

# A Statistical Analysis of the Vocabulary of Medical Research Articles (1): Comparison with the Cobuild Frequency Count

Faculty of Social Information Science, Kure University  
Simon A. Fraser

**Abstract** The nature of vocabulary selection has long been problematical for ESP practitioners such as those working in the field of Medical English, a genre of interest to the readers of this journal. In an attempt to quantify the vocabulary of medical research articles, this paper reports on a computer corpus comprising articles covering a range of specialized medical fields. The occurrences of the most frequent items on this medical list are compared with their corresponding totals on the Cobuild corpus of general English. Considerable differences are found between the two frequency lists, and the most important of these are discussed. It is suggested that a general English frequency list is inadequate as a source for vocabulary selection for a Medical English curriculum, and that a subject-specific medical word count has far greater applicability.

**Key Words:** English for Specific Purposes (ESP), Medical English, Frequency count, Cobuild corpus

## 1 INTRODUCTION

As Hitchcock<sup>1)</sup> points out, the nature of vocabulary selection for ESP is probably the most misunderstood aspect of the discipline because of the fears it engenders. Analysis of a word frequency count would appear to be a valid way of quantifying the vocabulary of scientific texts. However, most available word counts are not relevant to ESP because their samples are not specifically from technical sources.

This is the first in a series of three papers which will look in detail at a frequency word count derived from a corpus representing a sample of the language used in a genre of interest to the readers of this journal —that of the medical research article. In these papers, I would like to consider the following questions, and discuss the implications of my findings for

both ESP teachers and learners.

- 1) What are the most frequent words in a frequency list of a medical corpus, and how do these differ from a general frequency list (the most frequent words in the Cobuild corpus)?
- 2) How are these words distributed across the “IMRAD” (Introduction, Methods, Results, Discussion) structure of medical articles?
- 3) What is the frequency of occurrence and distribution across research articles in different medical fields of the following types of “subtechnical” words (common words which occur with special meanings in specific scientific and technical fields)? How are these words distributed across IMRAD, and how are they typically used?
  - a) discourse-structuring words (words which organize and structure the argument of a text)

b) "cryptotechnical" words (polysemous words whose technical sense is hidden and so can confuse the learner)

4) What are the most frequent "technical" words, and what are their distribution patterns?

(For further explanation of the above categories of words, see Fraser<sup>2)</sup>, elsewhere in this journal.)

The present paper will address the first of these questions, and questions 2, 3 and 4 will be considered in two papers to be published in the next issue of this journal.

## 2 THE ANALYSIS

### 2.1 The Medical Corpus

For this study, I used the "IMRAD" corpus compiled at the Institute of Applied Language Studies, Edinburgh University. This consists of medical articles which have been keyed in, and which all follow the layout of Introduction, Methods, Results and Discussion. The computer was used to analyze a total of sixteen medical articles. These articles covered a wide range of specialized medical fields, and were taken from seven different journals. Four of these journals are British: *The British Medical Journal*, *The Lancet*, *Scottish Medical Journal* and *Journal of General Virology*; and three are American: *The Journal of Bone and Joint Surgery*, *The New England Journal of Medicine* and *The Journal of Pediatrics*. Appendix 1 contains the complete list of articles. This selection of texts is fairly representative, although a more even balance of American and British journals would have been preferable. The texts are typical of the kind of literature which post-graduate students and researchers in the medical field will encounter.

The computer was instructed to produce a frequency word list of the entire corpus, and frequency lists for each of the four rhetorical sections: Introductions (I), Methods (M), Results (R), and Discussions (D). The following counts were obtained:

TABLE 1

Medical Corpus		
	No. of words	% of total
Total word count	21,732	100
I	2885	13.3
M	6349	29.2
R	4357	20.0
D	8141	37.5

It can be seen that relatively little space is taken up in introducing the article, whereas a large proportion of the article is given over to the Discussion section in which the author comments on the findings of the research. This is not altogether surprising; Swales<sup>3)</sup> points out that since 1930, Discussion and Conclusion sections have not only become more common but have also greatly increased in length and complexity. However, the finding that a high proportion of space is given to the Methods section would seem to contradict the observation made by Swales that the space given to this section has declined over the years.

Finally, the computer was instructed to produce frequency word lists for each of the sixteen articles:

TABLE 2

Article	No. of words
Paper 1 (New England Journal of Medicine)	2116
Paper 2 (BMJ)	2787
Paper 3 (BMJ)	562
Paper 4 (BMJ)	1418
Paper 5 (Journal of Pediatrics)	2263
Paper 6 (Journal of General Virology)	1473
Paper 7 (BMJ)	2647
Paper 8 (BMJ)	416
Paper 9 (Scottish Medical Journal)	597
Paper 10 (The Lancet)	2317
Paper 11 (Journal of Bone and Joint Surgery)	1248
Paper 12 (BMJ)	1971
Paper 13 (The Lancet)	2227
Paper 14 (The Lancet)	1216
Paper 15 (The Lancet)	594
Paper 16 (BMJ)	2445

Mean article length = 1644 words  
Standard deviation = 806.5

It is apparent from these figures that there is a very wide range in article length, with the longest article comprising 2787 words and the shortest only 416 words. Interestingly, both articles appear in the same journal, the *British Medical Journal*.

### 2.2 Problems with the Analysis

#### 2.2.1 The Definition of "Word"

There is a problem with the definition of

“word”, the concept of which remains blurred. The notion of “word” used in this analysis is *word-form* — an unbroken succession of letters. Although this is similar to the usual idea of “word”, it is not identical. *Report, reports, and reporting*, for example, are different word-forms, but may all be regarded as instances of the same word.

### 2.2.2 The Problem of Homographs

The computer is unable to differentiate between homographs. We need to know with which particular sense an item is being used. For example, is *mean* being used with its general sense or its technical meaning? Are modal verbs such as *may* being used deontically or epistemically?

### 2.2.3 Multi-word Items

Nominal compounds and specialized multi-word items will not be identified. Medical English often uses compound nominal phrases such as “ischaemic heart disease”, “blood vessel narrowing”, and so on.

## 3 COMPARISON WITH THE COBUILD FREQUENCY COUNT

The occurrences of the most frequent items on the medical list were compared with their corresponding totals on the 7.3 million word Cobuild corpus at Birmingham University (see Kennedy<sup>4</sup>) for an account of the design and development of the Cobuild project and other corpora). The frequency word list of the medical corpus is given in Appendix 2.

### 3.1 Was and Were

TABLE 3

	Medical			Cobuild		
	Freq.	%	Position	Freq.	%	Position
Was	288	1.33	7	54722	0.75	9
Were	263	1.21	8	18547	0.25	37

Table 3 shows how the frequencies of *was* and *were* differ between the medical corpus and the Cobuild corpus. The first column (Freq.) shows the actual frequency of occurrence in the

corpus, and the second column (%) is the frequency expressed as a percentage of the total number of words in the corpus. The third column (Position) is the position of the item in the word list, where words are ranked in order of frequency of occurrence.

