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

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 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 nomenclature, such as BANs or US Adopted Names (USANs). For example, we did not regard thiomersal (USAN) as a transpositional variant of thimerosal (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.

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.
- 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; propanolol to popranol 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 duplications (eg, l to ll and vice versa; n to nn and vice versa)
- Combination of changes (eg, gentamicin to gentamycin; amitriptyline to amitriptilin).

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 325979 hits; 160 hidden reference variants produced 3872 hits (1.19%; range 0-2068; median 49). Standard names of the control set gave 470064 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 |
| Amitripyline | 1 |
| Amitrpylin | 14 |
| Amtripylin | 1 |
| Amtripyline | 2 |
| Amitryline | 4 |
| Amtryline | 80 |
| Amtrypitline | 1 |
| Amtryptiline | 1 |
| Amtryptilin | 1 |
| Amtryptyl | 1 |
| Amtryptoline | 10 |
| Amtryptine | 2 |

| Table 2 | Numbers of PubMed hits of hidden reference variant spellings* of the standard name amitriptyline |
|---------|---------------------------------------------------------------|
| Variant spellings | No of hits |
| (l to y substitutions (and vice versa)) | |
| Spelling with single l | Spelling with double l |
| Without final e | Plus final e | Without final e | Plus final e |
| i→i | 8 | 8 | 0 | 0 |
| i→y | 14 | 4 | 0 | 2 |
| i→y | 1 | 1 | 0 | 1 |
| y→i | 2 | 10 | 0 | 0 |
| y→y | 1 | 10 | 0 | 0 |
| y→y | 0 | 2 | 0 | 0 |
| Total | 30 | 142 | 1 | 4 |

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

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

*Amitriptiline (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 (9.2%), respectively. The ending “mycin” was also often substituted in fidaxomycin (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, 0.4% of all hits for “gentamicin[w]”. 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.”

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 (%) |
|--------------------------|------------|--------------------|------------|-----------|
| Fidaxomycin              | 248        | Fidaxomycin        | 5          | 2.02      |
| Gentamicin               | 21 384     | Gentamicin         | 1977       | 9.25      |
| Netilmicin               | 1875       | Netilmicin         | 46         | 2.46      |
| Ozagamicin               | 539        | Ozagamicin         | 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 “micin” compared with hidden reference variants ending in “mycin”

| Standard name ("micin") | No of hits | Variant ("mycin") | No of hits | Ratio (%) |
|--------------------------|------------|--------------------|------------|-----------|
| Azithromycin             | 6797       | Azithromycin       | 4          | 0.06      |
| Bleomycin                | 17 575     | Bleomycin          | 8          | 0.05      |
| Cepromycin               | 570        | Cepromycin         | 1          | 0.18      |
| Clarithromycin           | 8827       | Clarithromycin     | 2          | 0.02      |
| Clindamycin              | 10 308     | Clindamycin        | 17         | 0.16      |
| Dijactinomycin           | 33 800     | Dijactinomycin     | 11         | 0.03      |
| Daptomycin               | 2394       | Daptomycin         | 1          | 0.04      |
| Erythromycin             | 23 605     | Erythromycin       | 13         | 0.06      |
| Fosfomycin               | 2501       | Fosfomycin         | 3          | 0.12      |
| Kanamycin                | 12 966     | Kanamycin          | 17         | 0.13      |
| Lincomycin               | 3114       | Lincomycin         | 2          | 0.06      |
| Mitomycin                | 18 503     | Mitomycin          | 15         | 0.08      |
| Neomycin                 | 12 678     | Neomycin           | 4          | 0.03      |
| Sptacinomycin            | 2467       | Sptacinomycin      | 1          | 0.04      |
| Spiramycin               | 1409       | Spiramycin         | 1          | 0.07      |
| Streptomycin             | 29 265     | Streptomycin       | 4          | 0.01      |
| Telithromycin            | 909        | Telithromycin      | 0          | 0         |
| Tobramycin               | 6822       | Tobramycin         | 30         | 0.44      |
| Vancomycin               | 23 905     | Vancomycin         | 23         | 0.10      |
| Total                    | 218 415     |                    | 157        | 0.07      |

