Title page

Title

Patient safety in medication nomenclature: orthographic and semantic properties of international nonproprietary names

Short title

Patient safety in international drug names

Article type

Research Article

Keywords

medical nomenclature; look-alike sound-alike errors; international nonproprietary names; patient safety; linguistics

MeSH terms (checked 29/07/2015)

confusion/prevention & control, drug labelling/methods, drug labelling/standards, humans, medication errors/prevention & control, nurses, patients, pharmacists, physicians, pharmaceutical preparations, prescription drugs, reading, risk factors, terminology as topic, visual perception, World Health Organization, writing

Corresponding author and first author

Rachel Bryan, Swansea University, Wales, UK

Order of authors

Rachel Bryan[a*], Jeffrey K. Aronson[b], Pius ten Hacken[c], Alison Williams[a], Susan Jordan[a]

Suggested reviewers

- Silvia Alessi-Severini
- Ryan M Carnahan
- Barbara Mintzes
- Elisabeth Hildt
- Allan V Kalueff
- Yukari C Manabe
- Lorenzo Giovanni Mantovani
- Kent E Vrana
- Bruce Lambert – bruce.lambert@northwestern.edu
- Professor Robin Ferner – r.e.ferner@bham.ac.uk

Opposed reviewers

None

Attachments

Fig1.tiff, Fig2.tiff, Fig3.tiff, Strikingimage.tiff

Manuscript

Patient safety in medication nomenclature: orthographic and semantic properties of international nonproprietary names

Rachel Bryan[a*], Jeffrey K. Aronson[b], Pius ten Hacken[c], Alison Williams[a], Susan Jordan[a]
[a] Swansea University, Wales, UK
[b] Nuffield Department of Primary Care Health Sciences, University of Oxford, UK
[c] Institute for Translation Studies, University of Innsbruck, Austria
* Corresponding author: r.bryan@swansea.ac.uk

Abstract

Background

Confusion between look-alike and sound-alike (LASA) medication names (such as mercaptamine and mercaptopurine) accounts for up to one in four medication errors, threatening patient safety. Error reduction strategies include computerised physician order entry interventions, and 'Tall Man' lettering. The purpose of this study is to explore the medication name designation process, to elucidate properties that may prime the risk of confusion.

Methods and findings

We analysed the formal and semantic properties of 7,987 international non-proprietary names (INNs), in relation to naming guidelines of the World Health Organization (WHO) INN programme, and have identified potential for errors. We explored: their linguistic properties, the underlying taxonomy of stems to indicate pharmacological interrelationships, and similarities between INNs. We used Microsoft Excel for analysis, including calculation of Levenshtein edit distance (LED). Compliance with WHO naming guidelines was inconsistent. Since the 1970s there has been a trend towards compliance in formal properties, such as word length, but longer names published in the 1950s and 1960s are still in use. The stems used to show pharmacological interrelationships are not spelled consistently and the guidelines do not impose an unequivocal order on them, making the meanings of INNs difficult to understand. Pairs of INNs sharing a stem (appropriately or not) often have high levels of similarity (<5 LED), and thus have greater potential for confusion.

Conclusions

We have revealed a tension between WHO principles stipulating use of stems to denote meaning, and the aim of reducing similarities in nomenclature. To mitigate this tension and reduce the risk of confusion, the stem system should be made clear and well ordered, so as to avoid compounding the risk of confusion at the clinical level. The interplay between the different WHO INN naming principles should be further examined, to better understand their implications for the problem of LASA errors.

1 Introduction

International Nonproprietary Names (INNs) constitute a nomenclature of over 8,000 generic names

for pharmaceutical substances. Some examples are given in Table 1. They are designated by the World Health Organization (WHO) and formally placed in the public domain to promote consistency in global communication between manufacturers, clinicians, prescribers, and patients. INNs are published in the six official languages of the United Nations and Latin, and are used by default as generic names in major national and regional pharmacopoeias, such as the British Pharmacopoeia and the European Pharmacopoeia [1]. Given their international status, the name designation process in place must encompass a wider conceptual system than that of regional naming councils, and naming guidelines must be robust and applied stringently.

INNs are designated in accordance with a set of naming principles, which give guidance on formal properties, such as spelling, phonology, hyphenation, and word length, and semantic properties, such as the use of stems to indicate pharmacological relationships between substances. Here we present an analysis of the formal and semantic properties of INNs, based on the naming guidelines of the WHO INN program, and discuss their clinical significance.