Both *was* and *were* occur relatively more frequently in the medical corpus than in the Cobuild corpus. *Were*, in particular, occurs almost five times as often in the medical corpus. This would suggest that there is, as is commonly believed, a tendency to use verbs in the passive voice more in medical research articles. If we look at how *was* and *were* are used in context, we do indeed see them commonly being used in passive constructions like the following:

1. Negative results were obtained for all controls... (Paper 8, p.722)
2. Statistical analysis was done with the McNemar's test for matched samples. (Paper 14, p.12)

There is support, then, for the findings of frequency counts performed on corpora of data from a variety of scientific and technical fields that show that the passive voice is used extensively in ESP (see, for example, Swales<sup>5</sup>) and Robinett<sup>6</sup>).

### 3.2 Pronouns

Interesting differences were found in the frequency of pronouns between the medical corpus and the Cobuild corpus. These are summarized in the following table:

TABLE 4

	Medical			Cobuild		
	Freq.	%	Position	Freq.	%	Position
I	6	0.03	-----	64849	0.89	8
They	17	0.08	159	29512	0.40	21
He	0	0	-----	42057	0.58	12
She	0	0	-----	20958	0.29	31
We	44	0.20	52	20964	0.29	30
Our	45	0.21	51	6189	0.08	98
It	54	0.25	38	61379	0.84	9

#### 3.2.1 I

*I* appears only six times in the entire medical corpus, although it occurs with high

frequency in the Cobuild corpus. The reason for its low frequency in the medical corpus is simply that the majority of articles are co-authored, rather than any reluctance to use the first person singular form. In fact, both of the articles which have a single author use *I* (in Paper 8 it occurs four times, and in Paper 16 twice).

### 3.2.2 *They*

*They* occurs frequently in the Cobuild corpus (the 21<sup>st</sup> most frequent form). The frequency is similar to that of the other functional words *be*, *had* and *but*. In the medical corpus, however, *they* is a relatively low frequency word (the 159<sup>th</sup> most frequent form). In medical writing, there is a tendency to avoid *they*, perhaps to ensure clarity and avoid misunderstanding. Instead, the names of people, or nouns such as *patients*, are repeated. Other, more specific expressions are favored, e.g.:

Two fresh morgue specimens were examined.  
*All of them* demonstrated...

(Paper 11, p.1100, R)

### 3.2.3 *He / She*

*He* and *she* are frequent forms in the Cobuild corpus (12<sup>th</sup> and 31<sup>st</sup> respectively), but they are not found at all in the medical corpus. This is not surprising, considering the impersonal nature of the subject matter; there is very little need to refer to people either anaphorically or deictically.

### 3.2.4 *We / Our*

*We* occurs fairly frequently in the medical corpus (the 52<sup>nd</sup> most frequent word), although not as often as in the Cobuild corpus, where it is the 30<sup>th</sup> most frequent word. Interestingly, *our* occurs relatively more frequently in the medical corpus than in the Cobuild corpus. If we look at the research articles to see why *our* occurs so frequently, we find that it is often used when the authors wish to express authorial comment, as in the following examples:

1. *Our* study confirmed this observation....

(Paper 13, p.551, D)

2. *Our* findings suggest that .... the complaint needs to be taken seriously by general practitioners and hospital doctors.

(Paper 2, p.1699, D)

*Our* was found to collocate particularly often with *findings*.

### 3.2.5 *It*

*It* occurs frequently in the Cobuild corpus. The 9<sup>th</sup> most frequent word, it occurs more often than either *is* or *was*. In the medical corpus, it is less frequent (the 38<sup>th</sup> most frequent word). This is probably because *it* is not often used anaphorically or deictically, but mainly as a dummy pronoun as in the following examples:

1. ...*it* is probably unrealistic to expect parents to provide constant supervision.

(Paper 7, p.1196, D)

2. *It* has also been shown that negative suggestions or events can have an adverse effect on patient wellbeing.

(Paper 4, p.789, D)

## 3.3 Lay-technical Words

*Patients* and *disease* are two words which occur very frequently in the medical corpus but are not found in the top 200 words in the Cobuild corpus.

TABLE 5

	Medical		
	Freq.	%	Position
Patients	168	0.77	13
Disease	63	0.29	30

There two words might be examples of "lay-technical" words – words which are conceptually accessible to the lay person, but found usually in limited contexts. We would certainly expect both *patients* and *disease* to occur with high frequency in a medical context.

### 3.4 Modals

The following table shows the frequency of modals in the two corpora:

TABLE 6

Cobuild	Position	Freq.	%
1. would	44	14687	0.20
2. can	60	11271	0.15
3. could	71	9214	0.13
4. will	77	8834	0.12
5. may	118	----	----
6. should	133	----	----
7. might	155	----	----
Medical	Position	Freq.	%
1. may	50	45	0.21
2. would	96	26	0.12
3. can	113	22	0.10
4. could	118	21	0.10
5. will	160	17	0.08
6. should	194	14	0.06
7. might	----	----	0.03

It is important in any analysis of modal verbs to recognize the epistemic vs. non-epistemic distinction. Huddleston<sup>7)</sup> groups modals under three headings:

#### (a) Epistemic uses

These involve implications concerning the speaker's knowledge of the situation in question, e.g.

- i *He may be ill*
- ii *He must be a friend of hers*
- iii *He will have finished by now*

#### (b) Deontic uses

These uses are concerned with obligation, prohibition and permission, e.g.

- i *You can / may have another apple*
- ii *You must be in bed before 8 o'clock*
- iii *You shall have your money back*

#### (c) Subject-oriented uses

These involve some property or disposition on the part of whoever or whatever is referred to by the subject, e.g.

- i *Liz can run faster than her brother*
- ii *Liz would not lend me the money [so I borrowed it from Ed]*

Although it was not possible to analyze every modal form in context, some examples of how these words are used in the articles are given

below:

#### 3.4.1 May

1. A gap in the tendon *may* not be convincingly palpable... (Paper 11, p.1099, I)
2. ....and practical ways of alleviating them *may* be useful adjuncts to anti-psychotic drugs. (Paper 3, p.327, D)

It seems that *may* is often used to express epistemic modality. The above examples show typical uses.

#### 3.4.2 Would

1. Effective strategies to protect children from accidental head injury *would* therefore appreciably reduce childhood mortality. (Paper 7, p.1193, I)
2. ...exclusion of the 7 patients *would* have biased the results in favor of DCBE alone. (Paper 13, p.550, D)

*Would* was found most often to express epistemic modality, as is shown in the two examples above.