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

The product of these factors, a variant index score, was on average much higher in the test group (range 36–4480; 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 Medline and misspellings of the word “random” and its derivatives in Medline and EMBASE. 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[w]” 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, 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...
| Standard name                  | No of PubMed hits | Hidden reference variant          | No of PubMed hits |
|--------------------------------|-------------------|-----------------------------------|-------------------|
| Acetylcholine                  | 87367             | Acetycholine/acetylchoine         | 59                |
| Acetylcysteine                 | 15552             | Acetylcysteine                    | 1                 |
| Acidinium                      | 143               | Acidinium                         | 2                 |
| Adrenaline                     | 18258             | Adrenaline                        | 1                 |
| Afibercept                     | 885               | Afibercept                        | 1                 |
| Agalsidase                     | 389               | Agalsidase                        | 1                 |
| Albendazole                    | 5206              | Albendazole                       | 2                 |
| Alglucosidase                  | 76                | Alglucosidase                      | 1                 |
| Alopurinol                     | 9168              | Alopurinol                        | 7                 |
| Alprazolam                     | 2408              | Alprazolam                        | 3                 |
| Aleplase                       | 1434              | Aleplase                          | 1                 |
| Amsulpride                     | 698               | Amsulpride                        | 2                 |
| Amitriptyline                  | 8530              | Amitriptyline                     | 1                 |
| Carmellose                     | 50                | Carmellose                        | 2                 |
| Chorambucil                    | 4744              | Chorambucil                       | 1                 |
| Chorphenicol                   | 38566             | Chorphenicol                      | 18                |
| Chlorhexidine                  | 9737              | Chlorhexidine                     | 1                 |
| Choroquine                     | 19061             | Choroquine                        | 5                 |
| Ending in “cilin” (15 names)   | 152776            | Ending in “cilin” (15 names)      | 168               |
| Clotrimazole                   | 2637              | Clotrimazole                      | 5                 |
| Colchicine                     | 18500             | Colchicine                        | 9                 |
| Cyclophosphamide               | 62552             | Cyclophosphamide                  | 2                 |
| Desflurane/sevorflurane        | 9533              | Desflurane/sevorflurane           | 3                 |
| Diclofenac                     | 10526             | Diclofenac                        | 3                 |
| Diethyldistibestrol            | 10000             | Diethyldistibestrol/diethyldistibestrol | 26 |
| Doxycycline*                   | 13592             | Doxycycline                       | 1                 |
| Enzalutamide                   | 615               | Enzalutamide                       | 1                 |
| Ethinyestradiol                | 2424              | Ethinyestradiol                   | 2                 |
| Filgrastim                     | 2220              | Filgrastim                        | 1                 |
| Ending in “foxacin” (5 names)  | 42376             | Ending in “foxacin” (5 names)     | 13                |
| Ending in “fozin” (3 names)    | 981               | Ending in “fozin” (3 names)       | 4                 |
| Fluorouracil                   | 46951             | Fluorouracil                      | 4                 |
| Flurbiprofen                   | 2376              | Flurbiprofen                      | 1                 |
| Flutamide                      | 3344              | Flutamide                         | 1                 |
| Folfotopin and urofolotopin    | 668               | Folfotopin and urofolotopin       | 11                |
| Gilbenclamide                  | 7421              | Gilbenclamide                     | 5                 |
| Gliclazide                     | 1138              | Gliclazide                        | 7                 |
| Hydrochlorothiazide            | 8093              | Hydrochlorothiazide               | 2                 |
| Hydroxyethylcellulose*         | 561               | Hydroxyethylcellulose/ hydroxyethylcellulose | 3 |
| Immunoglobulin                 | 283887            | Immunoglobulin                    | 4                 |
| Infliximab                     | 11000             | Infliximab                        | 5                 |
| Insulin                        | 351397            | Insulin                           | 1                 |
| Lefunomide                     | 2002              | Lefunomide                        | 3                 |
| Melphalan                      | 9699              | Melphalan                         | 7                 |
| Metyldopa                      | 4560              | Metyldopa                         | 1                 |
| Methylphenidate                | 7830              | Methylphenidate                   | 3                 |
| Naltrexone                     | 8669              | Naltrexone                        | 1                 |
| Olmesartan                     | 1321              | Olmesartan                        | 1                 |
| Oxalipatin                     | 8056              | Oxalipatin                        | 1                 |
| Penicillamine                  | 10852             | Penicillamine                     | 14                |
| Paclitaxel                     | 28899             | Paclitaxel                        | 4                 |
| Phenylephrine                  | 20623             | Phenylephrine                     | 4                 |
| Ending in “phyline” (3 names)  | 5306              | Ending in “phyline” (3 names)     | 12                |
| Propylthiouracil               | 5103              | Propylthiouracil                  | 2                 |
| Prostaglandin(s)               | 115753            | Prostaglandin(s)                  | 4                 |
| Repaglinide                    | 676               | Repaglinide                       | 1                 |
| Rosiglitazone                  | 1570              | Rosiglitazone                     | 1                 |
| Sildenafil                     | 6368              | Sildenafil                         | 3                 |
| Sulfadoxine                    | 2693              | Sulfadoxine                       | 1                 |
| Sulfamethoxazole               | 15466             | Sulfamethoxazole                  | 3                 |

(Continued)
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#l*.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 all variant spellings of the last few letters (for example, ending in “ln,” “line,” “llin,” “lline”) without sacrificing specificity, and gives further hits. However, this does not exhaust all variant forms. For example, the hidden reference variant amitipitline, 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 variants affords insight into the likelihood that word variants will 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#l*.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 all variant spellings of the last few letters (for example, ending in “ln,” “line,” “llin,” “lline”) without sacrificing specificity, and gives further hits. However, this does not exhaust all variant forms. For example, the hidden reference variant amitipitline, 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, 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, amenia and anaemia), and did not mention drug names.

**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.

### 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              | Sulphonamide                   | 1                 |
| Valproic             | 13432             | Valproic                       | 1                 |
| Zolmitriptan         | 565               | Zolmitriptan                   | 3                 |
| **Total**            | **1549326**       | **Total**                      | **442 (0.028)**   |

**X-miss variants**

| Amoxicillin          | 16411             | Amoxicillin                    | 1                 |
| Doxycycline*         | 13592             | Doxycycline                    | 1                 |
| Hydroxyethylcellulose* | 561              | Hydroxyethylcellulose          | 1                 |
| Oxygen               | 510268            | Oxygen                         | 2                 |
| **Total**            | **540832**        | **Total**                      | **5 (0.0009%)**   |

*Doxycycline* and hydroxyethylcellulose each feature twice.

---

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 SJ, Chang KY, 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:10.1097/00005630-199912000-00005.
7. Yujian L, Bo L. A normalized Levenshtein distance metric. IEEE Trans Pattern Anal Mach Intell. 2007;29:1091–9. doi:10.1109/TPAMI.2007/1078.
8. Ray JG, Vermeulen MJ, Mizispellin 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. BMJ Med J 2008;330:85-6.
10. Sampson M, McGowan J, Cogo E, Grimshaw J, Mohr D, Lefèvre 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.