Table 1: Examples of International Nonproprietary Names (INNs)

Year recommended	INN	Examples of current therapeutic indication(s) [2]
1955	chloramphenicol	Topical treatment of acute bacterial conjunctivitis
1965	betamethasone	Topical treatment of various dermatoses, including atopic dermatitis, psoriasis, and discoid lupus erythematosus
1975	levonorgestrel	72-hour emergency contraception
1985	mifepristone	Medical termination of developing intra-uterine pregnancy
1995	atorvastatin	Treatment of hypercholesterolaemia and prevention of cardiovascular disease
2005	golimumab	Treatment of rheumatoid arthritis, psoriatic arthritis, ankylosing spondylitis, and ulcerative colitis

1.1 Definitions

The EU directive 2001/83/EC [3] defined a *medicinal product* as:

> (a) any substance or combination of substances presented as having properties for treating or preventing disease in human beings; or (b) any substance or combination of substances which may be used in or administered to human beings either with a view to restoring, correcting or modifying physiological functions by exerting a pharmacological, immunological or metabolic action or to making a medical diagnosis.

OED (2015) defines 'medication' as "A drug or drugs prescribed or given as medical treatment; a medicine" [4]. In this paper we shall use 'medication' with the following definition [5], which was adapted to include placebos and other forms of medical interventions, such as inducing

anaesthesia:

> A manufactured article, intended to be taken by or administered to a person or animal, which contains a compound with proven biological effects, plus excipients, or excipients only, and may also contain contaminants.

'Medication' is often used interchangeably with 'medicine', 'medicinal product', and 'drug', but it is also used to denote the *process of giving* a medicine. Terminology usage in this paper is aligned to MeSH: 'medicine' refers to the practice and theory of medicine; 'drug' is normally restricted to adjectival use, e.g. 'drug-induced'; and 'medication' appears as a descriptor in multiple headings. MeSH terms without exception use 'drug' in relation to 'adverse', and so terms such as 'adverse drug reaction' will continue to be used. We are referring to prescribable medicinal products as components of therapeutic regimens.

By 'error' we mean:

> Something incorrectly done through ignorance or inadvertence; a mistake, e.g. in calculation, judgement, speech, writing, action, etc. or a failure to complete a planned action as intended, or the use of an incorrect plan of action to achieve a given aim. [5]

'Medication error' can thus be broadly defined as:

> A failure involving medication in the treatment process that leads to, or has the potential to lead to, harm to the patient.

Where 'medication' is defined as above (adapted from [5], to qualify the necessary condition of "involving medication").

1.2 Background

Medication errors make up a high proportion of all events related to patient safety [6,7], and are particularly common in intensive care, paediatrics/neonatology, care of the elderly, anaesthesia, and obstetrics [7,8]. Some medication errors will result in overdose, adverse drug reactions, or under-treatment, and cause serious harm to patients [9,10,11]. As more medications enter the market, with more variation in routes of administration, this problem is becoming increasingly complex [12].

Errors can occur when medications have similar-looking or similar-sounding names; these are called look-alike, sound-alike (LASA) errors. LASA errors are estimated to account for around one in every four medication errors in the USA [13], and they can occur during prescribing, transcribing, dispensing, and administration (examples in Table 2). Studies of United States Adopted Names (USANs), many of which take the form of INNs, have shown that the prescribing frequency of certain medications may prime the risk of LASA errors, and certain pre-approval strategies have been recommended, such as computerised searches, expert judgement, and psycholinguistic testing [14]. The bulk of the literature on LASA errors, involving confusion between both brand and generic names (brand-brand, generic-brand, and generic-generic), deals with mitigation strategies and regulatory obligations, such as 'Tall Man' lettering on packaging to highlight distinguishing characters (lamoTRIGine/lamiVUDine) and technological solutions, such as alerts built into prescription software and automated reporting systems [9,13,15,16,17].

Table 2: Examples of LASA errors

Reference	Medications	Type of	Clinical outcome

	involved (International Nonproprietary Name)	incident	
[18]	mercaptopurine (A); mercaptamine (B)	A prescribed instead of B by GP	Infant initially presented with nephropathic cystinosis. After one month on the wrong medication, the infant developed pancytopenia, but made a full recovery.
[19]	hydromorphone (C); morphine (D)	C administered instead of D by nurse	The patient (an elderly man) was discharged and suffered a fatal respiratory arrest on his way home.
[20]	gentamicin (E); clindamycin (F)	E administered instead of F	Not specified, but labelled 'low harm'.
[21]	cisatracurium (I); vecuronium (J)	I dispensed instead of J by pharmacy technician and was administered	The patient was a 30+1 week old neonate. The error was realised immediately, and no changes to vital signs were observed.

To date, a handful of studies have looked at the formal properties of generic names and their relation to LASA errors, and these concerned USANs, not INNs [22]. To our knowledge, this is the first study to contextualise the formal properties of INNs within the WHO naming guidelines, and the first to look at semantic properties, by exploring the underlying conceptual system that groups names according to their pharmacology. Since INNs are the global pharmaceutical nomenclature from which national nomenclature systems are derived, a study of the formal characteristics of the names is of real importance to those interested in medications management, LASA errors, and patient safety in general and would be expected to ascertain factors that should be considered in naming new medications.