#### 3.4.3 Can

1. The 60cm flexible sigmoidoscope *can* reach the descending colon in most patients... (Paper 13, p.549, I)
2. However, Tejani has suggested that even in steroid-responsive, minimal-change disease the prognosis *can* be uncertain... (Paper 10, p.256, I)

We can see that *can* is found in its subject-oriented uses (example 1) and its epistemic uses (example 2).

#### 3.4.4 Could

1. Patients *could* use the cassette player for pleasure as well as treatment... (Paper 3, p.327, M)
2. The environmental factor *could* be genetic inbreeding .... or a toxic substance. (Paper 5, p.752, D)

*Could* was found to occur with both deontic (example 1) and epistemic (example 2) uses.

### 3.4.5 *Will*

1. Although this approach *will* have major financial and logistic consequences...  
(Paper 13, p.551, D)
2. Many patients, however, *will* not be willing to continue this for long...  
(Paper 10, p.258, D)

The above two examples show that *will* is typically used epistemically in the articles.

### 3.4.6 *Should*

1. Possible sexual abuse *should* also be considered...  
(Paper 12, p.521, D)
2. We recommend that urine *should* be tested at least twice weekly... (Paper 10, p.258, D)

Most instances of *should* are, like the above examples, deontic – concerned with obligation.

The interesting difference between the two lists is that whereas *would* is the most frequent modal in the Cobuild frequency list, *may* is by far the most frequent modal in the medical list. This reflects the fact that there are significant differences in the likelihood of different modals being used in spoken and written contexts. Holmes<sup>8)</sup>, in her research on the relative frequencies of different modals, found that *may* occurs much more frequently in written data than in spoken. She also found that epistemic *would* occurs much more often in speech than in writing, which would explain its high frequency in the Cobuild corpus, which includes both written and spoken modes. The order in which all the other words occur is identical, and the figures from both corpora show similar differences in the frequency of occurrence.

## 3.5 Discourse Markers

All the words in Table 7 occur relatively frequently in the medical corpus, but do not occur

TABLE 7

	Medical Corpus		
	Freq.	%	Position
However	32	0.15	76
Since	19	0.09	140
Thus	18	0.08	147

in the first 200 words of the Cobuild corpus. As Salager<sup>9)</sup> says, we would expect such words to be found with high frequency because they are used to express “the grammatical-conceptual manifestations of advanced and complex thought”.

### 3.5.1 *However*

*However* is often used when adding a comment that seems to contradict or contrast with what has just been said. It is not surprising, then, that we find it so frequently in medical research articles, since we would expect the author to point out the limitations of both previous research and her own findings. *However* was found to be used often as in the following examples:

1. However, the diagnostic accuracy of DEBE has been questioned... (Paper 13, p.549, I)
2. These morphologic changes are not proof of a viral etiology, however, since identical findings have been induced in experimental animals. (Paper 5, p.752, D)

### 3.5.2 *Since*

*Since* was found mainly in its sense “because”. We might well expect a word like this to be frequent in research articles where the author is often required to state the reasons for his actions or findings.

1. *Since* the results of our study showed more cases of EHBA than expected, correction of any error in the assumption would tend to make the statistical analyses more significant. (Paper 5, p.752, D)
2. *Since* there were no cases in subjects outside the target population, the use of ....did not affect the P values.

(Paper 1, p.1103, M)

### 3.5.3 *Thus*

*Thus* is a formal word, and we would not expect to find it in the spoken language. However, it is not surprising to find it in the language of research articles where it is used to mean “therefore” or “as a result”.

1. Detection of IL-1 and TNF activities in BWs on day 4 p.i. coincides with and increase in the number of IL-1 and TNF-producing cells. Thus a replication stage would be necessary to trigger...

(Paper 6, p.478, D)

2. ...the total number of patients was about 175. Thus the predictions of Tejani from his small series of 48 patients have not been confirmed.

(Paper 10, p.258, D)

### 3.6 *During*

The following examples give us some clues as to why *during* should occur so frequently in medical research articles:

1. 97% of our cohort relapsed *during* follow-up.

(Paper 10, p.258, D)

2. We identified all children aged below 16 years who died with a head injury *during* 1979-86.

(Paper 7, p.1193, M)

3. Tejani showed that *during* follow-up ... renal death occurred in 25%.

(Paper 10, p.256, I)

In medical articles, it is often necessary to identify time periods in which deaths, relapses or complications occur. *During* was found to collocate with *follow-up* particularly frequently.

## 4 CONCLUSION

This study has reported on a statistical analysis of a computer corpus comprising sixteen medical research articles drawn from seven different journals. This highly specialized medical list and the Cobuild frequency list were compared and found to differ considerably.

To summarize, we can say the following

about vocabulary in medical journal articles:

1. There is a preponderance of verbs in their “past” forms, which indicates that there is a tendency to use the past tense and passive voice.
2. There are words occurring with particularly high frequency which could be considered to be general, but which are rare in a non-medical context.
3. In comparison with other modals, *may* is particularly favored.
4. “Discourse markers” are found to occur with high frequency.

### 4.1 Limitations of the Study

It has to be borne in mind when considering the results that the size of the sample is relatively small for a study using a computer corpus. A larger study would help to ensure representativeness and reduce the chance of high frequency words from one text causing unexpected results. Also, it would have been desirable for there to have been more American texts in the sample in order to avoid a bias towards British English. Another problem that we mentioned before was the inability of the computer to separate homographs. Finally, although the articles are fairly representative of the field of medical research, there are some obvious gaps. Gynaecology, for instance, is not represented.

### 4.2 Pedagogical Implications

A knowledge of the frequency and distribution of lexical items in the medical research genre should give the teacher a better idea of which items to teach. My findings have shown that a general frequency list is not adequate as a source for vocabulary selection for a medical English curriculum. We were concerned with only the top 200 words in the Cobuild list, but there is more than a little disparity between it and the medical list. Because of its subject-specificity, the medical word count has far greater applicability. The most frequent words with high distribution across the different medi-

cal fields will surely have the greatest claim to ensure that our teaching is more efficient.  
relevance. Giving priority to these items should

### References

- 1) J. Hitchcock. 1978. Reading and scientific English: prospects, problems and programs in Iran. In Trimble et al. (eds). *English for Specific Purposes: Science and Technology*. English Language Institute, Oregon State University.
- 2) S. A. Fraser. 2001. Vocabulary and the teaching of English for specific purposes. *Integrated Studies in Nursing Science*. 3: 32-37.
- 3) J. Swales. 1990. *Genre Analysis*. Cambridge University Press.
- 4) G. Kennedy. 1998. *An Introduction to Corpus Linguistics*. Longman.
- 5) J. Swales. 1976. Verb frequencies in English. *ESPMENA Bulletin* 4: 28-31.
- 6) B. Robinett. 1980. *Final Report: Delineation of Linguistic Features of Scientific and Technical English*. Unpublished manuscript submitted to Control Data Corporation.
- 7) R. D. Huddleston. 1988. *English Grammar: an Outline*. Cambridge University Press.
- 8) J. Holmes. 1988. Doubt and certainty in ESL textbooks. *Applied Linguistics* 9 (1): 21-44.
- 9) F. Salager. 1983. The lexis of fundamental English: classificatory framework and rhetorical function (a statistical approach). *Reading in a Foreign Language* 1 (1): 54-64.

