Upsala J Med Sci 109: 141–158, 2004

Using the MEDLINE® database to study the concept of urinary tract infections in different domains of medicine

Daniel Karlsson,[1] Olle Aspevall,[2] Hans Åhlfeldt,[1] Urban Forsum[3]

Department of Biomedical Engineering, Division of Medical Informatics, Linköpings universitet, Sweden,[1] Department of Immunology, Microbiology, Pathology and Infectious Diseases, Karolinska Institutet, Stockholm, Sweden,[2] Department of Biomedical Engineering, Division of Medical Informatics, Linköpings universitet, Sweden,[1] Department of Health and Environment, Division of Clinical Microbiology, Linköpings universitet, Sweden[3]

ABSTRACT

A free full-text copy of this article can be found at the web page of Upsala J Med Sci: http://www.ujms.se

As a way of exploring differences between medical domains regarding management of urinary tract infections, we investigated the MEDLINE® database for differences in indexing patterns. Further, our intention was to assess the MEDLINE® database as a source for studying medical domains. We examined the use of main headings, subheadings and the level of main headings in six medical domains that manage urinary tract infections. Many intuitive but also some counterintuitive results were found indicating that the MEDLINE® database is difficult to use for studying medical domains mainly due to unclear semantics both in the headings and the indexing process, which results in variability in indexing. This variability probably hides significant results. We also conclude that the differences found indicate that in addition to differences between domains, there are also large variations within domains.

1. INTRODUCTION

Urinary Tract Infection (UTI) is a typical example of a medical concept used in many different medical domains. Management of UTIs can include knowledge from several relevant domains such as infectious diseases, microbiology, urology, nephrology, internal medicine, and general practice. In order to investigate and demonstrate differences between the medical domains our research group has conducted a range of studies using different methods and source materials [1, 2]. This paper describes a part of this work specifically exploring semantic differences between domains be studying the MEDLINE® database.

Others have demonstrated differences both among as well as within the areas of knowledge constituting the field of UTI using other materials and methods. Olesen

Received 9 December 2003
Accepted 2 February 2004

and Östergaard studied the management of UTI patient cases in three different medical specialties with regard to the possibilities of providing CME [3]. Their study showed large variations in suggested strategies for diagnosis, treatment, and follow up both within and between specialties. Timpka and Bjurulf studied the agreement among physicians from three different specialties regarding management of female genitourinary infections and reported similar results [4]. However, while these studies demonstrated differences between domains, the perspective was limited to certain, selected aspects such as diagnosis, treatment and follow up in the case of Olesen and Östergaard and agreement on diagnosis in the case of Timpka and Bjurulf.

The aim of this study was to examine the assignment of MeSH® main headings and subheadings to articles concerning UTI that are published in serials[1] belonging to different medical domains and also to discuss the use of the MEDLINE® database as a source of knowledge for studying medical knowledge domains. Discovering differences in indexing patterns, such as different depths of knowledge, i.e. granularity, or different sets of subheadings reflecting aspects of interest, might be of help in the theorization of the area of UTI.

In this article, we will present the results on an aggregated level.

As mentioned above, our research group has performed other studies in this field. These studies include the construction of a categorical structure of the field of UTI [1] and a literature study discovering fundamental differences in the view of the concept of UTI and other related concepts such as bacteriuria and urethral syndrome [2]. This paper describes an attempt at using the knowledge implicit in the MEDLINE® headings and indexing patterns complementary to the other studies, aiding the construction of categorical structure of the field of UTI.

2. MATERIAL

The material consists of a subset of references extracted from the MEDLINE® bibliographic database concerning UTI. In order to retrieve this subset, a MEDLINE® query was formulated. The query consisted of the MeSH® main heading 'Urinary Tract Infections' in conjunction with other relevant main headings. The query was discussed with and revised in accordance with the views of the MEDLINE® indexers at the library at Karolinska institutet, the Swedish MEDLARS® center. The resulting query was the union of the sub-queries listed in Table 1. The resulting set consisted of 66,987 references from 1963 to 2000.

The references were classified into domains by the serial in which they were published. Of the 66,987 references, 2,027 (3.0%) had no associated serial, i.e. no ISSN (International Standard Serial Number) field, and could therefore not be classified, leaving 64,960 references for consideration. The frequency of missing ISSN fields was higher in the 1960s and early 1970s, with an average of 102.3 per year from

[1] 'Serials' denotes periodical publications in the MEDLINE® database, e.g. scientific journals or monograph series.

Table 1.

Search expression	References[6]	Date
explode Urinary Tract Infections	25,063	2000-03-23
explode Cystitis	4,328	2000-03-31
explode Anti-Infective Agents, Urinary	28,059	2000-03-31
explode Pyelitis	10,183	2000-03-31
explode Microbiology AND explode Urine	72	2000-03-31
Urinalysis	706	2000-03-31
Urethritis	2,798	2000-03-31
Total	66,987	

[6] Dupllicates and empty records excluded.

