



Nominal ISOMERs (Incorrect Spellings Of Medicines Eluding Researchers)—variants in the spellings of drug names in PubMed: a database review

Robin E Ferner,^{1,2} Jeffrey K Aronson³

¹West Midlands Centre for Adverse Drug Reactions, City Hospital, Birmingham B18 7QH, UK

²Institute of Clinical Science, University of Birmingham, Birmingham, UK

³Centre for Evidence Based Medicine, Nuffield Department of Primary Care Health Sciences, University of Oxford, Oxford, UK

Correspondence to: R E Ferner
r.e.ferner@bham.ac.uk

Cite this as: *BMJ* 2016;355:i4854
<http://dx.doi.org/10.1136/bmj.i4854>

Accepted: 18 July 2016

ABSTRACT

OBJECTIVE

To examine how misspellings of drug names could impede searches for published literature.

DESIGN

Database review.

DATA SOURCE

PubMed.

REVIEW METHODS

The study included 30 drug names that are commonly misspelt on prescription charts in hospitals in Birmingham, UK (test set), and 30 control names randomly chosen from a hospital formulary (control set). The following definitions were used: *standard names*—the international non-proprietary names, *variant names*—deviations in spelling from standard names that are not themselves standard names in English language nomenclature, and *hidden reference variants*—variant spellings that identified publications in textword (tw) searches of PubMed or other databases, and which were not identified by textword searches for the standard names. Variant names were generated from standard names by applying letter substitutions, omissions, additions, transpositions, duplications, deduplications, and combinations of these. Searches were carried out in PubMed (30 June 2016) for “standard name[tw]” and “variant name[tw] NOT standard name[tw].”

RESULTS

The 30 standard names of drugs in the test set gave 325 979 hits in total, and 160 hidden reference variants gave 3872 hits (1.17%). The standard names of the control set gave 470 064 hits, and 79 hidden reference variants gave 766 hits (0.16%). Letter substitutions (particularly i to y and vice versa) and omissions

together accounted for 2924 (74%) of the variants. Amitriptyline (8530 hits) yielded 18 hidden reference variants (179 (2.1%) hits). Names ending in “in,” “ine,” or “micin” were commonly misspelt. Failing to search for hidden reference variants of “gentamicin,” “amitriptyline,” “mirtazapine,” and “trazodone” would miss at least 19 systematic reviews. A hidden reference variant related to Christmas, “No-el”, was rare; variants of “X-miss” were rarer.

CONCLUSION

When performing searches, researchers should include misspellings of drug names among their search terms.

Introduction

Variant spellings of drug names can cause confusion, which could lead to serious harm.^{1,2} Nevertheless, these names are expected to be correctly spelled and indexed in published work. We have tested this assumption, which underlies many search strategies for systematic reviews and meta-analyses of therapeutic interventions.

Methods

We defined the following types of drug names:

- Standard name: the international non-proprietary name (INN)³ or (if there was no INN) the British Approved Name (BAN; box 1).
- Variant name: any deviation in spelling from the standard name that was not itself a standard name in English language nomenclatures, such as BANs or US Adopted Names (USANs). For example, we did not regard thimerosal (USAN) as a transpositional variant of thiomersal (INN), although many papers would be missed by not searching for both.
- Hidden reference variant: a variant spelling that, when used as a textword search term in PubMed and other databases, identified publications that were not identified by searching for the standard name as a textword.

