial genuses and chemical elements. Though this may be changed in later editions, frequency of appearance of concepts in journal articles is probably as good a criterion as any for determining an adequate level of specificity for a given concept in a subject heading list.

It was not always possible to find, in the large collection of material under a single subject heading, other concepts that were truly more specific in that they were smaller, wholly contained parts of the original idea. In such cases, it was necessary to resort to the device of precoordination. For instance, DIABETES MELLITUS attracted some 1,000 citations in 1962. This is a term that in itself describes a fairly specific entity, so that, to divide it further, we had to combine the diabetes concept with other concepts. Eight new diabetes-related headings were added: ACIDOSIS, DIABETIC; ALLOXAN DIABETES; DIABETES MELLITUS, JUVENILE; DIABETIC ANGIOPATHIES; DIABETIC RETINOPATHY; DIABETIC NEUROPATHIES; OBESITY IN DIABETES; and PREGNANCY IN DIABETES. Two of these precoordinated headings are closely related to previous main heading-subheading combinations. DIABETES MELLITUS, JUVENILE corresponds to the former DIABETES MELLITUS—in infancy and childhood; and PREGNANCY IN DIABETES more accurately describes the previous DIABETES MELLITUS—in pregnancy.

One of the major changes in the new edition was the decision to discontinue the use of subheadings. *Current List of Medical Literature* had not followed the practice of the Library of Congress and others of listing approved subheadings under each main heading for which their use was

allowed. With the 1954 *Subject Heading Authority List*, there appeared a "Categorical Listing" of standard subheadings. "Abnormalities," for instance, was listed as a standard subheading for use with terms for organs, tissues, and regions, and "anesthesia and analgesia" was to be used under surgical procedure headings. But such subheadings could be used only for subject headings which fell within the category of headings to which they were to be applied. There were over 100 such subheadings, some of which varied only slightly according to the category of main heading with which they were used. For instance, "therapeutic use" was used under physical agents and drugs and chemicals, and "therapy" was used with diseases.

In the 1960 *Medical Subject Headings*, the number of subheadings was reduced to sixty-seven, and they could be used under any kind of main heading if the combination was not patently foolish or impossible. These sixty-seven headings were applied with more generalized meanings. For instance, the subheading "therapy" was used to mean "therapy of," "therapeutic use of" or just "therapeutic aspects."

Though this solution was simpler, many problems still remained. The use of one subheading might prevent the use of another. For instance, if a paper covered the etiology, pathology, and therapy of a disease, it might occur "up front" without further subdivision, or it might occur under the subheading which seemed most appropriate to the indexer. If he chose "therapy" in this case, the article would be lost to the searcher looking for the etiology of the disease, if he searched only under the subheading "etiology." In addition, if the subheading "diseases" was appended to the term for an anatomic part, it would not be possible to subdivide further for the therapy or complications of such diseases. A related problem is the overlap in meaning of the subheadings themselves. It is difficult, for example, to decide whether a paper on chemical biosynthesis fits best under "chemistry" or "metabolism."

But a concerted effort to find a set or sets of subheadings which would be uniform in approach, mutually exclusive, few in number, and broad in concept, proved unsuccessful. One possible kind of subheading that would divide material evenly would be based on the kind of paper. It is not difficult to divide "experimental research" from "clinical research," and "general practice" and "reviews" can also be separated, although the distinction between "clinical research" and "general practice" might give rise to some disagreement. However, a sort on this basis closely parallels a sort by journals themselves. This kind of sort has already been provided in the new *Index Medicus*. Papers under a given heading will be arranged first by language and then by journal. Thus a given article will fall into the same position in relation to any other article, no matter under what subject headings both occur.

For a large proportion of the material in the 1960 and 1961 issues of

Cumulated Index Medicus, no further subdivision seemed really necessary. Only about 1,000 headings attracted as many as 100 articles. As already described, these popular headings were examined individually for the most appropriate method of division. Usually the division was different from that which subheadings would have provided, as in the diabetes example, where only two of the eight new combinations closely paralleled former main heading-subheading combinations.

When the former main heading-subheading combination was a natural one, however, as with DIABETES MELLITUS, JUVENILE and PREGNANCY IN DIABETES, we did retain a new combined main heading. You will note, for instance, many additions to anatomic terms of the word "diseases," or more specific disease terms such as "fractures" or "injuries" or "neoplasms."

In a sense, this system is not entirely unlike that of listing individual subheadings which may be used under each main heading, but it discards the false assumption that every main heading in a system is a single idea, and treats all concepts equally.