1963 to 1975, than today, with an average of 12.6 per year 1985 up until present ($p < 0.001$, χ^2 test).

The division of serials into domains was based on the 1999 list of currently indexed journals arranged by subject [5]. The subjects used in the list are a selection of 127 MeSH® main headings picked by the NLM™ to represent a classification of the areas of biomedicine. The list included 3,243 distinct serials. Of the 127 subjects, 18 were used to form the six domains.[2] This aggregation into six domains was performed by the authors and was considered uncontroversial, as it follows the internationally acknowledged fields of medical subspecialization. The domains were constructed in order to reflect distinct areas of medicine that manage UTI patients. The domains were Microbiology and Infectious Diseases (MI), Internal Medicine (IM), Pediatrics (Pe), Gynecology and Obstetrics (GO), Urology and Nephrology (UN), and General Practice[3] (GP). The extents of the domains MI, GO, and UN are motivated by frequent cross publishing between their sub-domains. For example, many articles[4] on microbiology are published in infectious diseases serials. The sixth domain, General Practice (GP), was constructed from the 'Family Practice' subject with the addition of a number of handpicked serials including, for example, the BMJ and other national medical serials assumed to have impact on the GP community. The selection process was discussed with opinion leading GPs.

Of the 3,243 serials, 715 (22.0%) belonged to one or two of the five domains. Of

---

[2] The term 'Domain' is used to denote the quantifications of the areas of medicine used specifically in this study.

[3] The term 'Family Practice' is used in the NLM™ list of subjects. The terms 'Family Practice' and 'General Practice' are not distinguished in this study.

[4] 'Article' is the term used by the NLM™ to denote the entity to which a reference refers. This also includes comments, editorials, guidelines, letters, monographs, reviews, etc.

Table 2.

Domain	Serials	References	References/Serial
Microbiology and Infectious Diseases (MI)	176	10,438	59.3
Internal Medicine (IM)	277	11,093	40.1
General Practice (GP)	20	1,941	97.0
Pediatrics (Pe)	102	2,028	19.88
Gynecology and Obstetrics (GO)	81	1,124	13.88
Urology and Nephrology (UN)	55	7,199	130.9

the 64,960 references, 31,797 (48.9%) came from those domains. Figures describing the distribution of serials and references in the domains can be seen in Table 2.

Twenty-two of the 715 serials (3.08%) were listed under two domains. Examples of these serials are 'Pediatric Nephrology' and 'Infectious Diseases in Obstetrics and Gynecology.' These serials were excluded when comparing the domains in question, as they were considered atypical for their domains. None of the serials was listed under more than two domains.

In the construction of the material, three selections were made:

1. A selection of all serials in biomedicine that publish articles concerning UTI were indexed in the MEDLINE® database. (Unknown number).
2. A selection of the references in the MEDLINE® database concerning UTI were found using our query. (66,987 publication items).
3. A selection of the references found using our query came from serials included in the domains. (31,797 publication items).

These selections may give rise to false negative references as well as false positive references.

First, we will discuss the false negatives. In selection no. 1, the selection of serials to be indexed in MEDLINE®, the NLM™ Literature Selection Technical Review Committee reviews and assesses the quality of the contents of the serials in question [6]. On the grounds of this assessment, a decision is made as to whether or not the serial should be indexed. The current selection of 3,243 indexed serials represents about 10% of the estimated 30,000 to 40,000 published biomedical serials [7]. Although the NLM™ serial selection process must be subjective to some extent, the resulting selection must be considered systematic. Thus, our results can only apply to serials indexed in the MEDLINE® database. However, we are of the opinion that the coverage of the MEDLINE® database is of such magnitude and substance that the contents of the covered serials sufficiently span those areas of medicine concerned by this study.

In selection no. 2 articles are selected with an unknown level of recall, which is

certainly less than 100%. In other words, not all articles concerning UTI will be found by searching the database with the query. False negatives arise mainly for two reasons. First, our query formulation may not fully span the area of UTI, and second, indexers may not always catch every important aspect of each indexed article. Assuming that our query is a reasonable approximation of the area of UTI, the only random error comes from variability and error in indexing. Thus, we stipulate that the area of UTI in our study equals the *expected* results of our query.

In selection no. 3, references from serials not belonging to any of the domains are excluded. Here, apart from the true negatives due to the fact that a serial may not concern the area corresponding to the domain, false negatives may arise for various reasons. The extensions of the domains in this study are stipulated by the NLM™ serial subject classification [5]. However, it is most probable that there are items published in excluded serials describing UTI in relation to some of the areas corresponding to the domains. In addition, the various areas of medicine are constantly evolving and often debated, so no consensus definitions of the domains are available, nor will they ever be.

