# Appendix E.— MEDLINE: Technical Processes

## Introduction

The processes involved in the creation and use of the data base MEDLINE were discussed in chapter *2.* Four aspects of these processes—literature selection, the development of medical subject headings (MeSH), indexing, and searching—are described here in more detail. The same search process is employed for all MEDLARS data bases.

## Literature Selection

The data tape used for the printing of Index Medicus is used as the source of data for MEDLINE. Thus, references to articles selected for Index *Medicus* are incorporated into MEDLINE. In selecting journals for *Index* Medicus/MEDLINE, the National Library of Medicine (NLM or the Library) receives the advice of seven to nine outside consultants who are acknowledged specialists in their fields. They are selected by the staff with the approval of the Director of NLM. Health educators, researchers, librarians and editors of medical and scientific journals have often been advisors. In 1981, in addition to representatives from these fields, the set of consultants included the vice-chairman of a health planning commission, who was formerly commissioner of a State health department. His appointment may indicate NLM's interest in health services research and delivery. Consultants are appointed on an ad hoc basis and serve no set term. They meet as a group three or four times a year, and the chief of the MeSH section/editor of *Index Medicus* acts as chairperson.

Journals considered for inclusion in *Index Medicus/ MEDLINE* are suggested to the editor by publishers, the selection and acquisition staff of NLM, and, occasionally, by users. Several issues of each journal title are reviewed and discussed at length by the consultants.

In the selection process, the consultants consider the scientific merit of the publications, and their relevance to NLM's objectives. Other than these general standards, and a scope and coverage manual for the selection of materials for the Library as a whole, there are no formal selection criteria. NLM believes that the diversity of material included in *Index Medicus/ MEDLINE* precludes the establishment of a set of criteria that would be relevant for all the categories included in the *Index.*

After each consultant rates each serial on a scale from 1 to 5, the chairperson averages the scores and ranks the serials according to their average ratings. With a score of 2.5, the serial is eligible for, but not assured, inclusion in the *Index.* In comparison to the number contained in NLM's collection, the number of serials indexed for *Index* Medicus/MEDLINE is very small. Only 2,664 of the more than 20,000 serials collected by the Library were indexed in 1980, an increase of slightly over 100 in the past 2 years. However, many of the serials are directories, annual reports, and newsletters. The number of articles indexed in 1980 was restricted to 273,750.

While management considerations (budget, number of positions, scope of the data base) have been important in delimiting the size of the universe of publications indexed for *Index* Medicus/MEDLINE, the question of quality control has dominated the restrictions on its size. The journal literature of the health sciences exhibits a high degree of repetition and a mixed degree of scientific excellence. The inclusion in or exclusion from *Index* Medicus/MEDLINE represents an effort to maintain a standard of scientific quality (3). With the purchase of the new IBM 370/168 computer, NLM expects to increase the number of articles indexed, and has requested that users, particularly foreign centers, update their equipment in the coming year to accommodate the change.

A similar mechanism is used to review journals that are being considered for deletion. In 1981, NLM added 145 new serials to *Index* Medicus/MEDLINE and deleted 142, 30 because they were no longer published. The consultants periodically review all the serials in an area, such as obstetrics-gynecology, and add or delete titles as appropriate.

As noted earlier, *Index* Medicus/MEDLINE has contained a small number of references to non journal literature such as monographs and proceedings of biomedical meetings. Because of the difficulty in selecting this type of material, the journal selection consultants recommended that this practice be discontinued; in 1981, it was. However, there is one type of document that is not covered in *Index* Medicus/MEDLINE and receives poor attention in NLM's entire collection, and that is the "fugitive" literature (see ch. 3).

## Medical Subject Headings (MeSH)

The terms that are most commonly used by the authors of English-language literature are used in the MeSH vocabulary. New terms for the MeSH vocabulary are suggested by the literature itself, MEDLINE users, professional associations, staff indexers, biomedical scientists, and special, NLM-concerned committees. With few exceptions, MeSH tends to follow the literature rather than to lead it. One exception occurred when the American Psychiatric Association re-

cently revised its classification of psychiatric disorders; MeSH was changed accordingly, even before the new classification had come into common usage. Similarly, NLM revised the MeSH terminology on the neurological aspects of speech disorders last year, with the assistance of experts in the field.

MeSH is arranged alphabetically and categorically. The 15 categories are further subdivided, and arranged in a hierarchical manner to show relationships between broader and narrower terms. The hierarchical structure and the indexing principles of specificity (i.e., indexing toward the most specific concept discussed) permits what is termed an "explode" capability during the search process. At the command of "explode," the system searches for all subcategories of a more general concept. For instance, if one were to search for the effect which a certain group of drugs, such as tranquilizing agents, had on animals, the "explode" feature would allow the searcher to specify only the phrase, "tranquilizing agents," and references indexed with the general term "tranquilizing agents," or specific agents would be retrieved.

