



Short communication

The impact on database searching arising from inconsistency in the nomenclature of parasitic diseases

Tibor Kassai^{1,*}

*Department of Parasitology and Zoology, Szent István University, Faculty of Veterinary Science,
1078 Budapest, István u. 2, Hungary*

Received 6 June 2005; accepted 21 February 2006

Abstract

The existing usage of disease names formed from the name of the parasite taxon is characterised by marked heterogeneity. This is largely due to the fact that, for coining disease names, four different suffixes, ‘-osis’, ‘-iosis’, ‘-asis’ and ‘-iasis’, are being used inconsistently. The result is that alternative terms are in use for naming the same disease, e.g. trypanosomosis and trypanosomiasis, fasciolosis and fascioliasis, ascariosis and ascariasis. In spite of the SNOAPAD/SNOPAD guideline (1988) which proposed the principles of a uniform and standard disease nomenclature, the actual usage depends largely on tradition, educational imprinting and personal preferences, showing great variation. By using alternative disease names as search terms the author investigates in four databases the impact of nomenclatural heterogeneity on information storage and retrieval. It is evident that the existence of alternative disease names in parasitology markedly interferes with the efficacy of online data retrieval. The value of a disease name as a search term was shown to be greatly different in various databases. Until we have to coexist with an inconsistent disease terminology we need to adopt specially structured database-search techniques to ensure a proper level of precision in searching. Such possible techniques are considered.

© 2006 Elsevier B.V. All rights reserved.

Keywords: Nomenclature; Parasitic diseases; Database searching; SNOAPAD/SNOPAD

1. Introduction

The evolutionary process, leading to the inconsistency which regrettably characterises the existing usage of nomenclature of parasitic diseases or

infections, is outlined in my recent review paper (Kassai, 2006). Although a proposal for establishing defined principles of a standardized disease nomenclature was put forward (Kassai et al., 1988; Kassai and Burt, 1994), the parasitologists’ community continues to use alternative or synonymous terms as disease names ad libitum. It was anticipated that the diverse use of terminology in primary documents poses problems in database online searching.

* Tel.: + 36 13559149; fax: +36 13559149.

E-mail address: tkassai@freemail.hu.

¹ Present address: 1125 Budapest, Rózse u. 19, Hungary.

2. Materials, methods and results

Data presented in Table 1 highlight the problem of data retrieval arising from the inconsistency in disease nomenclature. By using different search methods, the frequency of disease names ending in ‘-osis’ and ‘-iasis’ in the period 1 January, 1999–12 December, 2003 was compared in four databases concerned in handling data related to parasitology. Titles, abstracts and keywords were searched jointly.

The success of information retrieval is uncertain if disease names with diverse spellings are available as search terms. In this case web searches based on one version of spelling usually do not lead to sites using an other version. All possible synonyms should be considered if you want to minimise the possibility of losing relevant documents. For optimal retrieval, ambiguous search terms should be combined with the uppercase search operator OR, e.g. (fasciolosis OR fascioliasis). This is true when doing online search in some databases (e.g. Elsevier Science Direct, ISI Web of Science, CAB Abstracts). In PubMed, however, the automatic term mapping and translation system

provides a consistent way to retrieve information that may use different terminology for the same concept. The computer uses an algorithm for scoring the relatedness of two phrases by splitting them to three letter triplets or tri-gram units. This way, if one version of spelling is used as search term (such as fasciolosis) records with other spellings (such as fascioliasis) will automatically be retrieved (Table 1).

As regards disease names derived from the generic name *Trichinella*, the source of variation is not only in the ending of terms but also in the stem formation. Thus, trichinosis is an alternative term for trichinellosis. Surprisingly however, two other terms, trichinelliasis and trichiniasis, were also introduced yielding 393 records each (Table 1). This needlessly adds to the already existing inconsistency.

It is worth noting that in database records the English translation of non-English original titles does not necessarily reflect the preferred usage of the original author. The language of the author is ‘translated’ to the language of the database. It occurs that under the descriptor (DE) subject field of an abstract, the ‘-osis’ terms of the original paper are

Table 1

Comparison of frequency of disease names ending in -osis and -iasis by using different search methods in four databases in the period 1999–2003

Search terms and search methods used	Number of records in different databases			
	Elsevier Science Direct ^a	ISI Web of Science	CAB Abstracts	PubMed
Leishmaniosis OR leishmaniasis	456	500+	2110	2565
Leishmaniosis	109	75	60	2250
Leishmaniasis	421	500+	2096	2550
Leishmaniosis NOT leishmaniasis	35	36	14	15
Leishmaniasis NOT leishmaniosis	418	500+	2050	315
Fasciolosis OR fascioliasis	56	222	419	364
Fasciolosis	39	99	94	341
Fascioliasis	20	142	392	355
Fasciolosis NOT fascioliasis	36	80	27	9
Fascioliasis NOT fasciolosis	17	123	325	23
Trichinellosis OR trichinosis	43	228	419	428
Trichinellosis	34	147	182	412
Trichinosis	10	141	398	410
Trichinosis NOT trichinellosis	9	81	237	16
Trichinellosis NOT trichinosis	33	87	21	18
Trichinelliasis	0	0	1	393
Trichiniasis	0	1	1	393