False positives have four main sources. First, in selection no. 2, our query formulation may be too wide. Second, the classification of serials may introduce false positives in the sense that the serials may have been erroneously classified. Third, indexers may erroneously apply index headings to references that the corresponding articles do not cover. Fourth, serials may publish articles less related to the subject applied by the NLM™ and, hence, to the domain.

We consider the first two sources of false positives to be systematic. The second case, where serials were erroneously classified, was ruled out by inspection: the problem is not inaccurate classification but rather the fuzziness of the domains.

The third source of false positives introduces a systematic error in relation to the dimensions as discussed below (see paragraph 3.1, Dimensions). The fourth is considered a random error.

For two reasons the resulting set of references is also skewed towards recent years. First, the number of missing ISSN fields is lower today than in earlier years, and second, since the domain classification is based on the serials indexed today (1999), discontinued serials are excluded. The distribution of number of references per year can be seen in Figure 1.

Examining only major main headings could increase consistency [8] and should increase the precision as well as decrease the sensitivity. Since headings representing central concepts should still be more frequent than less important headings in our material, and in order not to decrease sensitivity, we chose to examine all main headings, not just the major ones.

The results of our query were seen as a sample of the true set of references in the MEDLINE® database concerning UTI as defined by the expected results of our query. This true set of references is the population in this study.

For all samples the unique ID, year of publication, and ISSN were collected together with all MeSH® main headings and subheadings.

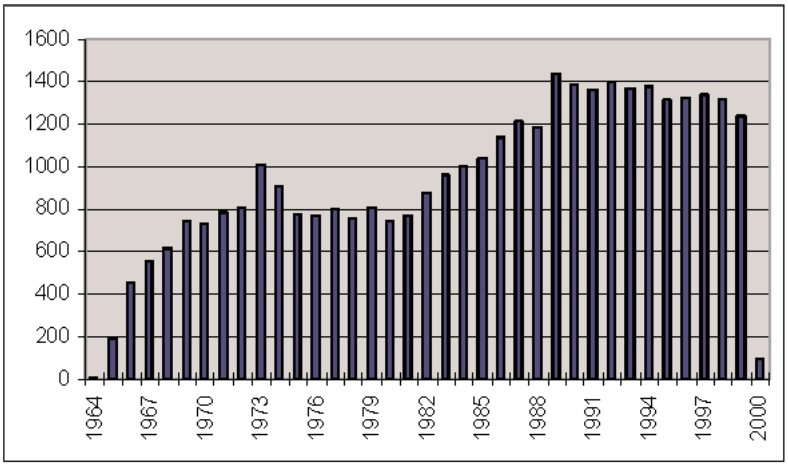

Fig. 1. References per year in resulting set.

3. METHOD

The references, sampled as described in chapter 2, were collected using NLMs™ free PubMed search system and imported into a Microsoft® Access database for re-organization of the material. The material was statistically analyzed using in-house developed software.

3.1 *Dimensions*

One of the aims of the study was to investigate the use of MeSH® main headings. To make use of the hierarchical tree structure of MeSH®, as defined by MeSH® tree numbers, entire sub-trees, here referred to as *dimensions*, were studied. The dimensions used in this study were constructed in order to reflect core aspects of UTI and were the result of a qualitative study of the knowledge domain of UTI performed by our research group [1]. The aim in using dimensions was also to reduce the effects of indexing variability. [1]The dimensions are listed in Table 3. Indention represents narrowness according to the MeSH® tree structure. For example, 'Antibiotics' is a narrower dimension than 'Chemicals and Drugs Category.'The fact that a MeSH® main heading can have more than one position in the tree structure did not influence to study. Since the dimensions are constructed from material relating to the area of UTI [1], the dimensions can be expected to appear more frequently in articles that belong to the study population and, hence, less frequently in articles not belonging to the population. Thus, the presence of false positives in the material, that is references not relating to the area of UTI, indicates that the dimension frequency results will be under-estimated. Metaphorically speaking, the material is diluted. However, the rate of false positives is assumed not to differ among domains.

Table 3. Dimensions.

Dimension	# MeSH® Main headings
Anatomy	1314
Bacteria	438
Infection	120
Pathological Conditions, Signs and Symptoms	494
Pathologic Processes	199
Disease Attributes	15
Signs and Symptoms	242
Chemicals and Drugs Category	6607
Antibiotics	24
Diagnostic Techniques and Procedures	354
Laboratory Techniques and Procedures	128
Specimen Handling	16
Therapeutics	321
Physiological Processes	102
Age Groups	28
Health Personnel	58

3.2 *Research questions*

The hypothesis is that conceptual differences exist between domains of medicine and that these differences can be seen and described as patterns of MEDLINE® indexing. To study the differences among domains, we statistically analyzed a selection of MEDLINE® database entries using three quantitative measures of domain differences: (1) use of headings, (2) granularity of headings used, and (3) use of subheadings. A 1-percent simultaneous degree of confidence was used in all analyses. For each quantitative measure and for each dimension, all pairs of domains were compared independently. To avoid dependencies between domains, articles published in serials belonging to more than one domain were excluded when those domains were compared, but not otherwise.