### APPENDIX 1

#### The Medical Journal Articles

##### PAPER 1

I. L. Acharya, C. U. Lowe et al. 1987. Prevention of typhoid fever in Nepal with the vi capsular polysaccharide of salmonella typhi. *The New England Journal of Medicine* 317 (18): 1101-1103.

##### PAPER 2

C. E. Price, R. J. Rona et al. 1988. Height of primary school children and parents' perceptions of food intolerance. *British Medical Journal* 296: 1696-1699.

##### PAPER 3

H. E. Nelson, S. Thrasher, T. R. E. Barnes 1991. Practical ways of alleviating auditory hallucinations. *British Medical Journal* 302: 327.

##### PAPER 4

T. T. C. McLintock, H. Aitken et al. 1990. Postoperative analgesic requirements in patients exposed to positive intraoperative suggestions. *British Medical Journal* 301: 788-790.

##### PAPER 5

A. D. Strickland, K. Shannon. 1982. Studies in the etiology of extrahepatic biliary atresia: Time-space clustering. *The Journal of Pediatrics* 100: 749-753.

##### PAPER 6

F. Vacheron, A. Rudent et al. 1990. Production of interleukin 1 and tumour necrosis factor activities in



bronchoalveolar washings following infection of mice by influenza virus. *Journal of General Virology* 71: 477-479.

PAPER 7

P. M. Sharples, A. Storey et al. 1990. Causes of fatal childhood accidents involving head injury in Northern region. *British Medical Journal* 301: 1193-1197.

PAPER 8

J. E. Hambidge. 1990. Use of skin thermometer to diagnose acute appendicitis. *British Medical Journal* 300: 722.

PAPER 9

G. Lamont, C. J. Simpson et al. 1986. Does cimetidine alter the prognosis following perforated duodenal ulcer? *Scottish Medical Journal* 31: 198.

PAPER 10

M. A. Lewis, E. M. Baildom et al. 1989. Nephrotic syndrome: From toddlers to twenties. *The Lancet, February* 4: 255-257.

PAPER 11

T. O'Brien. 1984. The needle test for complete rupture of the Achilles tendon. *The Journal of Bone and Joint Surgery* 66-A: 1099-1101.

PAPER 12

J. O. Beattie, D. Hull, F. Cockburn 1986. Children intoxicated by alcohol in Nottingham and Glasgow, 1973-84. *British Medical Journal* 292: 519-521.

PAPER 13

P. Durdey, P. M. T. Weston, N. S. Williams. 1987. Colonoscopy or barium enema as initial investigation of colonic disease. *The Lancet*, September 5: 549-551.

PAPER 14

R. A. Shinton, J. S. Gill et al. 1987. The frequency of epilepsy preceding stroke: Case control study in 230 patients. *The Lancet*, January 3: 11-12.

PAPER 15

Managing drug dealers who swallow the evidence. *British Medical Journal* (authors and date unknown).

PAPER 16

Snoring as a risk factor for ischaemic heart disease and stroke in men. 1985. *British Medical Journal, September* 7 (authors unknown).

## APPENDIX 2

### Frequency Word List of the Medical Corpus

The following list shows the word forms in the medical corpus, ranked in order of frequency of occurrence, with a cut-off point at 10. The distribution of each word across the “IMRAD” section is shown, and the second figure for each item (%) is the percentage of the total number of words in the section.

	Freq.	I	M	R	D
1. THE	1347	152 5.27%	415 6.53%	314 7.21%	466 5.72%
2. OF	1038	158 5.48%	264 4.16%	203 4.66%	413 5.72%
3. AND	588	83 2.88%	183 2.88%	116 2.66%	206 2.53%
4. IN	581	86 2.98%	111 1.75%	135 3.10%	249 3.06%
5. TO	415	56 1.94%	113 1.78%	65 1.49%	181 2.22%
6. A	411	60 2.08%	127 2.00%	61 1.40%	163 2.00%
7. WAS	288	7 0.24%	134 2.11%	84 1.93%	63 0.77%
8. WERE	263	8 0.28%	133 2.09%	67 1.54%	55 0.68%
9. WITH	235	34 1.18%	74 1.17%	49 1.12%	78 0.96%
10. FOR	228	32 1.11%	82 1.29%	32 0.73%	82 1.00%
11. IS	186	40 1.39%	22 0.35%	8 0.18%	116 1.42%
12. THAT	177	30 1.04%	26 0.41%	24 0.55%	97 1.19%
13. PATIENTS	168	19 0.66%	38 0.60%	46 1.06%	65 0.80%
14. CHILDREN	145	17 0.59%	24 0.38%	53 1.22%	54 0.66%
15. OR	136	8 0.28%	61 0.96%	26 0.60%	41 0.50%
16. BY	135	24 0.83%	49 0.77%	17 0.39%	45 0.55%
17. AS	122	19 0.66%	40 0.63%	17 0.39%	45 0.55%
18. ON	117	17 0.59%	45 0.71%	22 0.50%	33 0.41%
19. BE	115	20 0.69%	10 0.16%	6 0.14%	79 0.97%
20. NOT	99	14 0.48%	28 0.44%	6 0.14%	51 0.63%
21. FROM	98	9 0.31%	33 0.52%	18 0.41%	38 0.47%
22. THIS	95	9 0.31%	16 0.25%	13 0.30%	57 0.70%
23. AT	92	10 0.35%	35 0.55%	25 0.57%	22 0.27%

