Wayne State University
Digital Commons@Wayne State University

Nursing Faculty Research Publications | Nursing

7-1-2004

Comparison of Medical Subject Headings and Text-Word Searches in MEDLINE to Retrieve Studies on Sleep in Healthy Individuals

Elizabeth Sergeevna Jenuwine
Wayne State University

Judith A. Floyd
Wayne State University, ab9208@wayne.edu

Recommended Citation
Jenuwine, E. S., & Floyd, J. A. (2004). Comparison of Medical Subject Headings and text-word searches in MEDLINE to retrieve studies on sleep in healthy individuals. Journal of the Medical Library Association, 92(3), 349-353. Accessed at http://www.ncbi.nlm.nih.gov/pmc/articles/PMC442177/
Available at: http://digitalcommons.wayne.edu/nursingfrp/4

This Article is brought to you for free and open access by the Nursing at Digital Commons@Wayne State University. It has been accepted for inclusion in Nursing Faculty Research Publications by an authorized administrator of Digital Commons@Wayne State University. For more information, please contact digitalcommons@wayne.edu.

Comparison of Medical Subject Headings and text-word searches in MEDLINE to retrieve studies on sleep in healthy individuals*

By Elizabeth S. Jenuwine, PhD, MLS
aa8696@wayne.edu
Research Technician

Judith A. Floyd, PhD, RN
ab9208@wayne.edu
Professor of Nursing

College of Nursing
Wayne State University
5557 Cass Avenue, Room 300
Detroit, Michigan 48202

Objective: The objective was to investigate the performance of two search strategies in the retrieval of primary research papers containing descriptive information on the sleep of healthy people from MEDLINE.

Methodology: Two search strategies—one based on the use of only Medical Subject Headings (MeSH), the second based on text-word searching—were evaluated as to their specificity and sensitivity in retrieving a set of relevant research papers published in the journal *Sleep* from 1996 to 2001 that were preselected by a hand search.

Results: The subject search provided higher specificity than the text-word search (66% and 47%, respectively) but lower sensitivity (78% for the subject search versus 88% for the text-word search). Each search strategy gave some unique relevant hits.

Conclusions: The two search strategies complemented each other and should be used together for maximal retrieval. No combination of MeSH terms could provide comprehensive yet reasonably precise retrieval of relevant articles. The text-word searching had sensitivity and specificity comparable to the subject search. In addition, use of text words ''normal,'' ''healthy,'' and ''control'' in the title or abstract fields to limit the final sets provided an efficient way to increase the specificity of both search strategies.

INTRODUCTION

Recent years have seen a proliferation in the use of research synthesis methodologies such as meta-analyses and systematic reviews in the health care fields. Driven by the increased emphasis on evidence-based health care practices, practitioners need more works that summarize the current state of research in their areas of interest or that integrate the research findings.

The first important step of any systematic review or meta-analysis is identification and retrieval of relevant publications in a systematic, comprehensive, and reproducible way. This is usually achieved by searching bibliographic databases. Of biomedical bibliographic databases, MEDLINE is the largest and most widely used in the world. A number of researchers have developed general strategies for MEDLINE searches of randomized controlled trials [1–3]; diagnostic studies [4, 5]; etiological, therapeutic, or prognosis studies [6]; and systematic reviews [7]. These strategies rely on the use of subject headings or text words that define the methodologies, clinical applications, and publication types in addition to the specific subject terms to achieve comprehensive yet accurate retrieval.

In some cases, however, relevant research papers may not be limited to a particular study type or methodology. This happens when the goal of a meta-analysis is to synthesize research findings on a particular phenomenon or activity. The meta-analysis of descriptive data on sleep of healthy humans conducted by our research group is an example of this kind of study. It

* This work was supported by National Institutes of Health grant no. RO1 NR 03880.

Table 1
Sleep-wake characteristics

Sleep continuity variables	Sleep latency
	Frequency of awakenings after sleep onset
	Duration of time awake after sleep onset
	Early morning awakening (time awake before arising)
	Total night time sleep
	Time in bed
	Sleep efficiency
	Sleep maintenance
Sleep architecture variables	Sleep stages
	REM sleep
	Number of REM periods
	REM latency
	REM density
Sleep-wake circadian pattern variables	Bedtime
	Sleep-time
	Wake-time
	Arise-time
	Nap frequency
	Nap duration
	Circadian placement of first nap
	Total sleep time (in 24 hours)
Sleep-wake quality variables	Morning restedness
	Sleepiness
	Fatigue
	Sleep satisfaction
	Sleep quality

involves research synthesis of descriptive information on the sleep of healthy people of different ages to see whether and how these sleep characteristics change over time. The research synthesis methods used to combine findings reported in primary research papers are described elsewhere [8].