3.2.1 *Use of headings*

To estimate the importance of specific dimensions in the domains, the number of headings in a certain dimension was measured per article.

The number of headings in a dimension per article is dependent on the available number of headings in the MeSH® tree structure. As the number of headings per dimension varies from 15 for the dimension ‘Disease Attributes’ to 6607 for the dimension ‘Chemicals and Drugs Category,’ i.e. a factor of 440, dimensions should not be compared with each other.

To test the difference between domains, data was first trichotomized, i.e. the articles were categorized into three disjunct categories: those with no heading, those with one heading and those with two or more headings. Trichotomies were chosen in favor of dichotomies in order to consider cases with more than one heading per article. More than three categories greatly increased the number of cases where χ^2 test assumptions were violated, i.e. when the expected frequency was less than five for more than one fifth of the categories. The count of articles in these categories was used for testing. χ^2 values were computed for all unique unordered pairs of domains (15 pairs), except when χ^2 test assumptions were violated. No assumption regarding normality of data distribution was made.

3.2.2 *Use of subheadings*

The use of subheadings in combination with the main headings 'Urinary Tract Infections,' 'Cystitis,' and 'Pyelitis,' plus narrower headings was studied analogous to the case of MeSH® main headings (3.2.1). However, in contrast to dimensions, a subheading can only appear once per article whereas there can be several main headings in one dimension appearing in one article. Thus, dichotomies were used instead of trichotomies.

3.2.3 *Granularity of headings*

To estimate the granularity of knowledge among domains, the average heading level, i.e. the length of the path to the root in the MeSH® tree structure, was measured. In the tree structure, the heading level is generally not a good estimate of granularity since the "semantic distance" corresponding to one step down in the tree structure varies. For example, the heading pairs 'Arm' and 'Fingers', and 'Equipment and Supplies' and 'Air Bags'[5] have the same distance in the tree structure. However, if the heading level is studied in the context of just one dimension, level can be a

Table 4. Descriptive statistics for the number of headings per article for the dimension 'Bacteria'.

Percentiles	# Articles	# Headings	50%	75%	95%	Max	Mean	SD	Skew
MI	10138	9297	1	1	3	15	0.917	1.1526	2.7842
IM	9375	3043	0	0	4	12	0.325	0.8346	4.7283
Pe	1743	320	0	0	1	6	0.1836	0.5397	4.4115
GO	1092	279	0	0	2	7	0.255	0.7541	4.2796
UN	7072	1371	0	0	2	10	0.1939	0.6460	5.5961
GP	351	41	0	0	3	3	0.11681	1.1126	6.5876

[5] I.e. the "automotive safety devices."

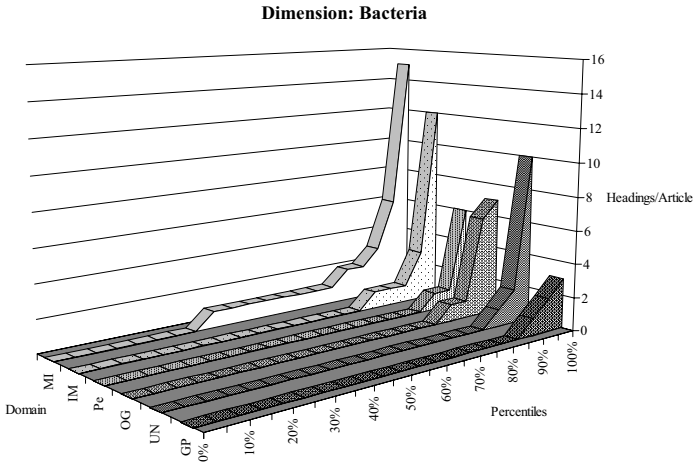

Fig 2. Percentiles plotted. Articles against number of MeSH® main headings for the dimension 'Bacteria.'

good estimate of granularity as the consistency is larger within our dimensions than within MeSH® as a whole. This also implies that dimensions should not be compared with each other.

When a heading in a certain dimension had two positions in the tree structure, the average level was used. For example, the heading 'Jaw' in the dimension 'Anatomy' is categorized both as a facial bone (level six) and as belonging to the stomatognathic system (level two). Consequently, the heading 'Jaw' is given the level four. The alternative aggregation-operators maximum and minimum are assumed to overestimate and underestimate the level, respectively.

To compare the actual use of main headings in the MEDLINE® database with the distribution of headings in the MeSH® tree structure, the MeSH® sub-tree corresponding to the dimension was included as a seventh "domain". Thus, the average level, corresponding to the granularity, of the main headings for a dimension was compared to actual use.

