Bahrain Medical Bulletin, Volume 30, No 3, September2008

Editorial

Cracking Up in the Search for Randomised Trials

Mike Clarke, BA, DPhil*, Liz MacKinnon BA**, Anne Eisinga BA Comb, MSc ***

Searches of electronic bibliographic databases are a key to finding articles in the healthcare literature. If records in these databases are incorrect because of spelling mistakes or transcription errors, users might fail to find them. We did a study to identify records in MEDLINE and EMBASE in which the word random (or its derivatives) had "cracked up" in the title or abstract in the database. This cracking up could include the introduction of spaces within the word, or the break-up of individual letters, such as m into rn. We found more than 60 occurrences in which the word, or letters within it, had cracked apart. Although we focused on records relating to reports containing the word random because of our particular interest in randomised trials, we expect that our findings highlight a more general problem with some records in the electronic bibliographic databases. This has implications for people using free text searching of these databases and for the quality control processes within them.

Indexed bibliographic databases of the healthcare literature provide the tools for most people trying to find articles of potential interest to them. However, the success of these searches depends on the inclusion criteria for the database being searched, the quality of the records within the database, and the quality of the search. If free text searching is done, misspellings in the records in the database or within the search may mean that the user does not find what they are looking for.

Ray and Vermeulen previously showed how misspellings of ten commonly used medical terms might mean that some relevant records would not be found in MEDLINE[1]. We subsequently showed how similar problems might occur with the word "random" and its derivatives (eg "randomisation"), with words such as "radnom" appearing in some abstracts[2]. Although the computerised search might fail to find such records because of the misspelling, a person reading the abstract might not be misled because of the ability of the human brain to overcome spelling mistakes and to interpret the correct meaning of the misspelt word. However, problems will arise if they never see the abstract because their search failed to find it.

---

* Director, UK Cochrane Centre
National Institute for Health Research Oxford, UK
** Senior Research Assistant, Clinical Trial Service Unit
University of Oxford, UK
*** Information Specialist, UK Cochrane Centre
National Institute for Health Research, Oxford, UK

Earlier this year, during the UK Cochrane Centre's systematic search of the database, EMBASE, for records that might relate to reports of trials to be included in the Cochrane Central Register of Controlled Trials some records were found in which words or letters in the abstract had cracked apart[3]. One record was found in which the m in randomised had become "rn", making the word "randornised" and, in another record, the word "random" had broken into "rand om". We sought to examine how widespread these mistakes are within EMBASE and MEDLINE.

One of the authors (LM) searched MEDLINE and EMBASE (from 1980 onwards) for the following character strings: randorn$, r andom$, ra ndom$, ran dom$, rand om$, rando m, rando mly, rando mi$, random ly, random ise$, random ize$, randomi se$, randomi ze$, randomis ed, randomiz ed, random isa$, random iza$, randomi sa$, randomi za$, randomis at$, randomiz at$, ranclom$.

These searches were done in OVID on July 17 2007. She did not restrict the searches by language of publication. She counted the number of records found with each search term in each database and examined whether any of the erroneous records were in both databases.

There are 64 records in which "random" and its derivatives have cracked up in MEDLINE (7 records) and EMBASE (57). These were all unique records with no duplication between the databases. Most of the examples arose because of a break within the word itself, but two (both in EMBASE) were caused by the letter m breaking into rn. The commonest problem was a break after the first three letters (ran), with 2 such errors in MEDLINE and 18 in EMBASE. In 9 records (MEDLINE: 5, EMBASE: 4) the break in the word was filled with a hyphen. In one case (EMBASE) it was filled by a hyphen and a space. We also found that the search for "random ly" found 5 records in MEDLINE and the same 5 in EMBASE due to the phrase "random(ly) amplified polymorphic DNA", which we did not count in the 64 errors.

Although we focused on records relating to the word random and its derivatives in MEDLINE and EMBASE, we expect that our findings highlight a more general problem with some records in electronic bibliographic databases. We do not believe that there will be anything unique or special about words relating to random events, random sampling or randomisation. We expect that there will be many other examples within these databases of words that have cracked up.

Our findings have implications for people using free text searching of these databases and for the quality control processes within these databases. As with misspellings which lead to words such as "radnom" we expect that a person reading the abstract would not be misled by character strings such as "rand om" and they might not even notice that "randomised" had become "randornised"[2]. However, problems will arise if the searcher never sees the abstract because their electronic search failed to find it.

People searching databases might wish to consider expanding their search strategies to include cracked up versions of the words they are searching for. We are aware that this

type of approach has been taken with common misspellings of the word “random” by, among others, the Cochrane Stroke Group[4]. In addition, the creators of electronic databases might wish to examine their procedures for quality assurance, especially when using character recognition software to input the abstracts[5].

Contribution of Authors

Mike Clarke conceived of this idea, drafted the manuscript and agreed the final version. Liz MacKinnon refined the idea, conducted the searches of MEDLINE and EMBASE, revised the manuscript and agreed the final version. Anne Eisinga refined the idea, revised the manuscript and agreed the final version.

REFERENCES

1. Ray JG, Vermeulen MJ. Mizspellin and Medicine. *BMJ* 1996; 313: 1658-9.
2. Clarke M, Greaves L, James S. MeSH terms must be used in MEDLINE searches. BMJ 1997; 314: 1203.
3. Eisinga A, Lefebvre C. Closing the Gap: Identifying Reports of Randomized Trials in EMBASE for Inclusion in the Cochrane Central Register of Controlled Trials (CENTRAL) [abstract]. 12th Cochrane Colloquium; 2004 Oct 2-6; Ottawa, Canada: 150-1.
4. Sandercock P, Algra A, Anderson C, et al. Cochrane Stroke Group. About The Cochrane Collaboration (Cochrane Review Groups (CRGs)) 2007; 4. Art. No: STROKE.
5. Morant A. How British Users Stay Ahead with ISDN - Integrated Services Digital Networks. http://findarticles.com/p/articles/mi_m0CMN/is_n8_v30/ai_14191165 (accessed 9 Nov 2007).

Correspondence address
Professor Mike Clarke
School of Nursing and Midwifery
Trinity College Dublin
24 D’Olier Street
Dublin 2
Ireland