The sleep characteristics of interest (Table 1) encompass a very broad area of sleep research. The reports containing relevant information include studies of normal sleep physiology and the effects of various conditions and/or substances on normal sleep, studies comparing the sleep of subject populations with various physical or psychological conditions to groups of healthy subjects, and large population surveys of sleep-related behaviors. They belong to various classes of studies, including diagnostic, therapeutic, and etiological studies. As a result, the approaches typically used to limit the retrieval of irrelevant studies from MEDLINE cannot be used in this case. On the other hand, selection of studies based on an important criterion relevant for our meta-analysis—whether or not they provide data for healthy subjects—is not readily supported by MEDLINE.

We developed two search strategies. The first one was based on the use of appropriate Medical Subject Headings (MeSH) (subject search), the second on the use of text words only (text-word search). We further evaluated the performance of these two search strategies by testing their ability to retrieve relevant research papers from the journal *Sleep*. This paper presents the results of a performance evaluation of two search strategies designed to achieve accurate and complete retrieval of primary research papers containing descriptive information on sleep and discusses the implications of this evaluation for achieving high retrieval in searching MEDLINE for broad subject areas.

METHOD

Design

The sleep parameters of interest include sleep continuity, architecture, circadian pattern, and subjective sleep quality characteristics (see Table 1 for the complete list). This list is the result of the previous work by one of the authors that identified and validated the terms used by sleep researchers to report their findings [9]. It represents a comprehensive list of quantitative characteristics of various aspects of sleep.

The studies containing any of these sleep parameters determined for healthy people and published in the journal *Sleep* from 1996 to 2001 were located by hand search and using electronic searching of MEDLINE. *Sleep* was selected, because it was indexed in MEDLINE and it was known by the authors to publish a significant number of relevant papers. In addition, the personal subscription of one of the authors made the journal readily available for hand searching.

The outcomes of the electronic searches were evaluated on the basis of their sensitivity and specificity. ''Sensitivity'' is the ability of a search to retrieve relevant articles. ''Specificity'' is the ability of the search to exclude irrelevant articles. The total number of articles published in *Sleep* from 1996 until 2001 (excluding editorials, reviews, practice guidelines, meta-analyses, animal studies, and papers on infants) was 575. Of them, 137 papers were relevant for our study and comprised the reference standard (see below), and 438 papers were irrelevant for our research purposes. We determined sensitivity to be the ratio of the number of the reference standard articles retrieved by the electronic search to the total number of articles in the reference standard. Specificity was determined to be the ratio of the number of irrelevant articles not retrieved by the search to the total number of irrelevant articles. The number of irrelevant articles not retrieved was determined as the difference between the total number of irrelevant articles (438) and the number of irrelevant articles retrieved by a search.

Reference standard

The ''reference standard'' is the complete set of relevant articles. In our study, the reference standard was composed of all primary research papers containing descriptive findings on sleep in healthy people published in *Sleep* from 1996 to 2001. One of the authors located these articles by a hand search of the journal issues for the time period indicated above. All papers—excluding editorials, reviews, practice guidelines, meta-analyses, and papers on infants and animals—were evaluated using a set of inclusion and exclusion criteria. To be included in the reference standard, the papers had to be primary research reports containing any number of the sleep-wake characteristics from Table 1 determined for healthy human sub-

Table 2
Subject search strategy for location of papers containing descriptive findings on sleep in healthy people published in *Sleep* from 1996 to 2001 in MEDLINE

	Search terms	Number of citations
#1	Sleep (exploded) (MeSH); 1996–2001, human, English	5,313
#2	Polysomnography (MeSH) OR Electroencephalography (MeSH) OR Circadian Rhythm (MeSH) OR Aging (MeSH) OR Time Factors (MeSH) OR Wakefulness (MeSH); 1996–2001, human, English	128,435
#3	#1 AND #2	3,086
#4	Sleep (jnl); 1996–2001, human; NOT age range = all infant (birth–23 months); NOT publication type = [review OR editorial OR practice guideline OR meta-analysis]	571
#5	#3 AND #4	254

jects excluding children under two. The final reference standard contained 137 papers.

