

Mechanization of Library Procedures in the Medium-Sized Medical Library:

XIV. Correlations between National Library of Medicine Classification Numbers and *MeSH* Headings*

BY RUTH E. FENSKE†, *Public Health Service Trainee
in Computer Librarianship*

*Washington University School of Medicine Library
St. Louis, Missouri*

ABSTRACT

The purpose of this study was to determine the amount of correlation between National Library of Medicine classification numbers and *MeSH* headings in a body of cataloging which had already been done and then to find out which of two alternative methods of utilizing the correlation would be best.

There was a correlation of 44.5% between classification numbers and subject headings in the data base studied, cataloging data covering 8,137 books. The results indicate that a subject heading index showing classification numbers would be the preferred method of utilization, because it would be more accurate than the alternative considered, an arrangement by classification numbers which would be consulted to obtain subject headings.

IN a talk at the Washington University School of Medicine, Miss Emilie Wiggins, of the National Library of Medicine, remarked that she had started annotating *Medical Subject Headings (MeSH)* with National Library of Medicine classification numbers. This raised the question of how much correlation actually does exist between classification numbers and subject headings. Further communication with Miss Wiggins revealed that Eugene Muench of the New York University Medical Center has already constructed a series of seven correlation indexes for *Medical Subject Headings* in English and Spanish and five biomedical classification

schemes (1). One of the Muench indexes is in subject heading order and shows, among other things, which National Library of Medicine classification numbers can be used with each *MeSH* term. Another is in National Library of Medicine classification number order, showing *MeSH* terms that can be used with each classification number. Mr. Muench, as did Miss Wiggins, annotated the *MeSH* list term by term with classification numbers. Their annotations are undoubtedly based on many years of experience and probably on some perusal of the card catalog.

The project I shall describe had as its purpose to investigate the extent of correlation between National Library of Medicine classification numbers and *Medical Subject Headings* in a body of cataloging which had already been done. Furthermore, it was intended to find out if, in exploiting any correlation that might exist, it would be better first to establish the classification number for a work and then to try to find the subject heading(s), or if it would be better to establish the subject heading(s) and then look for a classification number.

THE STUDY

Choosing a data base of manageable proportions was a major problem. The data base had to be large because no correlation based on precedent could be established unless a subject heading or classification number had been used more than once. I selected as a data base the cataloging that had been done at the Washington University School of Medicine Library from 1965, when its automated catalog-

* This work was supported by funds from the U.S. Public Health Service Grant no. 5 T01 LM000106-04.

† Present address: Middleton Medical Library, University of Wisconsin, Madison, Wisconsin.

TABLE 1
DATA BASE BY CLASSIFICATION NUMBER

Classif. no.	No. of books	Classif. no.	No. of books	Classif. no.	No. of books
A	13	QS	133	WJ	87
B	144	QT	130	WK	133
C	9	QU	233	WL	388
D	7	QV	301	WM	613
E	4	QW	198	WN	172
F	2	QX	31	WO	228
G	23	QY	78	WP	149
H	256	QZ	246	WQ	79
J	7	S	57	WR	65
K	2	T	36	WS	231
L	47	U	42	WT	93
N	6	V	2	WU	321
P	42	W	342	WV	88
Q	98	WA	204	WW	162
QA	24	WB	193	WX	79
QC	30	WC	109	WY	22
QD	125	WD	89	WZ	277
QE	2	WE	229	Z	294
QH	294	WF	117	ZQ	25
QK	11	WG	217	ZW	94
QL	91	WH	113		
QP	44	WI	156		
				Total	8,137

ing system began, until July, 1970, when the project was undertaken. In that period 8,137 books were cataloged. In using this data base I had the very practical advantage of having readily available through computer listing two lists which could be used without tying up the shelf list or the subject catalog. The first list was sorted first into classification number order and secondarily alphabetically by subject; the second list, first in subject heading order and secondarily by classification number.