# A Statistical Analysis of the Vocabulary of Medical Research Articles (1)

	Freq.	I	M	R	D
24. HAD	92	2 0.07%	34 0.54%	30 0.69%	26 0.32%
25. AN	89	18 0.62%	9 0.14%	9 0.21%	53 0.65%
26. ARE	80	18 0.62%	8 0.13%	6 0.14%	48 0.59%
27. HAVE	76	19 0.66%	8 0.13%	7 0.16%	42 0.52%
28. WHO	69	4 0.14%	18 0.28%	15 0.34%	32 0.39%
29. BEEN	68	22 0.76%	15 0.24%	8 0.18%	23 0.28%
30. DISEASE	63	20 0.69%	14 0.22%	7 0.16%	22 0.27%
31. ALL	61	3 0.10%	21 0.33%	17 0.39%	20 0.25%
32. THESE	60	7 0.24%	12 0.19%	12 0.46%	29 0.36%
33. TWO	60	9 0.31%	20 0.31%	22 0.50%	9 0.11%
34. YEARS	60	7 0.24%	15 0.24%	12 0.46%	26 0.32%
35. STUDY	57	10 0.35%	12 0.19%	1 0.02%	34 0.42%
36. THEIR	56	7 0.24%	19 0.16%	9 0.21%	30 0.37%
37. HAS	55	27 0.94%	7 0.11%	1 0.02%	20 0.25%
38. IT	54	4 0.14%	9 0.14%	7 0.16%	34 0.42%
39. THAN	54	4 0.14%	16 0.25%	9 0.21%	25 0.31%
40. THOSE	54	6 0.21%	11 0.17%	15 0.34%	22 0.27%
41. NO	51	8 0.28%	12 0.19%	15 0.34%	16 0.20%
42. AFTER	49	2 0.07%	17 0.27%	16 0.37%	14 0.17%
43. PATIENT	48	9 0.31%	17 0.27%	7 0.16%	15 0.11%
44. AGE	47	4 0.14%	14 0.22%	17 0.39%	12 0.15%
45. THREE	47	2 0.07%	22 0.35%	15 0.34%	8 0.10%
46. ALCOHOL	46	2 0.07%	7 0.11%	14 0.32%	23 0.28%
47. ONLY	46	5 0.17%	9 0.14%	13 0.30%	19 0.23%
48. WHICH	46	12 0.42%	12 0.19%	7 0.16%	15 0.18%
49. GROUP	45	5 0.17%	15 0.24%	17 0.39%	8 0.10%
50. MAY	45	8 0.28%	2 0.03%	0 0%	35 0.43%

**A Statistical Analysis of the Vocabulary of Medical Research Articles (1)**

	Freq.	I	M	R	D
51. OUR	45	2 0.07%	1 0.02%	1 0.02%	41 0.50%
52. WE	44	9 0.31%	8 0.13%	0 0%	27 0.33%
53. BUT	43	5 0.17%	7 0.11%	11 0.25%	20 0.25%
54. DCBE	43	5 0.17%	4 0.06%	14 0.32%	20 0.25%
55. MOST	43	12 0.42%	5 0.08%	8 0.18%	18 0.22%
56. CASES	42	4 0.14%	5 0.08%	13 0.32%	20 0.25%
57. BETWEEN	38	5 0.17%	12 0.19%	10 0.23%	11 0.14%
58. DURING	38	9 0.31%	8 0.13%	9 0.21%	12 0.15%
59. VI	38	11 0.38%	10 0.16%	5 0.11%	12 0.15%
60. ALSO	37	4 0.14%	12 0.19%	4 0.09%	17 0.21%
61. HOSPITAL	36	6 0.21%	17 0.27%	6 0.14%	7 0.09%
62. IF	36	1 0.03%	11 0.17%	1 0.02%	23 0.28%
63. INTOLERANCE	36	9 0.31%	11 0.17%	4 0.09%	12 0.15%
64. MORE	36	6 0.21%	11 0.17%	3 0.07%	16 0.20%
65. OTHER	36	3 0.10%	15 0.24%	5 0.11%	13 0.16%
66. STROKE	35	7 0.24%	14 0.22%	14 0.09%	10 0.12%
67. TEST	35	4 0.14%	12 0.19%	12 0.28%	7 0.09%
68. USE	35	0 0%	5 0.08%	4 0.09%	16 0.20%
69. WHEN	35	3 0.10%	15 0.24%	5 0.11%	12 0.15%
70. SYMPTOMS	34	4 0.14%	14 0.22%	2 0.05%	14 0.17%
71. THERE	34	9 0.31%	5 0.08%	9 0.21%	11 0.14%
72. TREATMENT	34	2 0.07%	11 0.17%	11 0.25%	10 0.12%
73. COLONOSCOPY	33	3 0.10%	5 0.08%	10 0.23%	15 0.18%
74. EHBA	33	0 0%	6 0.09%	4 0.09%	23 0.28%
75. NUMBER	33	0 0%	9 0.14%	10 0.23%	14 0.17%
76. HOWEVER	32	8 0.28%	0 0%	4 0.09%	20 0.25%
77. INTO	32	1 0.03%	20 0.32%	7 0.16%	4 0.05%

# A Statistical Analysis of the Vocabulary of Medical Research Articles (1)

	Freq.	I	M	R	D
78. DIAGNOSIS	31	5 0.17%	10 0.16%	6 0.14%	10 0.12%
79. POSITIVE	31	4 0.14%	8 0.13%	7 0.16%	12 0.15%
80. TABLE	31	0 0%	5 0.08%	26 0.60%	0 0%
81. EACH	30	1 0.03%	22 0.35%	4 0.09%	3 0.04%
82. RESULTS	30	2 0.07%	6 0.08%	5 0.11%	17 0.21%
83. FIRST	29	2 0.07%	6 0.09%	7 0.11%	14 0.21%
84. GROUPS	29	3 0.10%	15 0.24%	6 0.14%	5 0.06%
85. HEIGHT	29	3 0.10%	17 0.27%	1 0.02%	8 0.10%
86. ANY	28	0 0%	10 0.16%	5 0.11%	13 0.16%
87. MEN	28	3 0.10%	9 0.12%	6 0.14%	10 0.12%
88. CELLS	27	5 0.17%	8 0.13%	3 0.07%	11 0.14%
89. RELAPSE	27	1 0.03%	3 0.05%	2 0.05%	21 0.26%
90. AGED	26	4 0.14%	4 0.06%	12 0.28%	6 0.07%
91. ETAL	26	3 0.10%	3 0.05%	2 0.05%	18 0.22%
92. ITS	26	10 0.35%	7 0.11%	2 0.05%	7 0.09%
93. PER	26	1 0.03%	8 0.13%	12 0.28%	5 0.06%
94. REPORTED	26	6 0.21%	4 0.06%	3 0.07%	13 0.16%
95. SUCH	26	10 0.35%	4 0.06%	0 0%	12 0.15%
96. WOULD	26	2 0.07%	4 0.06%	1 0.02%	19 0.23%
97. ADMITTED	25	3 0.10%	7 0.11%	9 0.21%	6 0.07%
98. BECAUSE	25	6 0.21%	2 0.03%	6 0.14%	11 0.14%
99. OVER	25	2 0.07%	5 0.02%	10 0.23%	8 0.10%
100. TIME	25	1 0.03%	10 0.16%	8 0.18%	6 0.07%
101. TNF	25	2 0.07%	4 0.06%	10 0.23%	9 0.11%
102. ACTIVITY	24	1 0.03%	3 0.05%	13 0.30%	7 0.09%
103. CHILD	24	1 0.03%	8 0.13%	8 0.18%	7 0.09%
104. FOUND	24	6 0.21%	2 0.03%	5 0.11%	11 0.14%