The process of discrimination between relevant and irrelevant papers was rather straightforward and included no subjective judgment. The papers included in the reference standard were further evaluated for use in the meta-analysis; thus, we could be certain that the set contained no false positives. To ensure that no paper that satisfied our inclusion criteria was erroneously not included in the reference standard, the screening procedure was performed twice. If any of the relevant papers were still erroneously missed, this could slightly decrease the selectivity values reported here and increase the precision values.

MEDLINE search strategies

MEDLINE searches were performed via PubMed. Two search strategies were developed. In the subject search strategy (Table 2), only MeSH terms were used to locate the relevant studies, while the text-word strategy (Table 3) used free-text searching of all relevant fields to retrieve the studies of interest.

The text-word terms were selected based on the names of the sleep parameters under investigation (Table 1) and their slight variations. Some of the terms from Table 1 were not used in the text-word search, because they were known from the authors' previous experience to produce very few or no relevant hits (e.g., ''restedness,'' ''arise-time,'' and ''time in bed'') or be too unspecific (e.g., ''fatigue'' and ''awakening(s)'').

The choice of MeSH terms was based on the analysis of their definitions and the authors' experience. We combined the term ''Sleep (exploded)'' with MeSH terms for methodologies used for determination of sleep continuity and architecture variables (''Polysomnography'' and ''Electroencephalography''), the terms related to sleep-wake cycle (''Circadian Rhythm,'' ''Wakefulness,'' and ''Time Factors''), and the term ''Aging'' to retrieve studies on the effect of age on sleep characteristics. The search strategies were developed blind to the contents of the reference standard.

RESULTS

The results on sensitivity and specificity of the two search strategies are presented in Table 4. The text-word search had higher sensitivity. It retrieved 120 out of 137 relevant articles (88%). Of the 17 relevant articles not retrieved, 14 did not contain any of the specified terms, while the remaining 3 included infants as one of the age groups (one erroneously).

The subject search retrieved 107 relevant papers or 78% of the reference set (Table 4). Eight of these papers were retrieved only by the subject search but not by the keyword search. We further analyzed the performance of individual MeSH terms used in the subject search in combination with ''Sleep (exploded)'' in retrieval of papers from the reference set (Table 5). The analysis showed that each term provided some unique relevant hits. The best performing MeSH term, ''Polysomnography,'' provided only 50 (36%) of the relevant hits. The sensitivity of each tested MeSH term ranged from 5% to 36%, while the specificity varied from 85% to 99% (Table 5).

Analysis of the MeSH terms assigned to the 30 papers from the reference set not found by the subject search showed that the only headings that most of these papers shared were ''Sleep (exploded)'' or

Table 3
Text-word search strategy for location of papers containing descriptive findings on sleep in healthy people published in *Sleep* from 1996 to 2001 in MEDLINE

	Search terms	Number of citations
#1	sleep efficiency (tw) OR sleep latency (tw) OR sleep stages (tw) OR sleep maintenance (tw) OR sleep onset (tw) OR total sleep (tw) OR sleep-time (tw) OR REM (tw) OR sleepiness (tw) OR sleep duration (tw) OR sleep satisfaction (tw) OR tiredness (tw) OR sleep quality (tw) OR quality of sleep (tw) OR night waking (tw) OR sleep behavior (tw) OR sleep-wake cycle (tw) OR sleep length (tw) OR [(bedtime (tw) OR nap (tw) OR wake-time (tw)) AND sleep (tw)]; 1996–2001, human, English	4,636
#2	sleep (jnl); 1996–2001, human; NOT age range = all infant (birth–23 months); NOT publication type = [review OR editorial OR practice guideline OR meta-analysis]	571
#3	#1 AND #2	354

Table 4
Results of subject and text-word search strategies for location of papers containing descriptive findings on sleep in healthy people published in *Sleep* from 1996 to 2001 in MEDLINE