Senior pharmacists from hospitals in Birmingham, UK provided 30 examples of drug names that were commonly misspelt on hospital prescription charts. We then chose a control set of 30 drugs at random from the Sandwell and West Birmingham Hospitals NHS Trust formulary. We ran a search in PubMed⁴ on 30 June 2016 for textword instances of the standard name of each drug and for spelling variants created by the following types of changes:

- Substitutions (eg, i to y and vice versa; one unaccented vowel to another vowel or y; soft c to s and vice versa; hard c to k and vice versa; ch to k; f to ph

WHAT IS ALREADY KNOWN ON THIS TOPIC

Spelling errors are not uncommon in databases such as PubMed and Medline
Drug names are frequently misspelt in these databases and in hospital prescription charts

WHAT THIS STUDY ADDS

Database searches using only drug names spelt correctly will miss relevant references in which the names are spelt incorrectly
These references, which include systematic reviews, will remain hidden unless searches are also undertaken using possible misspellings
Authors and editors should be more vigilant about spelling drug names correctly, and indexers of databases such as PubMed should cross index incorrect spelling variants to correctly spelt names in both directions
When performing searches involving drug names, researchers should include incorrect spellings among their search terms

Box 1: Some national drug naming systems

- A panel of international nomenclature experts assigns recommended international non-proprietary names (rINNs) to drugs, under the aegis of the World Health Organization.
- Occasionally, an objection is raised to a name. If agreement cannot be reached, the name remains a proposed INN (pINN). Nearly 5% of all INNs are pINNs. For example, amantadine was proposed in 1965, but it has not become a rINN because an objection remains on file.
- The best known national drug naming systems are the British Approved Name (BAN), dénomination commune française (DCF), Japanese Accepted Name for pharmaceuticals (JAN), and US Adopted Name (USAN).
- The UK uses the INN as the BAN, except for adrenaline and noradrenaline (INNs epinephrine and norepinephrine). That is not the case elsewhere. For example, compare paracetamol (INN) and acetaminophen (USAN); salbutamol (INN) and albuterol (USAN); rifampicin (INN) and rifampin (USAN); glibenclamide (INN), and glyburide (USAN).
- Some compounds that do not have INNs can still have a BAN, using a chemical name—for example, acetylsalicylic acid and glyceryl trinitrate.
- Mixtures of drugs do not have INNs. In some cases, BANs have been specially created for such mixtures (eg, co-codamol is the BAN for a mixture of codeine and paracetamol).

and vice versa; ide to ine and vice versa; m to n and vice versa; th to t; x to ks)

- Omissions (eg, prednisolone to pednisolone; propranolol to popranolol or propanolol; omission of final e)
- Additions (eg, cotrimoxazole/clotrimazole to clotrimoxazole; addition of final e)
- Transpositions (eg, furosemide to fruosemide; filgrastim to filgastrim)
- Duplications and deduplications (eg, l to ll and vice versa; n to nn and vice versa)
- Combination of changes (eg, gentamicin to gentamycine; amitriptyline to amytriptilin).

We searched for “standard name[tw]” (where tw=textword) and noted the number of hits. We then searched for “variant spelling[tw] NOT standard name[tw]” and added together the number of hits for each name over all its variant spellings. We thus determined the number of hits that would have been missed by searching only for the standard name. We checked whether the retrieved references were systematic reviews, including meta-analyses.

Results

Numbers of hits in PubMed after use of standard names and hidden reference variant spellings

Standard names of the test set of 30 drugs gave 325 979 hits; 160 hidden reference variants produced 3872 hits (1.19%; range 0-2068; median 49). Standard names of the control set gave 470 064 hits. Of 208 possible hidden reference variants, we found 79, which gave 766 hits (0.16%; range 0-115; median 16). Amitriptyline (8530 hits) had 18 hidden reference variant spellings (179 hits; 2.06%), the most variant names for a single standard name (tables 1 and 2).

Types of variant

Table 3 shows frequencies of the different types of spelling variants.