Because the MeSH vocabulary is used for *Index Medicus,* for MEDLINE, and for cataloging, it has inherent problems; what is optimum for one may not be optimum for the other. For example, the *Index Medicus* user usually has a general interest in a subject and wants to search in fewer places than does the MEDLINE user who may have a specific topic of interest. Thus, the 14,000 MeSH descriptors are divided into 9,000 major and 5,000 minor descriptors. Whenever an indexer assigns a specific minor descriptor for the on-line searcher, the computer adds an appropriate predetermined, more general major descriptor under which the citation may appear in Index *Medicus.*

## Indexing

There are prescribed qualifications for indexers who assign headings from MeSH to articles. In 1981, NLM had 21 full-time and one part-time indexers/revisers on its staff. Almost all of them had undergraduate training in the biological sciences; some held masters and doctoral degrees. One or two of the foreign language experts have had the science requirement waived. As noted in chapter 2, NLM and some foreign centers contract out some of the indexing to U.S. commercial firms. The commercial contractors are required to hire individuals with biomedical backgrounds acquired either through formal education or comparable experience.

All domestic indexers must take an NLM-operated 2-week formal training course. After completing the course, each new indexer is assigned to a reviser, who continues the training in an individual one-on-one setting. This one-to-one relationship lasts anywhere from **2 to 6** months depending on the reviser's estimate of the new indexer's ability. The training of foreign indexers has varied over time. In the late **1960's** and early **1970's,** many foreign centers sent their staff to NLM for training. Later, NLM sent experienced indexers abroad to some of the centers, to train the respective centers' staff. For the most part, new foreign indexers now receive their training from indexers previously trained at or by NLM.

NLM continually provides training to indexers. In the United States and abroad, indexers are in weekly or monthly contact by telephone or mail with their revisers. There is, as well, a day-long training session for indexers in the United States each fall, when the new MeSH is published. NLM sends orientation packages to indexers (usually from foreign centers) who do not attend the session. In addition, seminars with the revisers are held periodically in order to increase consistency of indexing. Technical notes and memoranda are published, generally monthly, to increase indexing consistency.

NLM staff indexers have specific performance standards to fulfill. They are expected to index four articles of medium difficulty per hour and five articles of lesser difficulty per hour. Revisers, who are highly qualified indexers, are responsible for reviewing and revising 15 articles per hour that are indexed by new indexers, and scanning **25** articles per hour that are indexed by more experienced indexers. Unlike NLM staff indexers, contract indexers are not held to index a set number of articles per hour, but are paid for each article indexed.

NLM exercises a high degree of control over the quality of the indexing by *means* of a sequence of computerized validation routines, by proofreading (which is performed at a number of stages in the flow of material through the indexing section), by controlling the qualifications and the training of the indexers, and by the use of revisers. In addition, NLM provides the commercial contractors with NLM indexing tools, such as MeSH, and requires the firms to supply the indexers with dictionaries, textbooks, and other aids.

Despite their qualifications and the training they receive, indexers make errors: a few find their way into *Index Medicus* and MEDLINE. Library personnel ascribe this to the complexity of the indexing processes and policies and to what they perceive as the inconsistencies of MeSH. There are, in particular, subtle differences between some of the subheadings, which may cause the indexers to be inconsistent in assigning them to MeSH. For example, the indexer may perceive the difference in the subheadings of metabolism or physiology as being so slight that the choice between them

may be arbitrary in some instances. A computer program checks all indexing terms on the citation forms against MeSH, identifying spelling errors, ineligible subject headings/subheading combinations, and non-MeSH terms for correction.

In late 1981, an average of 69 days passed between the time a top priority journal was received at NLM and its entry into MEDLINE. Other journals took even longer. Since Index Medicus and MEDLINE are updated monthly, it would take the minimum of 30 days to process a journal. Concerned with the flow-through time, NLM is currently planning to partially computerize the indexing process.

## Searching and Retrieval

Searchers formulate a search on the basis of knowledge of indexing principles, such as specificity, the use of subheadings, MeSH, and Boolean logic, and transmits the search statement to the computer via a keyboard terminal. (Searches can also be formulated on-line and then stored in the computer for reference and later use. ) The computer searches the data bases and produces an individualized bibliography.

On-line search requests are generally broken down by trained information specialists into concepts that can be translated into MeSH terms and retrieved by querying the computer using one MeSH term or a combination of MeSH terms combined according to Boolean logic, a system that uses the connective "and," "or," and "not" to express relationships between concepts. For example, to obtain every article discussing either potassium or cyanide, a searcher would ask for "potassium cyanide." To obtain only discussions including both poisons, a searcher would ask for "potassium *and* cyanide. " To obtain articles dealing with potassium only when cyanide does not also appear, a searcher would ask for "potassium *and not* cyanide. "

For new or abstract concepts that cannot be found in MeSH, the searcher uses text word searching. A text word search usually results in fewer, more specific citations than a search using MeSH. It also allows for retrieval of very new concepts that MeSH may not have incorporated into its vocabulary, and allows for the occasional user who may not know medical terminology to use the system. Unlike MeSH, text word searching does not have the "explode" feature and requires searching for a concept under all its possible expressions.