Search strategy	Total retrieved	Relevant retrieved	Sensitivity	Specificity
Subject searches:				
1. Total	254	107	78%	66%
2. Limited to ''normal,'' ''healthy,'' or ''control'' in titles or abstracts	130	77	56%	88%
Text-word search:				
1. Total	354	120	88%	47%
2. Limited to ''normal,'' ''healthy,'' or ''control'' in titles or abstracts	161	80	58%	81%

''Sleep Disorders (exploded).'' The search for papers containing MeSH terms of ''Sleep'' or ''Sleep Disorders'' allowed us to retrieve 133 out of 137 articles from the reference set (Table 5). The four papers not retrieved by this broad search included three papers with ''infants'' as one of the age groups and one paper that did not contain either of the terms. While the 97% sensitivity of this search is very impressive, these terms are too general to be practically used. In fact, the specificity of this search was only 2%. It retrieved 564 out of the total of 575 research papers published in *Sleep* from 1996 to 2001 (Table 5). Combining both the subject and keyword search strategies retrieved 128 out of 137 articles, thus providing a combined sensitivity of 93%.

To find the ways to increase the specificity of the searches, we analyzed titles and abstracts of the false positives retrieved by both strategies. Often these papers did not meet our inclusion criteria, because they did not contain information for healthy people. The health status of the research subjects is not considered when papers are assigned subject headings by MEDLINE indexers. Thus, no appropriate subject headings were available to discriminate between papers that contained information on healthy individuals and those that did not. In an attempt to select the papers that reported data on healthy individuals, we limited the final sets to the papers that contained the words ''normal,'' ''healthy,'' or ''control'' in titles or abstracts, because these words were likely to be used to describe healthy or general-health population groups. This approach significantly increased the specificity of both searches (Table 4). However, it also resulted in decreased sensitivity. Both searches together retrieved eighty-seven relevant articles (64%).

DISCUSSION

In this paper, we compared the performance of two MEDLINE search strategies in locating research papers containing descriptive findings on sleep. Each strategy offered some advantages and some drawbacks. Text-word searching was more successful than subject searching in locating the relevant papers but had lower specificity. It also required us to use a large number of terms to account for the variables we were interested in as well as differences in the use of terminology by different authors. Just as expected, the use of MeSH terms resulted in more precise searching but at a cost of lower sensitivity. Each search strategy provided some unique relevant hits. The text-word search accounted for twenty-one unique relevant hits and the subject search for eight. Thus, the search strategies complemented each other and should be used together for maximal retrieval.

We analyzed the sleep-related MeSH terms of the papers in the reference standard. The subject headings were either too narrow or too general to be completely satisfactory for our purposes. To achieve an almost comprehensive retrieval, one would have to use a combination of very general terms: ''Sleep'' and ''Sleep Disorders,'' but the very low specificity of such a search (Table 5) makes it inefficient for practical purposes. On the other hand, MEDLINE has no mechanism to select for papers that contain data on healthy individuals, one of the selection criteria in our study.

Table 5
Performance of different MeSH terms in locating papers containing descriptive findings on sleep in healthy people published in *Sleep* from 1996 to 2001 in MEDLINE

Search strategy	Total retrieved	Relevant retrieved	Unique relevant	Sensitivity	Specificity
Sleep (exploded) [MeSH] AND Aging [MeSH]	11	7	1	5%	99%
Sleep (exploded) [MeSH] AND Circadian Rhythm [MeSH]	59	31	2	23%	94%
Sleep (exploded) [MeSH] AND Electroencephalography [MeSH]	90	38	5	28%	88%
Sleep (exploded) [MeSH] AND Polysomnography [MeSH]	117	50	12	36%	85%
Sleep (exploded) [MeSH] AND Time Factors [MeSH]	77	40	5	29%	92%
Sleep (exploded) [MeSH] AND Wakefulness [MeSH]	107	47	6	34%	86%
Sleep (exploded) [MeSH] OR Sleep Disorders (exploded) [MeSH]	564	133	25	97%	2%

We relied on text words describing the health status of research subjects. However, according to our data, only about 65% to 70% of papers that presented data for healthy subjects mentioned this fact in their titles or abstracts (Table 4).

Our analysis is limited in that we tested only one journal. Different editors may have different standards regarding study description and use of terminology. A larger set of journals should be tested to achieve more general results. We should note, however, that our results are likely to represent the best-case scenario. *Sleep* is one of the major journals in the area of sleep research. It uses structured and, therefore, more descriptive abstracts. In addition, a number of research groups published multiple papers in this journal during the period tested, which increased the likelihood that the papers in our reference set would use consistent terminology. Thus, we believe that when used on a larger set of journals, our search strategies are likely to perform less well.