Using a local data base, even though it is based on National Library of Medicine cataloging, has one major disadvantage. The acquisitions policy of any library affects the number of works assigned to any area of the classification scheme and the number of times any subject heading is used. For instance, Washington University has few books classified in WY, and the subdivision NURSING is rarely used because Washington University does not have a nursing school and, therefore, does not acquire nursing books. Cataloging in this area may be particularly high in correlation. This all affects the results. Table 1 shows how many books in

each major area of the classification scheme there are in the data base.

The data are analyzed on the basis of classification number-subject heading (CN-SH) pairs. These pairs are always one-to-one. Therefore, a book having classification Z 699 and subject headings INFORMATION RETRIEVAL SYSTEMS, AUTOMATIC DATA PROCESSING, and LIBRARIES would involve three CN-SH pairs. A distinction is made between classification number groups and total classification numbers, and between subject heading groups and total subject headings. *Classification number groups* refers to distinct, nonrepeated classification numbers. In the list sorted by classification all WO 100s are together. This is the WO 100 classification number group. *Total classification numbers* refers to the total number of classification numbers assigned, whether or not some classifications are repeated. There are 2,033 classification number groups and a total of 8,137 classification numbers. That means each classification number occurs on the average four times. Similarly, *subject heading groups* refers to distinct, nonrepeated subject headings. In the list sorted by subject heading, all the occurrences of RETINAL DETACHMENT are together. This is the RETINAL DETACHMENT subject heading group. *Total subject headings* means the total number of subject headings assigned, counting one subject head as many times as it is used. There are 5,251 subject heading groups and 13,527 total subject headings. Each subject thus occurs 2.6 times on the average. From these figures we can also tell that each book was assigned an average of 1.7 subject headings. Table 2 summarizes the basic characteristics of the data base.

Correlation is considered in terms of classification number-subject heading (CN-SH) pairs. To find these pairs the following steps were taken. (1) Each classification number group was considered separately. For example, WO 100 was considered separately from WO

TABLE 2
BASIC CHARACTERISTICS OF THE DATA BASE

A. Total books cataloged.....	8,137
B. Total classification numbers.....	8,137
Classification number groups.....	2,033
C. Total subject headings.....	13,527
Subject heading groups.....	5,251

NLM CLASSIFICATION NUMBERS AND SUBJECT HEADINGS

TABLE 3
EXAMPLE OF SORTING BY CLASSIFICATION NUMBER SUBJECT HEADINGS

Classification Numbers	Subject Headings
WL 600 B348E	AUTONOMIC NERVOUS SYSTEM BRAIN STEM MEDULLA OBLONGATA NEUROPHYSIOLOGY
WL 600 B963A	AUTONOMIC NERVOUS SYSTEM NEUROCHEMISTRY
WL 600 G795M	MOTOR NEURONS MUSCLES—CONGRESSES
WL 600 G984E	MUSCLES—CONGRESSES NERVOUS SYSTEM—CONGRESSES
WL 600 T828C	AUTONOMIC NERVOUS SYSTEM NEUROPHYSIOLOGY

101. (2) Then, for each classification number group, the number of CN-SH pairs was counted. As an example, five works cataloged under WO 100 might yield a total of eight CN-SH pairs. (3) After that, the number of CN-SH pairs that agreed with any other CN-SH pair was counted. For example, under WL 600 it was found that five books were assigned this classification number and that a total of twelve subject headings were assigned these five books (Table 3). This makes a total of twelve CN-SH pairs for the WL 600 classification number group. AUTONOMIC NERVOUS SYSTEM occurs three times, NEUROPHYSIOLOGY two times, and MUSCLES—CONGRESSES two times. This makes a total of seven CN-SH pairs that agree with at least one other CN-SH pair; i.e., there are seven correlations. With this method correlations did not need to be limited to an ideal combination of one subject heading and one classification number. If none of the CN-SH pairs for a specific classification number agreed with another CN-SH pair for that classification number, the figure 0 was recorded to indicate that there was no correlation.