Table 1 | Hidden reference variants and number of PubMed hits of the standard name amitriptyline

Variant name	No of hits
Amitriptyline	1
Amitriptilin	8
Amitriptiline	8
Amitriptyine	1
Amitriptylin	14
Amitriptyllin	1
Amitriptylline	2
Amitriptylin	4
Amitriptyline	80
Amitriptylline	1
Amitriptylin	1
Amitriptyline	32
Amitriptylline	1
Amytriptilin	2
Amytriptiline	10
Amytriptylin	1
Amytriptyline	10
Amytriptiline	2

Table 2 | Numbers of PubMed hits of hidden reference variant spellings* of the standard name amitriptyline

Variant spellings (i to y substitutions and vice versa)	No of hits			
	Spelling with single l		Spelling with double ll	
	Without final e	Plus final e	Without final e	Plus final e
i-i-i	8	8	0	0
i-i-y	14	†	1	2
i-y-i	4	80	0	1
i-y-y	1	32	0	1
y-i-i	2	10	0	0
y-i-y	1	10	0	0
y-y-i	0	2	0	0
y-y-y	0	0	0	0
Total	30	142	1	4

*Included one case of the variant “amitriptyline” lacking the r, and one of “amitriptyine” lacking the l. There was no variant in which the final i was replaced by a y.

†Amitriptyline, the standard name, gave 8530 hits.

Table 3 | Frequencies of 239 different types of hidden reference variants of 60 drug names (test and control sets combined). We found no variants of aprepitant

Single changes (n=207) and combinations (n=31)	Frequencies (No (%))
Substitutions (total)	108 (45)
Single	91
Double	16
Treble	1*
Omissions	67 (28)
Additions	20 (8)
Transpositions	11 (5)
Duplications	2 (1)
Deduplications	0
Any combination of these aberrations	31 (13)

*Amytriptiline (two hits).

We examined names ending in “micin” in detail. Most of the errors occurred with the standard form “gentamicin” compared with the variant “gentamycin,” which resulted in 21 384 and 1977 hits (9.25%), respectively. The ending “mycin” was also often substituted in fidaxomicin (2.02%) and netilmicin (2.46%; table 4). In contrast, in 19 standard drug names ending in “mycin” (218 415 hits), the hidden reference variant “micin” was rare (157 hits (0.07%); table 5).

Names ending in “in” or “ine” were also likely to generate hidden spelling variants by addition or omission of the final “e.” The 28 standard names of this type in the test and control sets combined yielded 296 973 hits and hidden spelling variants yielded 3450 hits (1.16%), compared with 499 070 hits and 1188 hits (0.24%), respectively, for the other 32 names.

Searches for systematic reviews

We found 87 systematic reviews or meta-analyses that mentioned the standard name gentamicin, 0.41% of all hits for “gentamicin[tw].” We found six further systematic reviews (6.5% of the total) in PubMed after searching for hidden reference variants of gentamicin. In Medline, the equivalent search for “gentamicin.af.”

(where af=all fields) identified 141 systematic reviews, with 19 782 hits (0.71%). The hidden reference variants, with 863 hits, identified 15 additional systematic reviews (9.6% of the total).

Similarly, for amitriptyline, we found 179 systematic reviews in PubMed and another five as hidden reference variants. Corresponding numbers were 110 and six for mirtazapine and 47 and two for trazodone. Thus, for these drugs, 19 systematic reviews of 455 (4.2%) would have been missed by searching for the standard spellings only.

A variant index score

We scored various features of standard names as follows:

- Number of letters
- Number of syllables
- Number of unaccented vowels + 1
- Number of incidences of i and y + 1
- Number of incidences of f or ph + 1
- Number of potential duplications or deduplications (l, m, n, s, t) + 1
- Standard names ending in “in” or “ine” (no=1, yes=2)
- Standard names ending in “micin” (no=1, yes=2).

The product of these factors, a variant index score, was on average much higher in the test group (range 54-4480; median 524) than the control group (range 36-1440; median 272).

Discussion

We have uncovered a potential indirect harm from incorrect variant spellings of drug names that has not previously been investigated, to our knowledge, although others have reported misspelt general medical textwords in Medline⁸ and misspellings of the word “random” and its derivatives in Medline and EMBASE.⁹ Difficulties in recognising and distinguishing drug names can lead to clinical harm directly, for example, when one drug name is read as another. Here, we demonstrate the extent to which medical literature searches can be frustrated by textword searches that fail to include variant spellings, since articles referenced only by the variant spelling will remain hidden. PubMed offers the correct spelling (eg, gentamicin) when you enter an incorrect one (eg, gentamycin), but not the other way round—searching for “gentamicin[tw]” does not yield incorrect spellings.