**A Statistical Analysis of the Vocabulary of Medical Research Articles (1)**

	Freq.	I	M	R	D
105. UNDER	24	2 0.07%	6 0.09%	12 0.28%	4 0.05%
106. VIRUS	24	6 0.21%	7 0.11%	6 0.14%	5 0.06%
107. BOTH	23	3 0.10%	4 0.06%	7 0.16%	9 0.11%
108. INTOLERANT	23	1 0.03%	4 0.06%	1 0.02%	17 0.21%
109. SNORERS	23	1 0.03%	10 0.16%	3 0.07%	9 0.11%
110. SNORING	23	8 0.28%	5 0.02%	1 0.02%	9 0.11%
111. TENDON	23	5 0.17%	4 0.06%	11 0.25%	3 0.04%
112. ANALYSIS	22	1 0.03%	15 0.24%	1 0.02%	5 0.06%
113. CAN	22	7 0.24%	0 0%	1 0.02%	14 0.17%
114. CONTROL	22	1 0.03%	10 0.16%	6 0.14%	5 0.06%
115. FEVER	22	10 0.35%	4 0.06%	4 0.09%	4 0.05%
116. OCCURRED	22	2 0.07%	5 0.08%	13 0.30%	2 0.02%
117. SUBJECTS	22	0 0%	10 0.16%	5 0.11%	7 0.09%
118. COULD	21	1 0.03%	5 0.08%	1 0.02%	14 0.17%
119. FIVE	21	1 0.03%	10 0.16%	10 0.23%	0 0%
120. ISCHAEMIC	21	2 0.03%	10 0.16%	2 0.05%	7 0.09%
121. RISK	21	3 0.10%	2 0.03%	0 0%	16 0.20%
122. TEXAS	21	1 0.03%	11 0.17%	2 0.05%	7 0.09%
123. USED	21	4 0.14%	12 0.19%	2 0.05%	3 0.04%
124. AMONG	20	0 0%	3 0.05%	5 0.11%	12 0.15%
125. EPILEPSY	20	5 0.17%	2 0.03%	1 0.02%	12 0.15%
126. FLEXIBLE	20	2 0.07%	2 0.03%	6 0.14%	10 0.12%
127. LESS	20	1 0.03%	5 0.08%	4 0.09%	10 0.12%
128. MANY	20	6 0.21%	0 0%	2 0.05%	12 0.15%
129. SUGGESTIONS	20	3 0.10%	2 0.03%	2 0.05%	13 0.16%
130. TYPHOID	20	11 0.38%	2 0.03%	3 0.07%	4 0.05%
131. ABOUT	20	1 0.03%	7 0.11%	1 0.02%	10 0.12%

# A Statistical Analysis of the Vocabulary of Medical Research Articles (1)

	Freq.	I	M	R	D
132. CLINICAL	19	5 0.17%	6 0.09%	2 0.05%	6 0.07%
133. COMPLETE	19	3 0.10%	5 0.08%	3 0.07%	8 0.10%
134. HEART	19	2 0.07%	9 0.14%	2 0.05%	6 0.07%
135. HOME	19	0 0%	3 0.05%	8 0.18%	8 0.10%
136. NEGATIVE	19	1 0.03%	6 0.09%	3 0.07%	9 0.11%
137. OFTEN	19	3 0.10%	2 0.03%	4 0.09%	10 0.12%
138. RUPTURE	19	7 0.24%	2 0.03%	4 0.09%	6 0.07%
139. SIGMOIDOSCOPY	19	2 0.07%	6 0.09%	3 0.07%	8 0.10%
140. SINCE	19	2 0.07%	4 0.06%	0 0%	13 0.16%
141. URINE	19	2 0.07%	11 0.17%	2 0.05%	4 0.05%
142. FINDINGS	18	4 0.14%	6 0.09%	1 0.02%	7 0.09%
143. HEAD	18	6 0.21%	5 0.08%	5 0.11%	2 0.02%
144. INJURY	18	5 0.17%	7 0.11%	4 0.09%	2 0.02%
145. POPULATION	18	2 0.07%	10 0.16%	1 0.02%	5 0.06%
146. SINGLE	18	1 0.03%	5 0.08%	4 0.09%	8 0.10%
147. THUS	18	1 0.03%	3 0.05%	3 0.07%	11 0.14%
148. YEAR	18	2 0.07%	2 0.03%	12 0.28%	2 0.02%
149. ACCIDENTS	17	1 0.03%	0 0%	10 0.23%	6 0.07%
150. BLOOD	17	2 0.07%	6 0.09%	6 0.14%	3 0.04%
151. DATA	17	2 0.07%	9 0.14%	3 0.07%	3 0.04%
152. DIAGNOSTIC	17	10 0.35%	2 0.03%	2 0.05%	3 0.04%
153. DID	17	1 0.03%	4 0.06%	4 0.09%	8 0.10%
154. FOLLOW-UP	17	4 0.14%	2 0.03%	5 0.11%	6 0.07%
155. PARENTS	17	1 0.03%	4 0.06%	3 0.07%	9 0.17%
156. RATE	17	2 0.07%	4 0.06%	4 0.09%	7 0.09%
157. STUDIES	17	6 0.21%	4 0.06%	0 0%	7 0.09%
158. THEN	17	2 0.07%	14 0.22%	0 0%	1 0.01%

**A Statistical Analysis of the Vocabulary of Medical Research Articles (1)**

	Freq.	I	M	R	D
159. THEY	17	1 0.03%	10 0.16%	6 0.14%	0 0%
160. WILL	17	1 0.03%	3 0.05%	1 0.02%	12 0.15%
161. ACHILLES	16	4 0.14%	3 0.05%	6 0.14%	3 0.04%
162. ANALGESIA	16	5 0.17%	2 0.05%	1 0.02%	8 0.10%
163. HISTORY	16	1 0.03%	11 0.17%	1 0.02%	3 0.04%
164. HOURS	16	1 0.03%	10 0.16%	3 0.07%	2 0.02%
165. INCREASED	16	4 0.14%	2 0.05%	3 0.07%	7 0.09%
166. MICE	16	2 0.07%	6 0.09%	7 0.16%	1 0.01%
167. ml.	16	0 0%	14 0.22%	2 0.05%	0 0%
168. NEEDLE	16	0 0%	7 0.11%	6 0.14%	3 0.04%
169. SO	16	1 0.03%	4 0.06%	1 0.02%	10 0.12%
170. SOME	16	5 0.17%	5 0.08%	1 0.02%	5 0.06%
171. WELL	16	3 0.10%	3 0.05%	1 0.02%	9 0.11%
172. ALONE	15	2 0.07%	0 0%	7 0.16%	6 0.07%
173. CAUSE	15	6 0.21%	2 0.03%	2 0.05%	5 0.06%
174. CHILD'S	15	0 0%	15 0.24%	0 0%	0 0%
175. CONSIDERED	15	1 0.03%	6 0.09%	4 0.09%	4 0.05%
176. GENERAL	15	3 0.10%	1 0.02%	2 0.05%	9 0.11%
177. INCLUDED	15	0 0%	8 0.13%	2 0.05%	5 0.06%
178. INFLUENZA	15	4 0.14%	5 0.08%	4 0.09%	2 0.02%
179. mg.	15	0 0%	14 0.22%	1 0.02%	0 0%
180. MONTHS	15	1 0.03%	7 0.11%	5 0.11%	2 0.02%
181. OLDER	15	0 0%	1 0.02%	10 0.23%	4 0.05%
182. RECEIVED	15	0 0%	11 0.17%	3 0.07%	1 0.01%
183. RESPONSE	15	3 0.10%	6 0.09%	2 0.05%	4 0.05%
184. SIGNIFICANT	15	2 0.07%	2 0.03%	4 0.09%	7 0.09%
185. TAKEN	15	1 0.03%	10 0.16%	2 0.05%	2 0.05%