Even in this case, a number of titles and abstracts lacked sufficient descriptiveness. Both strategies combined failed to retrieve 7% of relevant studies when we did not limit to studies containing ''normal,'' ''healthy,'' or ''control'' in titles or abstracts. This number increased to 36% when we imposed this limit.

Researchers in other biomedical fields also indicate that lack of descriptiveness and consistent terminology in titles and abstracts impedes their ability to locate relevant information in MEDLINE [10–12], especially when selection of relevant papers is based on attributes for which no subject terms are available. One way to address this problem is to create new MeSH terms as the need arises. NLM continuously follows developments in the biomedical fields and implements changes to the MeSH thesaurus as necessary. Our experience suggests that subject terms that describe health status of human research subjects could be useful for projects similar to ours. Further research is needed to determine whether other users of MEDLINE would benefit from this capability.

However, the number of MeSH terms cannot be expanded endlessly to satisfy any possible research need. With proliferation of research-synthesis approaches, the possibilities of what information in the documents may become selection criteria for researchers are endless. Text-word searching provides greater versatility and adaptability to particular research needs. Maximizing the potential of text-word searching requires titles and abstracts of the searched articles to be more descriptive and to use terminology more consistently. Authors of scientific publications should be aware that the quality of titles and abstracts is likely to affect retrieval of their works from electronic databases. Medical librarians can play an important role in conveying this message to prospective authors of research literature.

ACKNOWLEDGMENTS

We thank Lynda Baker for critically reviewing the manuscript.

REFERENCES

1. Adams CE, Power A, Frederick K, Lefebre C. An investigation of the adequacy of MEDLINE searches for randomized controlled trials (RCTs) of the effects of mental health care. Psychol Med 1994 Aug;24(3):741–8.
2. Nwosu CR, Khan KS, Chien PF. A two-term MEDLINE search strategy for identifying randomized trials in obstetrics and gynecology. Obstet Gynecol 1998 Apr;91(4):618–22.
3. Watson RJ, Richardson PH. Identifying randomized controlled trials of cognitive therapy for depression: comparing the efficiency of Embase, MEDLINE and PsycINFO bibliographic databases. Br J Med Psychol 1999 Dec;72(Pt 4): 535–42.
4. Devillé WL, Bezemer PD, Bouter LM. Publications on diagnostic test evaluation in family medicine journals: an optimal search strategy. J Clin Epidemiol 2000 Jan;53(1):65–9.
5. Van der Weijden T, Ijzermans CJ, Dinant GJ, van Duijn NP, 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 1997 Jun;14(3):204–8.
6. Felber SH. Searching for evidence-based oncology: tips and tools for finding evidence in the medical literature. Cancer Control 2000 Sep–Oct;7(5):469–75.
7. Boynton J, Glanville J, McDaid D, Lefebre C. Identifying systematic reviews in MEDLINE: developing an objective approach to search strategy design. J Info Sci 1998 Jun; 24(3):137–57.
8. Floyd JA, Janisse JJ, Medler SM, Ager JW. Nonlinear components of age-related change in sleep initiation. Nurs Res 2000 Sep–Oct;49(5):290–4.
9. Floyd JA, Falahee ML, Fhobir RH. The use of the arcs software system to store and examine sleep research results. Comput Nurs 1999 Nov–Dec;17(6):259–67.
10. Murphy LS, Reinsch S, Najm WI, Dickerson VM, Seffinger MA, Adams A, Mishra SI. Searching biomedical databases on complementary medicine: the use of controlled vocabulary among authors, indexers and investigators. BMC Complement Altern Med 2003 Jul 7;3(1):3.
11. Dijkers MP. Searching the literature for information on traumatic spinal cord injury: the usefulness of abstracts. Spinal Cord 2003 Feb;41(2):76–84.
12. Derry S, Kong Loke Y, Aronson JK. Incomplete evidence: the inadequacy of databases in tracing published adverse drug reactions in clinical trials. BMC Med Res Methodol 2001 Sep 3;1(1):7.

Received August 2003; accepted March 2004