Information in systematic reviews can be lost if the review is indexed under a hidden reference variant and not under the textword for the standard name. The problem is not limited to PubMed. In Medline, 13 systematic reviews were hidden under the variant spelling “gentamycin.” In the Cochrane Database of Systematic Reviews,⁵ there were 15 systematic reviews of “gentamicin,” but use of the term “gentamycin” identified four otherwise hidden reviews.

The most obvious way to mitigate this problem is for authors and editors to take care over the correct spellings of drug names. Indexing could be improved, especially by ensuring that standard names are always used

Table 4 | Number of PubMed hits for drugs with standard names ending in “micin” compared with hidden reference variants ending in “mycin”

Standard name (“micin”)	No of hits	Variant (“mycin”)	No of hits	Ratio (%)*
Fidaxomicin	248	Fidaxomycin	5	2.02
Gentamicin	21 384	Gentamycin	1977	9.25
Netilmicin	1875	Netilmycin	46	2.46
Ozogamicin	539	Ozogamycin	13	2.41
Total	24 046	—	2041	8.49

*Number of hits of variant divided by number of hits of standard name.

Table 5 | Number of PubMed hits for drugs with standard names ending in “mycin” compared with hidden reference variants ending in “micin”

Standard name (“mycin”)	No of hits	Variant (“micin”)	No of hits	Ratio (%)*
Azithromycin	6797	Azithromicin	4	0.06
Bleomycin	17 575	Bleomicin	8	0.05
Capreomycin	570	Capreomicin	1	0.18
Clarithromycin	8827	Clarithromicin	2	0.02
Clindamycin	10 308	Clindamicin	17	0.16
[D]actinomycin	33 800	[D]actinomicin	11	0.03
Daptomycin	2394	Daptomicin	1	0.04
Erythromycin	23 605	Erythromicin	13	0.06
Fosfomycin	2501	Fosfomicin	3	0.12
Kanamycin	12 966	Kanamicin	17	0.13
Lincomycin	3114	Lincomicin	2	0.06
Mitomycin	18 503	Mitomicin	15	0.08
Neomycin	12 678	Neomicin	4	0.03
Spectinomycin	2467	Spectinomicin	1	0.04
Spiramycin	1409	Spiramicin	1	0.07
Streptomycin	29 265	Streptomicin	4	0.01
Telithromycin	909	Telithromicin	0	0
Tobramycin	6822	Tobramicin	30	0.44
Vancomycin	23 905	Vancomicin	23	0.10
Total	218 415	—	157	0.07

*Number of hits of variant divided by number of hits of standard name.

Table 6 | “No-el” and “X-miss” hidden reference variants of some drug names, mostly not included in the main study