# A Statistical Analysis of the Vocabulary of Medical Research Articles (1)

	Freq.	I	M	R	D
186. WHOM	15	3 0.10%	1 0.02%	7 0.16%	4 0.05%
187. ASSOCIATED	14	3 0.10%	2 0.03%	1 0.02%	8 0.10%
188. BEFORE	14	1 0.03%	8 0.13%	1 0.02%	4 0.05%
189. DOSES	14	1 0.13%	8 0.13%	2 0.05%	3 0.04%
190. EFFECT	14	5 0.17%	1 0.02%	2 0.14%	6 0.06%
191. INTOXICATION	14	2 0.07%	1 0.02%	6 0.14%	5 0.06%
192. MEAN	14	1 0.13%	5 0.08%	4 0.09%	4 0.05%
193. SAME	14	1 0.13%	4 0.06%	3 0.07%	6 0.07%
194. SHOULD	14	1 0.13%	0 0%	0 0%	13 0.16%
195. SHOWED	14	5 0.17%	1 0.02%	1 0.02%	7 0.09%
196. THROUGH	14	1 0.13%	9 0.14%	1 0.02%	3 0.04%
197. TREATED	14	0 0%	9 0.14%	1 0.02%	4 0.05%
198. TYPES	14	0 0%	5 0.08%	2 0.05%	7 0.09%
199. WHILE	14	0 0%	0 0%	8 0.18%	6 0.07%
200. CONTROLLED	13	6 0.21%	1 0.02%	1 0.02%	5 0.06%
201. DRUGS	13	3 0.10%	5 0.08%	0 0%	5 0.06%
202. EVIDENCE	13	5 0.17%	1 0.02%	0 0%	7 0.09%
203. FOUR	13	0 0%	4 0.06%	7 0.16%	2 0.02%
204. HABITUAL	13	1 0.03%	5 0.08%	1 0.02%	6 0.07%
205. IMPORTANT	13	2 0.07%	0 0%	1 0.02%	10 0.12%
206. INFANTS	13	1 0.03%	4 0.06%	2 0.05%	6 0.07%
207. MADE	13	0 0%	4 0.06%	2 0.05%	7 0.09%
208. OBTAINED	13	0 0%	6 0.09%	6 0.14%	1 0.01%
209. OVERALL	13	2 0.07%	2 0.03%	4 0.09%	5 0.06%
210. PART	13	3 0.10%	5 0.08%	3 0.07%	2 0.02%
211. SEEN	13	2 0.07%	5 0.08%	3 0.07%	3 0.04%
212. SHOWS	13	0 0%	0 0%	11 0.25%	2 0.02%

**A Statistical Analysis of the Vocabulary of Medical Research Articles (1)**

	Freq.	I	M	R	D
213. USING	13	3 0.10%	5 0.08%	1 0.02%	4 0.05%
214. ACCURACY	12	4 0.14%	1 0.02%	2 0.05%	5 0.06%
215. AREA	12	2 0.07%	6 0.09%	1 0.02%	3 0.04%
216. AREAS	12	2 0.07%	4 0.06%	0 0%	6 0.07%
217. AVOIDED	12	0 0%	6 0.09%	0 0%	6 0.07%
218. BEING	12	1 0.03%	2 0.03%	6 0.14%	3 0.04%
219. CLINICALLY	12	2 0.07%	1 0.02%	5 0.11%	4 0.05%
220. COMPLICATIONS	12	3 0.10%	0 0%	3 0.07%	6 0.07%
221. EARLY	12	1 0.03%	0 0%	2 0.05%	9 0.11%
222. EITHER	12	1 0.03%	3 0.05%	3 0.07%	5 0.06%
223. EXAMINATION	12	1 0.03%	6 0.09%	0 0%	5 0.06%
224. FOLLOWING	12	2 0.07%	5 0.08%	4 0.09%	1 0.01%
225. GIVEN	12	1 0.03%	9 0.14%	1 0.02%	1 0.01%
226. HEALTH	12	4 0.14%	6 0.09%	0 0%	2 0.02%
227. INCIDENCE	12	0 0%	0 0%	5 0.11%	7 0.09%
228. INFECTION	12	3 0.10%	3 0.05%	2 0.05%	4 0.05%
229. LEAST	12	1 0.03%	5 0.08%	1 0.02%	5 0.06%
230. OBSERVED	12	0 0%	1 0.02%	3 0.07%	8 0.10%
231. ONSET	12	1 0.03%	2 0.03%	4 0.09%	5 0.06%
232. OPERATION	12	2 0.07%	7 0.11%	3 0.07%	0 0%
233. OUT	12	1 0.03%	4 0.06%	2 0.05%	5 0.06%
234. PERCENT	12	4 0.14%	0 0%	6 0.14%	2 0.02%
235. PERIOD	12	1 0.03%	5 0.08%	4 0.09%	2 0.02%
236. PLAYED	12	1 0.03%	4 0.06%	4 0.09%	3 0.04%
237. POSSIBLE	12	1 0.03%	6 0.08%	0 0%	6 0.07%
238. REDUCTION	12	4 0.14%	1 0.02%	1 0.02%	6 0.07%
239. REMISSION	12	0 0%	1 0.02%	0 0%	11 0.14%