You will observe that the decision to discontinue use of subheadings was based almost entirely on its effect on the printed *Index Medicus*. Some users have assumed that the decision was made because subheadings were not needed in the computer or because the computer system was better without them. Actually, while subheadings are not necessary, the computer system still carries provision for them, and if they were used it would be possible to search on either the main heading, the subheading, or the combination of both.

Another change in the new *Medical Subject Headings* is the designation of four types of cross-reference instead of two. The "see" references are now divided into "see" and "see under" references. The former refers from a synonymous to a preferred term, and the latter directs one from a specific to a more general term. As stated earlier, the "see under" references were examined carefully and many specific terms replaced "see under" references in the revised list.

"See also" references were divided into "see also related" and "see also specific" cross-references. "See also related" references direct the user to headings where related materials may be found. Often the terms connected by the reference approach closely related or overlapping subject areas on different axes and thus fall into different categories. Obvious cross-references, as that from an organ to its diseases, were avoided.

"See also specific" references direct one from a general term to all the specific subterms included in the system. There are only a couple of dozen such references in the *Medical Subject Headings* list, however, because this type of reference is now taken care of by the separate categorized lists. The

"see also specific" references which do remain are those in which the general terms appear in different categories from the specific ones, such as references from plants, body fluids, etc., to their specific chemical constituents.

Categorized lists of terms were printed for the first time in the 1963 *Medical Subject Headings*. As has already been described, preparation of these lists was accomplished during the examination of the 1960 list for revision. The final published listing contains thirteen main categories and a total of fifty-eight separate groups in subcategories and main categories. These categorized lists have made it possible for the user to find many more related terms than were in the former cross-reference structure. We are at present further refining them into four-level hierarchies for use in computer search.

In the subcategory B4, for instance, VIRUSES is a "first-level" term. ANIMAL VIRUSES is at the second level, ENTEROVIRUS at the third, and the three specific kinds of Enterovirus are fourth-level terms. We could ask for all citations in the Virus category B4 by simply requesting an "exploded" search on the B4 category. Similarly, we could ask for an "exploded" search on ANIMAL VIRUSES and obtain all the citations under each of the headings listed thereunder. Or we could get all the citations under all headings in the ENTEROVIRUS group, by asking for an "exploded" search on ENTEROVIRUS.

The 1963 *Medical Subject Headings* contains 5,700 descriptors, compared with 4,400 in the 1960 edition. Of the headings used in the 1960 list, 113 were withdrawn in favor of newer terms or terms more in conformity with the system of the 1963 edition. With the change from a printed index only to a printed index as merely one product of a computer-based publication and search system, the problems involved are much greater than merely those of learning to use 1,400 new headings.

The computer allows for deeper indexing than can be published in a printed index that covers between 150,000 and 250,000 items. We estimate that there will be an average of three headings per article in *Index Medicus*, but that an average of from eight to ten headings will be used for each article in the system.

One continuing important problem for the indexer is deciding for each descriptor assigned to an article whether the article should be placed in *Index Medicus* under that heading. Certain terms, such as those for geographic areas and others describing the kind of paper, cannot be used in *Index Medicus*. And such broad, general terms as METABOLISM or PHYSIOLOGY would usually be used as "non-print" coordinates. When both specific and general terms are used in indexing, the citation would probably be placed in *Index Medicus* under the more specific concepts. As

we gain more experience, our principles are being codified, but the need to decide whether a heading should be "print" or "non-print" will probably always add to the time required to index an article.

A justifiable criticism frequently levelled against a classic library subject heading list is that it is static. With MEDLARS it will be possible to update *Medical Subject Headings* more easily and effectively. The subject heading list will be constantly in the computer and the terms assigned each citation will be automatically checked against this list as it enters the system. At the same time the computer will be able to report regularly the number of times each heading is used. We can find not only to what citations a given heading was assigned, but the frequency with which other headings were assigned with it. This should make it easy to stay alert to the need for new headings.

The need for new concepts in *Medical Subject Headings* is frequently expressed within and outside the Library and each such recommendation is carefully considered. Each new authoritative classification or nomenclature list is also thoroughly reviewed for recently accepted terms or changes in approved usage.

And finally, it is possible to use provisional headings in the computer. An article may be indexed not only by existing terms in the list, but also by provisional headings which will not appear in *Index Medicus* or in *Medical Subject Headings* but may be used in searching. Periodically all the provisional terms in the system will be examined together with the citations that have been assigned to them. In this way, we should be able to select important new terms to be added to *Medical Subject Headings* as soon as they begin to be used.

The librarian who uses *Medical Subject Headings* as an authority will have to balance the disadvantages of having to keep up with our changes and additions against the advantages of having new and specific terms to use in searching *Index Medicus*. We hope he will find that the advantages are preponderant.