Standard name	No of PubMed hits	Hidden reference variant	No of PubMed hits
No-el variants			
Acetylcholine	87367	Acetylcholine/acetylchoine	59
Acetylcysteine	15552	Acetycysteine	1
Acidinium	143	Acidinium	2
Adrenaline	18258	Adrenaine	1
Aflibercept	885	Afibercept	1
Agalsidase	389	Agasidase	1
Albendazole	5206	Abendazole	2
Alglucosidase	76	Aglucosidase	1
Allopurinol	9168	Alopurinol	7
Alprazolam	2408	Aprazolam	3
Alteplase	1434	Ateplase	1
Amisulpride	898	Amisupride	2
Amitriptyline	8530	Amitriptyine	1
Carmellose	50	Carmelose	2
Chlorambucil	4744	Chorambucil	1
Chloramphenicol	38566	Choramphenicol	18
Chlorhexidine	9737	Chorhexidine	1
Chloroquine	19061	Choroquine	5
Ending in “cillin” (15 names)	152776	Ending in “cilin” (15 names)	148
Clotrimazole	2637	Cotrimazole	5
Colchicine	18500	Cochicine	9
Cyclophosphamide	62552	Cycophosphamide	2
Desflurane/sevoflurane	9533	Desfurane/sevofurane	3
Diclofenac	10526	Dicofenac	3
Diethylstilbestrol	10000	Diethylstilbestrol/diethylstibestrol	26
Doxycycline*	13592	Doxycycine	1
Enzalutamide	615	Enzautamide	1
Ethinylestradiol	2424	Ethinyestradiol	2
Filgrastim	2220	Figrastim	1
Ending in “floxacin” (5 names)	42376	Ending in “foxacin” (5 names)	13
Ending in “flozin” (3 names)	981	Ending in “fozin” (3 names)	4
Fluorouracil	46951	Fuorouracil	4
Flurbiprofen	2376	Furbiprofen	1
Flutamide	3344	Futamide	1
Follitropin and urofollitropin	668	Folitropin and urofolitropin	11
Glibenclamide	7421	Glibencamide	5
Gliclazide	1138	Glicazide	7
Hydrochlorothiazide	8093	Hydrochorothiazide	2
Hydroxyethylcellulose*	561	Hydroxyethylcelluose/ hydroxyethylcellulose	3
Immunoglobulin	283887	Immunogobulin	4
Infliximab	11000	Infiximab	5
Insulin	351197	Insuin	1
Leflunomide	2002	Lefunomide	3
Melphalan	9699	Mephalan	7
Methyldopa	4560	Methydopa	1
Methylphenidate	7830	Methyphenidate	3
Naltrexone	8669	Natrexone	1
Olmесartan	1321	Omesartan	1
Oxaliplatin	8056	Oxalipatin	1
Penicillamine	10852	Penicilamine	14
Paclitaxel	28899	Pacitaxel	4
Phenylephrine	20623	Phenyephrine	4
Ending in “phylline” (3 names)	5306	Ending in “phyline” (3 names)	12
Propylthiouracil	5103	Propythiouracil	2
Prostaglandin(s)	115753	Prostagandin(s)	4
Repaglinide	676	Repaginide	1
Rosiglitazone	5570	Rosigitazone	1
Sildenafil	6348	Sidenafil	3
Sulfadoxine	2693	Sufadoxine	1
Sulfamethoxazole	15466	Sufamethoxazole	3

(Continued)

Table 6 | “No-el” and “X-miss” hidden reference variants of some drug names, mostly not included in the main study

Standard name	No of PubMed hits	Hidden reference variant	No of PubMed hits
Sulphonamide	8063	Sufonamide	1
Valproic	13 432	Vaproic	1
Zolmitriptan	565	Zomitriptan	3
Total	1 549 326		442 (0.028)
X-miss variants			
Amoxicillin	16 411	Amoicillin	1
Doxycycline*	13 592	Doicycline	1
Hydroxyethylcellulose*	561	Hydroyethylcellulose	1
Oxygen	510 268	Oygen	2
Total	540 832		5 (0.0009%)

*Do[x]cyc[l]ine and hydro[x]yethy[l]cellu[l]ose each feature twice.

when it is possible to identify them. However, even with scrupulous indexing, orthographic variants will pose challenges, because one cannot expect indexers to seek out all variant spellings in a paper for inclusion under a MeSH term heading. Researchers could also search for all likely variants as textwords, although this would pose challenges for names with many potential variants. For example, 18 variants of amitriptyline returned 179 hits that would have been hidden using only the standard name.

Another solution is to use wild cards, if available. Medline allows users to search for words that are spelt with alternative letters. For example, a search for “amitriptyline.af.” yields 8092 hits. Searching for “am#tr#pt#line.af.” uncovers all variant spellings with i to y substitutions (and vice versa) in amitriptyline (table 2), revealing 123 hidden reference variants. The textword “am#tr#pt#l*.af.” truncated at the letter l uncovers variants of the last few letters (for example, ending in “lin,” “line,” “llin,” “lline”) without sacrificing specificity, and gives further hits. However, this does not exhaust all variant forms. For example, the hidden reference variant amitriptyline, generated by omission, was missed.