Note the marked differences in the distribution of hits in Science Direct, Web of Science and CAB versus PubMed when ‘leishmaniosis’, ‘fasciolosis’ and ‘trichinellosis’ are used as search terms. This is largely due to the PubMed’s automatic term mapping and translation system which presents an attempt to recognise the synonymy of disease names with diverse spellings leading to larger yield of records. Titles, abstracts and keywords were searched

^a Except Abstract Database.

replaced by disease names ending in ‘-iasis’, in accordance with the preferred terminology adopted in the thesaurus, under the false belief that the ‘-iasis’ term is the correct equivalent of the original ‘-osis’ term. Thus, it may occur, that in the record of a single paper two sets of different disease names are recorded, and in the descriptor field the ‘-iasis’ terms only are listed if they correspond to the preferred terminology of the database. In handling the records on bacterial diseases such ‘translation’ or ‘transliteration’ does not occur thanks to the uniform nomenclature adopted by our bacteriologist colleagues.

The magnitude of possible variations in the results of data retrieval, in different databases arising from the use of separate search terms with diverse spellings, is illustrated in Table 2. Uniformly high yields of records were achieved in all databases (from 93 to 99%) with the single search term ‘leishmaniasis’ because the use of that term is highly dominant in the literature over its alternative ‘leishmaniosis’. However, in the case of all the three parasitoses it is also seen that the yield of records is greatly dependent on the search term used, at least in three databases. For example, if ‘fasciolosis’ alone is used as a search term, recoveries from 24% (CAB Abstracts) to 95% (PubMed) of the total numbers of relevant records will be obtained. Should we use ‘fascioliasis’ alone as search term, the yield will vary from 37% (Science Direct) to 98% (PubMed). This implies that when online search is based on one version only, of the disease names with diverse spellings, part of the relevant records remains hidden! It is also seen that the usefulness of the same disease name as a search term may be greatly different in various databases.

Database users are not necessarily aware of all the variations of terms available for naming a particular disease or infection. Therefore, when generic derivative disease names are used as search terms, the

adoption of one or a combination of several of the following search options is needed.

- For optimal retrieval, the search should be directed to the fields including jointly the titles, abstracts and keywords or medical subject headings (MeSH).
- For a single search query it is recommended that the alternative search terms combined with the operator OR, and possibly also by AND, e.g. enter (trichinosis OR trichinellosis) AND swine, be used.
- Truncation is the mode of search on the root word. It is a search used for all the words starting with a row of characters, e.g. enter trichin* for trichinosis and trichinellosis as well as for many other aspects related to *Trichinella* spp. In this case irrelevant records may appear in disturbingly large numbers, and refining the search is needed.
- Use wildcards in the middle of terms to reach for terms that have alternate spellings. The asterisk (*) wildcard represents zero to multiple characters, e.g. enter trypanosom*sis for trypanosomosis and trypanosomiasis. The other wildcard is the question mark (?) which stands for a single character, e.g. enter giardi?sis for giardiosis and giardiasis. The dollar sign (\$) stands for a single or no character, e.g. enter trichin\$\$\$osis for finding both trichinosis and trichinellosis. Note that the meaning of wildcards may vary between individual databases.
- A search can also be based on using the Latin name of the parasite (genus or genus and species, e.g. *Trichinella* or *Trichinella spiralis*), and then limit data fields according to one’s interest. This is perhaps the best way to focus your search on items that deal in depth with the topic you are looking for.

Such a structured search should retrieve all records of interest. For building up proper search strategies it

Table 2

Comparison of percentage distribution of records yielded by using different search terms ending in ‘-osis’ or ‘-iasis’ in four databases in the period 1999–2003 (based on data presented in Table 1)

Databases	Percentage distribution of records yielded by using different search terms					
	Leishmaniosis	Leishmaniasis	Fasciolosis	Fascioliasis	Trichinellosis	Trichinosis
Science Direct	8	93	68	37	80	23
Web of Science	3	98	45	64	64	62
CAB Abstracts	3	99	24	93	49	93
PubMed	89	99	95	98	97	96

The number of records obtained by using the search formula (-osis OR -iasis), e.g. (leishmaniosis OR leishmaniasis), was considered to be 100%.

is strongly advised to consult the related manuals and the guidelines for users of individual databases.

The software of databases needs regular updating, sometimes even developments to adapt tools, such as consistent controlled vocabulary of disease names, favouring optimal recovery of relevant data for the users. On the other hand, those concerned in parasitology must keep in mind that, we – when using online data-retrieval services – need to adopt carefully structured database search techniques for ensuring the proper level of precision in searching.

Acknowledgements

The author is most grateful to Zsolt Abonyi (Department of Biomathematics, Szent István Uni-

versity, Faculty of Veterinary Science, Budapest, Hungary) and Éva Orbán (Veterinary Science Library, Szent István University, Budapest, Hungary) for their helpful suggestions and critical reading of the manuscript.

References

- Kassai, T., 2006. Nomenclature for parasitic diseases: cohabitation with inconsistency for how long and why? *Vet. Parasitol.* 138, 169–178.
- Kassai, T., Burt, M.D.B., 1994. A plea for consistency. *Parasitol. Today* 10, 127–128.
- Kassai, T., Cordero del Campillo, M., Euzeby, J., Gaafar, S., Hiepe, Th., Himonas, C.A., 1988. Standardized nomenclature of animal parasitic diseases (SNOAPAD). *Vet. Parasitol.* 29, 299–326.