# A Statistical Analysis of the Vocabulary of Medical Research Articles (1)

	Freq.	I	M	R	D
240. REQUIREMENTS	12	6 0.21%	0 0%	1 0.02%	5 0.06%
241. RESULT	12	1 0.03%	3 0.05%	2 0.05%	6 0.07%
242. STATISTICAL	12	0 0%	7 0.11%	0 0%	1 0.01%
243. TAPE	12	1 0.03%	6 0.09%	2 0.05%	2 0.04%
244. THEREFORE	12	3 0.10%	0 0%	0 0%	9 0.11%
245. VACCINES	12	3 0.10%	4 0.06%	0 0%	5 0.06%
246. YOUNG	12	0 0%	1 0.02%	7 0.16%	4 0.05%
247. ACUTE	11	2 0.07%	7 0.11%	1 0.02%	1 0.01%
248. ADMISSION	11	0 0%	2 0.03%	7 0.16%	2 0.02%
249. ADULTS	11	0 0%	5 0.08%	0 0%	6 0.07%
250. ANTIBODIES	11	1 0.03%	6 0.09%	0 0%	4 0.05%
251. BOWEL	11	3 0.10%	2 0.03%	2 0.05%	4 0.05%
252. CHILDREN'S	11	2 0.07%	6 0.09%	0 0%	3 0.04%
253. CIMEDITINE	11	3 0.10%	4 0.06%	3 0.07%	1 0.01%
254. COLON	11	3 0.10%	0 0%	3 0.07%	5 0.06%
255. COURSE	11	2 0.07%	3 0.05%	1 0.02%	5 0.06%
256. DETECTED	11	1 0.03%	1 0.02%	4 0.09%	5 0.06%
257. DIFFERENCE	11	0 0%	1 0.02%	3 0.07%	7 0.09%
258. HYPERTENSION	11	3 0.10%	5 0.08%	0 0%	3 0.04%
259. MEDICAL	11	2 0.07%	4 0.06%	2 0.05%	3 0.04%
260. MISSED	11	1 0.03%	1 0.02%	4 0.09%	5 0.06%
261. PAIN	11	1 0.03%	3 0.05%	2 0.05%	5 0.06%
262. QUESTIONS	11	0 0%	9 0.14%	2 0.05%	0 0%
263. RANGE	11	0 0%	5 0.08%	4 0.09%	2 0.02%
264. REFERRED	11	0 0%	3 0.05%	0 0%	8 0.10%
265. REPORTS	11	7 0.24%	1 0.02%	0 0%	3 0.04%
266. ROUTINE	11	3 0.10%	6 0.09%	0 0%	2 0.02%

**A Statistical Analysis of the Vocabulary of Medical Research Articles (1)**

	Freq.	I	M	R	D
267. SERIES	11	0 0%	1 0.02%	1 0.02%	9 0.11%
268. SERUM	11	1 0.03%	3 0.05%	1 0.02%	6 0.07%
269. SHOWN	11	0 0%	4 0.06%	0 0%	7 0.09%
270. SIMPLE	11	3 0.10%	2 0.03%	2 0.05%	4 0.05%
271. SIX	11	0 0%	7 0.11%	2 0.05%	2 0.02%
272. SKIN	11	1 0.03%	9 0.14%	0 0%	1 0.01%
273. SOCIAL	11	2 0.07%	4 0.06%	2 0.05%	3 0.04%
274. TEN	11	1 0.03%	3 0.05%	5 0.11%	2 0.02%
275. THERAPY	11	0 0%	8 0.13%	1 0.02%	2 0.02%
276. TOTAL	11	1 0.03%	1 0.02%	6 0.14%	3 0.04%
277. VARIABLES	11	0 0%	9 0.14%	2 0.05%	0 0%
278. WITHIN	11	0 0%	3 0.05%	5 0.11%	3 0.04%
279. ACCIDENT	10	0 0%	2 0.03%	6 0.14%	2 0.02%
280. ALWAYS	10	1 0.03%	3 0.05%	2 0.05%	4 0.05%
281. ASKED	10	0 0%	10 0.16%	0 0%	0 0%
282. ASSOCIATION	10	4 0.14%	0 0%	0 0%	6 0.07%
283. COHORT	10	1 0.03%	1 0.03%	0 0%	8 0.10%
284. COMMON	10	5 0.17%	0 0%	2 0.05%	3 0.04%
285. COUNTIES	10	1 0.03%	7 0.11%	1 0.02%	1 0.01%
286. DIFFERENT	10	1 0.03%	6 0.09%	0 0%	3 0.04%
287. DONE	10	1 0.03%	8 0.13%	1 0.02%	0 0%
288. DRINKING	10	0 0%	0 0%	8 0.18%	2 0.02%
289. DRUG	10	0 0%	6 0.09%	2 0.05%	2 0.02%
290. EXAMINED	10	4 0.14%	3 0.05%	3 0.07%	0 0%
291. EXPECTED	10	0 0%	1 0.02%	3 0.07%	6 0.07%
292. FOODS	10	1 0.03%	6 0.09%	0 0%	3 0.04%
293. HIS	10	1 0.03%	2 0.03%	2 0.05%	5 0.06%

# A Statistical Analysis of the Vocabulary of Medical Research Articles (1)

	Freq.	I	M	R	D
294. IDENTIFIED	10	1 0.03%	5 0.08%	2 0.05%	2 0.02%
295. INJURIES	10	1 0.03%	1 0.02%	6 0.14%	2 0.02%
296. MACROPHAGES	10	4 0.14%	1 0.02%	5 0.11%	0 0%
297. MORPHINE	10	0 0%	3 0.05%	4 0.09%	3 0.04%
298. NORMAL	10	2 0.07%	1 0.02%	4 0.09%	3 0.04%
299. NORTH	10	1 0.03%	4 0.06%	2 0.05%	3 0.04%
300. REACTIONS	10	2 0.07%	4 0.06%	0 0%	4 0.05%
301. REGION	10	1 0.03%	5 0.08%	1 0.02%	3 0.04%
302. RELAPSES	10	0 0%	2 0.03%	3 0.07%	5 0.06%
303. REMAINING	10	1 0.03%	4 0.06%	3 0.07%	2 0.02%
304. SAMPLE	10	0 0%	4 0.06%	1 0.02%	5 0.06%
305. SECOND	10	1 0.03%	3 0.05%	0 0%	6 0.07%
306. SLEEP	10	3 0.10%	0 0%	0 0%	7 0.09%
307. STATE	10	0 0%	8 0.13%	2 0.05%	2 0.02%
308. SUSPECTED	10	1 0.03%	2 0.03%	6 0.14%	1 0.01%
309. SYSTEM	10	2 0.07%	5 0.08%	1 0.02%	2 0.02%
310. TRAFFIC	10	0 0%	0 0%	2 0.05%	8 0.10%
311. TWINS	10	0 0%	1 0.02%	0 0%	9 0.11%
312. TYPHI	10	6 0.21%	4 0.06%	0 0%	0 0%
313. VALUE	10	2 0.07%	3 0.05%	1 0.02%	4 0.05%
314. WARDS	10	0 0%	9 0.14%	0 0%	1 0.01%