The variant index score that we have calculated from eight important features associated with hidden reference variants affords insight into the likelihood that newly coined names might prove problematic. Combining the index score with Trigram-2b or the Levenshtein distance, which measure how likely names are to be confused,^{6,7} could help reduce problems with new names.

It has been suggested that all relevant spelling variants should be included in search strategies. However, this recommendation did not refer to incorrectly spelt variants as opposed to variants in standard spelling, such as those between US and UK English (eg, anemia and anaemia), and did not mention drug names.¹⁰

Limitations

Although we systematically generated variants of standard names of drugs (as described in the methods), we could have missed some variants, and underestimated the frequencies of hidden reference variants. In the Xmas spirit, we offer table 6, illustrating other variant spellings.

Note added in proof: Both authors found it hard to proofread an article intended to contain many variant spellings. We apologise if, inadvertently, we have sometimes spelt drug names correctly.

Contributors: The authors contributed equally to the genesis, analysis, and interpretation of the data and the writing of the paper. REF is the guarantor.

Funding: There was no specific funding for this study.

Competing interests: Both authors have completed the ICMJE uniform disclosure form at www.icmje.org/coi_disclosure.pdf and declare: REF and JKA have had their names misspelt from time to time; both have occasionally undertaken research and medicolegal work that has entailed literature searches; both have published articles on nomenclature and definitions. JKA is chairman of the expert advisory group on nomenclature of the British Pharmacopoeia Commission; however, the opinions expressed in this article do not necessarily represent those of that organisation or other members of that group. They declare no other interests.

Ethical approval: Approval was not required.

Data sharing: No additional data available.

The lead author (the manuscript's guarantor) affirms that the manuscript is an honest, accurate, and transparent account of the study being reported; that no important aspects of the study have been omitted; and that any discrepancies from the study as planned (and, if relevant, registered) have been explained.

This is an Open Access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 3.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/3.0/>.

- 1 Aronson JK. Confusion over similar drug names. Problems and solutions. *Drug Saf* 1995;12:155-60. doi:10.2165/00002018-199512030-00001.
- 2 Bryan R, Aronson JK, ten Hacken P, Williams A, Jordan S. Patient safety in medication nomenclature: orthographic and semantic properties of international nonproprietary names. *PLoS One* 2015;10:e0145431. doi:10.1371/journal.pone.0145431.
- 3 World Health Organization. International non-proprietary names. www.who.int/medicines/services/inn/en.
- 4 PubMed. www.ncbi.nlm.nih.gov/pubmed.
- 5 Cochrane Library. Cochrane Database of Systematic Reviews. www.cochranelibrary.com/cochrane-database-of-systematic-reviews.
- 6 Lambert BL, Lin S-J, Chang K-Y, Gandhi SK. Similarity as a risk factor in drug-name confusion errors: the look-alike (orthographic) and sound-alike (phonetic) model. *Med Care* 1999;37:1214-25. doi:10.1097/00005650-199912000-00005.
- 7 Yujian L, Bo L. A normalized Levenshtein distance metric. *IEEE Trans Pattern Anal Mach Intell* 2007;29:1091-5. doi:10.1109/TPAMI.2007.1078.
- 8 Ray JG, Vermeulen MJ. Mizspellin and Medline. *BMJ* 1996;313:1658-9. doi:10.1136/bmj.313.7072.1658.
- 9 Clarke M, MacKinnon L, Eisinga A. Cracking up in the search for randomised trials. *Bahrain Med J* 2008;30:85-6.
- 10 Sampson M, McGowan J, Cogo E, Grimshaw J, Moher D, Lefebvre C. An evidence-based practice guideline for the peer review of electronic search strategies. *J Clin Epidemiol* 2009;62:944-52. doi:10.1016/j.jclinepi.2008.10.012.