After this exercise for classification numbers, the same method was used for the subject heading list. An example of the method is that six books were assigned the heading BREAST NEOPLASMS as at least one of the subject headings (Table 4). Four of these books were classified WP 870, and none of the other CN-SH pairs were repeated. Therefore, the number

TABLE 4
EXAMPLE OF SORTING BY SUBJECT HEADING

BREAST NEOPLASMS	QZ 200 N277A34
BREAST NEOPLASMS	QZ 268 J 95E
BREAST NEOPLASMS	WP 870 I34C WP 870 M821C WP 870 S767C WP 870 S989C

four was recorded. The numbers were recorded subject heading by subject heading, considering a subject heading as being different from the same subject heading with a subheading; e.g., RETINAL DETACHMENT and RETINAL DETACHMENT—SURGERY were considered two distinct subject headings.

After obtaining these figures group by group, the data as a whole were examined. The number of classification number groups which contained any repeated CN-SH pairs was counted. This is the same as counting how many classification number groups show correlations by subject (Table 5-A). Then, looking at the data from the other direction, the number of subject heading groups containing repeated CN-SH pairs was counted. This gave the number of subject heading groups showing correlation by classification number (Table 6-A).

The classification and subject heading groups showing no correlation were examined to find

RUTH E. FENSKE

TABLE 5
ANALYSIS OF 2,033 CLASSIFICATION NUMBER GROUPS

A. Show correlation by subject	37.5%	763	
B. Show no correlation by subject	62.5%	1,270	
1. One book only with this classification number		47.2%	959
2. More than one book with this classification number		15.3%	311
	100.0%	2,033	

TABLE 6
ANALYSIS OF 5,251 SUBJECT HEADING GROUPS

A. Show correlation by classification number	24.3%	1,275	
B. Show no correlation by classification number	75.7%	3,976	
1. One book only with this subject heading		61.8%	3,243
2. More than one book with this subject heading		13.9%	733
	100.0%	5,251	

TABLE 7
ANALYSIS OF 1,270 CLASSIFICATION NUMBER GROUPS SHOWING NO CORRELATION BY SUBJECT

A. One book only with this classification number	75.5%	959
B. More than one book with this classification number	24.5%	311
	100.0%	1,270

TABLE 8
ANALYSIS OF 3,976 SUBJECT HEADING GROUPS SHOWING NO CORRELATION BY CLASSIFICATION

A. One book only with this subject heading	81.6%	3,243
B. More than one book with this subject heading	18.4%	733
	100.0%	3,976

out how many of the groups showed no correlation because they represented the cataloging of only one book and therefore contained no other CN-SH pairs to compare (Table 5-B-1, 6-B-1, 7-A, 8-A).

The final figures focus on the classification number-subject heading pairs. Table 9-A shows

the total number of CN-SH pairs which agree with at least one other CN-SH pair; Table 9-B and 9-C show the total number of CN-SH pairs in classification and subject heading groups which show some correlation, and Table 10 shows how many of these CN-SH pairs actually agreed with another CN-SH pair in that group.

TABLE 9
ANALYSIS OF 13,527 CLASSIFICATION NUMBER-SUBJECT HEADING PAIRS

A. Correlated	44.5%	6,013
B. In classification number groups which show correlation	77.9%	10,543
C. In subject heading groups which show correlation	62.5%	8,449

TABLE 10
ACTUAL CORRELATION OF CN-SH PAIRS IN GROUPS WHICH SHOW CORRELATION

A. In classification number groups	57.0%	6,013 (of 10,543)
B. In subject heading groups	71.2%	6,013 (of 8,449)

NLM CLASSIFICATION NUMBERS AND SUBJECT HEADINGS

DISCUSSION

These figures reveal that 37.5% (Table 5-A) of the classification number groups that exist at the Washington University School of Medicine Library could be listed in an index which would lead the cataloger to a list of suggested subject headings (only those subject headings which had been correlated under that classification number would be in the index). But even if the cataloger found a classification number in the index, only 57% (Table 10-A) of the time would the subject heading which had actually been assigned be in the index. This is because only 57% of the CN-SH pairs in classification number groups which show correlation actually are correlated. A further complication is that one book may involve up to three or four CN-SH pairs. For a particular book the cataloger may find that the classification number index lists either all the subject headings he needs, part of the subject headings he needs, or none of the subject headings he needs. It would involve a great deal of judgment on the part of the cataloger to choose correctly one of the alternatives, especially when in 43% of the cases the subject heading he needs will not be there. Looking at the shelflist would be similar to this approach, but the shelflist would show every subject heading that had ever been used with this classification number, regardless of whether the combination had been repeated or not.