To test differences between domains, two-tailed confidence intervals for differences between means were constructed using Z-distributions for each pair of

Table 5. Number of articles with zero, one, and two or more headings, 'Bacteria' dimension, 'Microbiology and Infectious Diseases' and 'Internal Medicine' compared.

	0	1	2-	Total
MI	4119	4597	1650	10366
IM	8665	1789	567	11021
$\chi^2 = 3363.41, p \approx 0$				

domains (21 pairs including MeSH® comparison). No assumption regarding normality of data distribution was made.

4. RESULTS

Selected examples of significant results are presented here. The general principle used to select examples for presentation in this paper is that one or two domains should differ significantly from all but, at most, one domain. Some additional examples have been provided where this was considered appropriate. In addition, estimates of the magnitude of the differences are provided.

An average of 58% of the main headings in any of the dimensions were used in any of the articles published in a serial belonging to a domain. For all articles, 69% of the headings were used.

4.1 *Use of headings*

Of the 16 dimensions, all except for the 'Health Personnel' dimension allowed analysis. The distribution of number of main headings per article had a high positive skewness, as is exemplified in Table 4 and Figure 2. Further, many of the articles were not indexed under any heading relating to the dimension at hand. Outliers were frequently present, as indicated by the difference between the 95th percentile and the maximum value in Table 4. All dimensions showed a similar shape in their distribution. An example of a trichotomization is shown in Table 5. Here the use of headings in the dimension 'Bacteria' in the domains of 'Microbiology and Infectious Diseases' and 'Internal Medicine' is compared. In this example, 4,119 of the 10,366 articles belonging to the MI domain had no 'Bacteria' main heading, 4,597 had one heading, and 1,650 had two or more 'Bacteria' headings. The results concerning the use of headings are presented in Table 6.

4.2 *Use of subheadings*

Thirty-eight different subheadings were found, 18 of which were frequent enough to allow analysis and show some significant differences. There were 19,410 articles which were indexed under the main headings 'Urinary Tract Infections', 'Cystitis,' and 'Pyelitis' that belonged to one of the domains. Of these, 19,050 (98.1%) had at least one subheading. The results are presented in Table 7.

4.3 *Granularity of headings*

Of the 16 dimensions, all except the 'Health Personnel' dimension allowed analysis. One of the dimensions showed no significant differences, and another dimension showed only scattered significant differences, i.e. no patterns. For each dimension, both the results for the domains in relation to the MeSH® tree structure and the results of inter-domain comparisons are presented. These results are presented in the upper and lower right hand side of each row respectively. The results are presented in Table 8.

Table 6. Results for use of main headings.

Dimension	Significant ($\chi 2$, $p < 0.01$) differences
Anatomy	OG and UN more headings than all others.
Bacteria	MI more headings than all others (median of MI = 1, median of others = 0).
Infection	MI fewer headings than all others (median of MI = 0, median of others = 1), OG more headings than all others (75-percentile of OG = 2, 75-percentile of others = 1).
Pathological Conditions, Signs and Symptoms	MI fewer headings than all others, OG and UN more headings than all others (75-percentile of MI = 0, 75-percentile of others = 1).
Pathologic Processes	MI fewer headings than all others (85-percentile of MI = 0, 85-percentile of others = 1), OG and UN more headings than all others.
Disease Attributes	MI fewer headings than all others (90-percentile of MI = 0, 90-percentile of others = 1).
Signs and Symptoms	MI fewer headings than all others. (95-percentile of MI = 0, 95-percentile of others $\geq$ 1)
Chemicals and Drugs Category	MI more headings than all others (median of MI = 3, median of others = 3), IM more headings than Pe, OG, UN and GP (75-percentile of IM = 3, 75-percentile of Pe, OG, UN and GP = 2)
Antibiotics	MI more headings than all others (80-percentile of MI = 1, 80-percentile of others = 0), UN fewer headings than all except for OG (90-percentile of UN = 0, 90-percentile of others = 1).
Diagnostic Techniques and Procedures	MI fewer headings than all others (90-percentile of MI = 0, 90-percentile of others = 1).
Laboratory Techniques and Procedures	MI fewer headings than IM and UN.
Specimen Handling	MI fewer headings than IM.
Therapeutics	MI fewer headings than all except for Pe (90-percentile of MI = 0, 90-percentile of others = 1).
Physiological Processes	OG more headings than all except for Pe (95-percentile of OG = 2, 95-percentile of others = 0).
Age Groups	MI fewer headings than others (median of MI = 0, median of others $\geq$ 1), Pe more headings than others (median of Pe = 2, median of others $\leq$ 1)
Health Personnel	-

5. DISCUSSION

An important question here is whether MEDLINE® is useful as a source for analyzing medical knowledge domains. This issue can be subdivided into two questions, one regarding the quality of the MEDLINE® database and the other the adequacy of the MeSH® thesaurus.

Table 7. results for use of subheadings.

Subheading	Significant ($\chi 2$, $p < 0.01$) differences
Blood	MI fewer subheadings than Pe and UN (MI 0.6%, Pe, UN > 1.5%). Pe and UN more subheadings than GP (GP 0.2%, Pe 1.63%, UN 1.56%).
Complications	MI fewer subheadings than others (MI 2.9%, others 8.9% – 20%), Pe, OG, and UN more subheadings than others (Pe 19.0%, OG 16.7%, UN 20%, others 2.9% –11.2%).
Diagnosis	MI fewer subheadings than others (MI 4.3%, others 13.8% – 24%), Pe and GP more subheadings than all but OG (Pe 24%, GP 23%, others 4.3% – 18.0%)
Drug therapy	IM and GP more subheadings than others (IM 22%, GP 24%, others 14.6% – 16.4%)
Epidemiology	MI fewer subheadings than all but UN (MI 3.5%, others 5.2% – 9.8%).
Etiology	MI fewer subheadings than others (MI 3.9%, others 11.70% – 25%), OG and UN more subheadings than others (OG 23%, UN 25%, others 3.9% – 15.6%).
Immunology	UN more subheadings than others (UN 4.4%, others 0.89% – 2.5%)
Metabolism	UN more subheadings than all but Pe and OG (UN 1.88%, others except for PG and OG 0.21% – 0.74%).
Microbiology	MI more subheadings than others (MI 17.7%, others 9.5% – 12.8%).
Pathology	UN more subheadings than others (UN 9.1%, others 0.99% – 3.0%).
Physiopathology	MI fewer subheadings than all but GP (MI 0.42%, 0.48%, others but GP 1.62% – 4.6%).
Prevention & control	MI fewer subheadings than others (MI 2.4%, others 4.2% – 10.1%).
Radiography	MI fewer subheadings than all but OG (MI 0.15%, others except for OG 1.97% – 4.8%).
Radionuclide imaging	Pe more subheadings than others (Pe 3.7%, others 0.067% – 0.92%).
Surgery	UN more subheadings than others (UN 5.5%, others 0.057% – 1.63%).
Therapy	MI fewer subheadings than others (MI 0.58%, 0.59%, others 3.2% – 5.6%).
Ultrasonography	Pe more subheadings than MI, IM, and UN (Pe 2.2%, MI 0.038%, IM 0.22%, UN 0.64%).
Urine	MI fewer subheadings than IM, Pe and UN (MI 1.47%, IM 3.0%, Pe 4.7%, UN 4.0%)

Table 8.Results for average MeSH® main heading level.

Dimension	Significant (99% 2-sided CI) differences
Anatomy	MeSH® higher than others (mean of MeSH® 4.22, others 3.42 – 3.57).
	UN higher than MI, IM, and Pe (mean of UN 3.57, MI 3.48, IM 3.46, Pe 3.42).
Bacteria	MI higher than IM (mean of MI 5.36, IM 5.21). No other significant differences.
Infection	MeSH® higher than others (mean of MeSH® 4.58, others 3.36 – 3.69).
	MI higher than others (mean of MI 3.69, others 3.36 – 3.54).
Pathological Conditions, Signs and Symptoms	MeSH® higher than others (mean of MeSH® 4.59, others 4.06 – 4.12). No other significant differences.
Pathologic Processes	MeSH® higher than OG, UN, and GP (mean of MeSH® 4.17, OG 3.83, UN 3.70, GP 3.91).
	MI higher than all except for Pe (mean of MI 4.03, mean of others except for Pe 3.70 – 3.91), UN lower than others (mean of UN 3.70, others 3.83 – 4.03).
Disease Attributes	No significant differences.
Signs and Symptoms	MeSH® higher than others (mean of MeSH® 5.02, others 4.04 – 4.25).
	UN lower than all except for OG (mean of UN 4.04, others except for OG – 4.25).
Chemicals and Drugs Category	MeSH® higher than others (mean of MeSH® 4.79, others 4.46 – 4.66).
	Pe higher than others (mean of Pe 4.66, others 4.46 – 4.52).
Antibiotics	MeSH® higher than all except for MI (mean of MeSH® 3.58, others except for MI 2.50 – 2.75).
	MI higher than others (mean of MI 3.06, others 2.50 – 2.75).
Diagnostic Techniques and Procedures	MeSH® higher than others (mean of MeSH® 4.82, others 4.08 – 4.49).
	UN higher than others (mean of UN 4.49, others 4.08 – 4.29).
Laboratory Techniques and Procedures	MeSH® higher than all except for MI (mean of MeSH® 4.79, others except for MI 4.08 – 4.39).
	MI higher than others (mean of MI 4.90, others 4.07 – 4.39).
Specimen Handling	MeSH® higher than all except for UN (mean of MeSH® 4.94, others except for MI 3.81 – 3.99).
Therapeutics	MeSH® higher than all except for MI (mean of MeSH® 3.35, others except for MI 3.04 – 3.13).
	MI higher than others (mean of MI 3.25, others 3.04 – 3.13).
Physiological Processes	Scattered significant differences.
Age Groups	Pe higher than all but MeSH® (mean of Pe 3.83, others 3.69 – 3.71).
Health Personnel	-