Approaching the problem from the other direction, it is seen that 24.3% (Table 6-A) of all subject heading groups could be listed in an index (probably an annotated *MeSH*) which would lead the cataloger to a list of suggested classification numbers. If a given subject heading is in the index, the classification number to go with it would be listed in the index 71.2% of the time (Table 10-B). Using this index would not involve as much judgment on the part of the cataloger as would using the previously mentioned index, since he would need only to select one classification number from the list presented rather than up to three or four subject headings from a list of subject headings. Furthermore, if there is more than one subject heading, the cataloger can look under both terms in the index to find what classification number is common to them. Using this index would be like using a subject catalog to see what classi-

fication numbers had been assigned books with a certain subject heading. This usually involves going to the public catalog or leafing through several volumes of *Current Catalog*.

CONCLUSIONS

There is a correlation of 44.5% (Table 9-A) between classification numbers and subject headings; i.e., 44.5% of the classification number-subject heading pairs agree with some other CN-SH pair within a specific classification number or subject heading group. And the results obtained definitely indicate that in exploiting this correlation, going from subject heading to classification number would be better than going from classification number to subject heading, because going from subject heading to classification number would be more accurate. Although a subject heading list would cover only 24.3% (Table 6-A) of the subject headings and 62.5% (Table 9-C) of the CN-SH pairs, it would be more accurate than a classification number list for what is in it: 71.2% for the subject heading list to show classification numbers versus 57% for the classification number list to suggest subject headings (Table 10). Another advantage of using a subject heading list rather than a classification number list is that, as explained above, using the subject heading-to-classification number list would not involve as much judgment on the part of the user as would using a classification number-to-subject heading list.

Annotating a subject list with classification numbers is not a new idea. The Library of Congress subject headings are selectively annotated with classification numbers. An annotated subject heading list would augment the tools readily available to the medical cataloger. The shelflist can serve approximately the same function as an annotated classification number list, but unfortunately an annotated *MeSH* is not one of the present tools of the medical cataloger.

It may be argued that the only classification numbers that would be in the projected list would be the obvious ones. This could well be true, but a person cannot remember too many classification numbers. Other catalogers may argue that they do not have trouble assigning classification, that it is the rarely used, but perfect, subject heading which is difficult to find. The method described here probably

would not help these catalogers to the extent that the *MeSH* categorized lists do or that the *MeSH* tree structure would if it were available to them. The very specific subject heading that is not used very often in book cataloging could be hand-edited into the list if the broader term showed correlation under some classification number.

Having seen the results and having considered some of the pros and cons of the projected list, it would seem that annotating *MeSH* selectively with one or several classification numbers might prove to be a practical aid to cataloging, especially if it could be shown that adding more cases would increase the percentage of correlation to more than the 44.5% this study has established. One can observe that only 13.9% (Table 6-B-2) of the subject heading groups showed no correlation when more than one book had been assigned that subject heading; most of these subject heading groups will probably never show a correlation. But one can also observe that 61.8% (Table 6-B-1) of all subject heading groups in the data base had only been used once; many of these would

probably show a correlation if more cases were added. Considering this, it would seem that further study is justified.

SUGGESTIONS FOR FURTHER STUDY

Further extensions of this work might include a more careful scrutiny of the present data. Is the correlation particularly strong in a given area of the scheme, such as in the WMs, or in a given category of the *MeSH* categorized list? How do subheadings affect correlation? One could also examine a larger data base to see how a greater number of cases affects the total number of CN-SH pairs in correlation. Hand-editing could be done on the present data base along the lines suggested above, using the categorized lists and the tree structure to establish the broader terms.

REFERENCES

1. MUENCH, EUGENE V. A computerized English/Spanish correlation index to five biomedical library classification schemes based on *MeSH*. *Bull. Med. Libr. Assoc.* 59: 404-419, July, 1971.