To estimate the quality of the MEDLINE® database one could examine the rates of false negatives and positives as well as the variability of MEDLINE® indexing. To get an approximation of the rate of false negatives (1 – Sensitivity) and rate of false positives (1 – Positive Predictive Value) we examined earlier studies of the MEDLINE® sensitivity and specificity of the MEDLINE® database [10–15]. [13] The studies show that the rate of false negatives can be expected to be about 20%, while the rate of false positives is often much higher. None of the studies concerned general search tasks like the query in this study but, on the contrary, rather specific search tasks. Thus, interpretation of the results of these studies in the context of the present study must be done with some reservation. However, the rate of false positives should be lower in the case of general search tasks than in the case of specific tasks, since more articles should fit into the more open restrictions of the general search task.

The variability of indexing in the MEDLINE® database may also have influenced the material. The inter-indexer consistency has been reported to be 48.2% for non-major main headings [8]. Sources of this variability could be that indexers choose headings from a set of nearly equally suitable headings, or that indexers are inconsistent in their choice of whether or not to include a specific heading. However, by examining aggregations of headings such as, for example, our dimensions, instead of specific headings, we believe that the former source of variability can be reduced.

In conclusion, the MEDLINE® database must be used with an awareness of the fact that the databases, like other databases, have problems concerning precision and recall and that there is a significant amount of variability in indexing.

The other question concerned the adequacy of the MeSH® thesaurus and, more specifically, the thesaurus as it is used in representing the contents of the articles in the indexing process.

The material presented here was sampled over a time span of about 38 years (1963 – March 2000). The results reflect a mean for this period. During this time, the MeSH® thesaurus has changed considerably: the volume has increased and new levels of depth have been added. The changes in MeSH® will thus affect the standard deviation of the results. The standard deviation also reflects the shifts of focus in the research and practice regarding UTI management over the years, both directly and indirectly via the changes in MeSH®. It is interesting to note that the length of this period of time is about the same as the careers of the presently active physicians. The physicians who are now about to retire began working in the beginning of the 1960s. Thus, although the material might not reflect the current status of the area of UTI, it might better reflect the practice of UTI management in health care today. A study of the current concept of UTI should span a shorter time period.

Most of the domain differences are between microbiology/infectious diseases and the other domains. This is evident in the comparison of the use of subheadings, where the MI domain has a lower relative frequency of clinically related subhead-

ings such as ‘Complications,’ ‘Diagnosis,’ and ‘Prevention & Control.’ This difference most probably reflects a greater influence of the basic sciences in that domain, which is not as clear in other domains.

Most of the significant results have apparent and intuitive interpretations. For example, in the microbiology/infectious disease domain more ‘Bacteria’ and ‘Antibiotics’ headings per article are used than in other domains. In pediatrics, more age related headings are used per article than in other domains. Since most patients with UTI are treated by general practitioners, it is reasonable that there are more ‘Drug Therapy’ subheadings in that domain. Also, it is intuitive that in the ‘Laboratory Techniques and Procedures’ dimension, the level of main headings used in the microbiology/infectious diseases domain is higher than in other domains. Concerning the main heading levels, only small mean differences between domains can be noted. Typically, the mean differences are between 0.1–0.3, with the ‘Laboratory Techniques and Procedures’ dimension, with a width of range of 1.0, as the exception. The standard deviation is typically between 1.0–1.5, indicating that the variation is greater within the domains than between the domains. The question is then whether these results are sound. Obviously, at least some variations in granularity should exist. Thus, we conclude that the heading level does not adequately reflect differences in granularity of knowledge among domains. However, granularity may be expressed in other ways when indexing articles, such as, for example, through co-ordination [15].

In some dimensions, interpretation of the heading level as being related to granularity can be misleading. For example, in the ‘Anatomy’ domain the tree structure is mostly a part-whole hierarchy and not a “narrowness” hierarchy as is usual in other dimensions. The heading ‘Finger’ alone is not necessarily more narrow than the heading ‘Arm.’ However, the heading level is higher in the UN domain, a domain given by anatomy. The interpretation of a main heading may also depend on co-occurring headings. For example, when an ‘Anatomy’ heading co-occurs with a ‘Disease Category’ heading, the intended interpretation may be that the disease resides in the anatomical region or system. A disease that resides in the finger can be seen as having a “narrower” relation to a disease that resides in the arm. Thus, headings may be given new semantics by the co-occurring headings. The semantics may be further clarified by the use of subheadings, a fact that was utilized by Cimino and Barnett in the extraction of knowledge from the MEDLINE® database [16].

The results also show that the mean level is often higher in the MeSH® thesaurus than in actual use. This may be seen to conflict with the indexer directives to use the most specific heading that is still adequate [15]. This phenomenon can be explained, however, in that many more specific, i.e. higher level, headings are infrequently used but needed for coverage.

Some of the results are, however, counterintuitive. For example, there are fewer ‘Infection’ headings and ‘Etiology’ subheadings in the microbiology/infectious diseases domain than in other domains. The only explanation we can give here, assuming that more is said about the etiology of UTI in the microbiology/infectious dis-

eases domain, is that we either have misinterpreted the use of the headings or that MeSH® and the indexing process do not sufficiently reflect those aspects of medicine.

Thus, due to problems primarily concerning unclear semantics of in both the headings and the indexing process, we conclude that the MEDLINE® database is difficult to use for examining aspects of medical knowledge domains. The variability in the expression of semantics probably hides significant results, although the significant results that have been found are mostly reliable. The method seems to be primarily sound, although it is not complete. In addition, we find no reason to believe that these problems should not exist for other terminological systems. We believe that the problems found in this study do not only stem from problems specifically with MeSH® or MEDLINE®, but from the inexactness of the semantics of medicine [17].

ACKNOWLEDGEMENTS

This work was funded by the Knowledge Foundation (KK-stiftelsen) under the IT in Health Care (ITHS) programme.

REFERENCES

1. Aspevall O, Karlsson D, Forsum U. Building a concept system to structure the contents of a decision support system – A grounded theory study of concepts in the knowledge domain of urinary tract infection. 2001, Med Inform 26:115–129.
2. Liss P-E, Aspevall O, Karlsson D, Forsum U. Terms used to describe urinary tract infections – the importance of conceptual clarification, 2003, APMIS 111:291–299
3. Olesen F, Oestergaard I. Patients with urinary tract infection: proposed management strategies of general practitioners, microbiologists and urologists, Br. J. Gen. Pract. 45 (1995) 611–613.
4. Timpka T, Bjurulf P. The semantics of diagnosis and management of genitourinary infections: a cross-specialty study, Fam. Pract. 6 (1989) 279–285.
5. NLM™, List of Journals Indexed in Index Medicus®, [cited 2000 April 19] Available from: URL: http://www.nlm.nih.gov/tsd/serials/lji.html
6. NLM™, Journal selection for Index Medicus MEDLINE®, [cited 2000 April 19] Available from: URL: http://www.nlm.nih.gov/pubs/factsheets/jsel.html
7. Wyatt J. Use and sources of medical knowledge, Lancet 338 (1991) 1368–1373.
8. Funk M E, Reid C A. Indexing consistency in MEDLINE®, Bull. Med. Libr. Assoc. 71 (1983) 176–183.
9. Deville W L, Bezemer P D, Bouter L M. Publications on diagnostic test evaluation in family medicine journals: an optimal search strategy, J. Clin. Epidemiol. 53 (2000) 65–69.
10. van der Weijden T, IJzermans C J, Dinant G J, van Duijn N P, de Vet R, Buntinx F. Identifying relevant diagnostic studies in MEDLINE. The diagnostic value of the erythrocyte sedimentation rate (ESR) and dipstick as an example, Fam. Pract. 14 (1997) 204–208.
11. Nwosu C R, Khan K S, Chien P F. A two-term MEDLINE search strategy for identifying randomized trials in obstetrics and gynecology, Obstet. Gynecol. 91 (1998) 618–622.
12. Marson A G, Chadwick D W. How easy are randomized controlled trials in epilepsy to find on Medline? The sensitivity and precision of two Medline searches. Epilepsia 37 (1996) 377–380.
13. Dickersin K, Scherer R, Lefebvre C. Identifying relevant studies for systematic reviews, B.M.J. 309 (1994) 1286–1291.
14. Kirpalani H, Schmidt B, McKibbon K A, Haynes R B, Sinclair J C. Searching MEDLINE for randomized clinical trials involving care of the newborn, Pediatrics 83 (1989) 543–546.

15. Bachrach C A, Charen T. Selection of MEDLINE contents, the development of its thesaurus, and the indexing process, Med. Inform. 3 (1978) 237–254.
16. Cimino J J, Barnett G O. Automatic knowledge acquisition from MEDLINE, Methods Inf. Med. 32 (1993) 120–130.
17. Rector A L. Clinical Terminology: Why Is it so Hard? Methods Inf. Med. 38 (1999) 239–252.

Corresponding author: Urban Forsum
Avdelningen för klinisk mikrobiologi
Linköpings universitet
581 85 Linköping
Sweden
Phone: +46 13 22 20 90
Fax: +46 13 22 45 96
e-mail: urban.forsum@imk.liu.se