INNs are designated and promoted for international use and are restricted by a set of principles. Principles 1 and 2 are marked as 'guiding principles', and the WHO stipulates that "these primary principles are to be implemented by using the following secondary principles" [23]. The WHO naming principles in designating INNs that we explore here are provided in Table 3, and the full list of principles in Appendix 1. The other five principles were not looked at, since they pertained to particular classes (such as acids and salts), the regulation of INN designation, or purely phonetic aspects. Motivation behind these selected principles falls into four categories:

1. Usability: how easily can the name be used in the four modalities of language: reading, writing, listening, and speaking? Is the name memorable, and can it be printed on packaging?
2. Taxonomy: does the name indicate its position in the conceptual system, and interrelationships? This is a two-fold condition: there must be a robust and consistent conceptual system, and the formal properties of the names should be exploited to map on to the underlying conceptual system.

3. Clarity: how liable is the name to be confused with other names, both in the same system and in other systems?
4. International use: does the name adhere to the phonotactics of all languages in which it is used, and can it be easily transliterated into other languages with different alphabets or writing systems? (As indicated above, INNs are published in Latin, English, Spanish, French, Russian, Chinese, and Arabic.)

Table 3: Selected WHO naming principles for designation of INNs (taken from [23]; sub-categorised here using square brackets)

Principle 1	International Nonproprietary Names (INNs) should be distinctive in [a] sound and [b] spelling. [c] They should not be inconveniently long and [d] should not be liable to confusion with names in common use.
Principle 2	[a] The INN for a substance belonging to a group of pharmacologically related substances should, where appropriate, show this relationship. [b] Names that are likely to convey to a patient an anatomical, physiological, pathological, or therapeutic suggestion should be avoided.
Principle 6	The use of an isolated letter or number should be avoided; hyphenated construction is also undesirable.
Principle 7	To facilitate the translation and pronunciation of INNs, "f" should be used instead of "ph", "t" instead of "th", "e" instead of "ae" or "oe", and "i" instead of "y"; the use of the letters "h" and "k" should be avoided. When devising an INN it is important to be aware of possible language problems. Since the name is used worldwide, not only should certain letters be avoided, but experts need to be aware of unsuitable connotations in the major languages spoken in the world.

Here we address the overarching research question: In relation to the WHO principles of designation of INNs, are there any threats to interpretation or translation in the form of:

1. Isolated numbers, isolated characters, or hyphens present in INNs (principle 6)
2. Prohibited graphs and digraphs present in INNs (principle 7)
3. Word length statistics (principle 1c)
4. Use of stems to indicate pharmacological relationships (principle 2a)
5. Patterns of similarity between INNs (principle 1d)

2 Methods

The present analysis is concerned with the formal and semantic properties of International Nonproprietary Names, and was undertaken within the framework of naming guidelines ('principles') set out by the WHO [23,24]. In linguistics, the descriptors *formal* and *semantic* are often dichotomised to compare, respectively, the written or phonetic form of a word and its underlying conceptual meaning(s). These are inseparable facets of natural language, but the distinction is useful for analytical purposes [25].

As a starting point for the analysis, all INNs (n=7,987) published in Recommended Lists from 1952 (when the INN program began) to August 2012 were digitised into an Excel spreadsheet. They were cross-verified on WHO MedNet. Two Excel databases were created, the

first containing all single-word INNs (n=7,111) and the second containing multi-word INNs (n=876). The multi-word database was used for analysis under Question 1 concerning isolated numbers, characters, or hyphens. Any names containing a space or a non-alphanumeric character (such as a hyphen) were included in the multi-word database. The single-word database was used for analysis of Questions 2–5.

The use of stems (Question 4), was explored qualitatively on a randomly selected 1% segment of the single-word database (n=71), as it was decided that for this question depth of analysis was preferable over breadth. The WHO Stembook [23] was used to verify that each INN in the 1% segment of the single-word database (n=71) had been named in accordance with its chemical structure and with WHO guidelines on the use of stems [23]. Of the 71 INNs, 43 were listed as correct in the WHO Stembook, and these were selected for analysis. The Stembook was referred to for definitions and guidance on the use of stems, and names in the sample were further investigated in the context of their wider conceptual relations.

For Question 5 concerning patterns of similarity between INNs, a further 1% sample of names (n=71) was chosen at random from the single-word database, and pairwise similarity was measured against the entire database, using Levenshtein edit distance (LED). LED can account for differences in word length [26] and is used in spell checking and predictive text software. It computes the number of insertions, deletions, or substitutions necessary to transform one string into another. For example, to transform *book* into *back*, *o* would be substituted for *a*, and the next *o* would be substituted for *c*, and the LED is 2. The LED is sensitive to differences in word length; for example, to transform mercaptopurine into mercaptamine, *a* replaces *o*, *m* replaces *p*, and *u* and *r* are deleted, so the LED is 4. Since the function's algorithm is processed linearly through the characters in each string, matching sequences will decrease the LED.

For ease of reference, this paper will refer to hyperonyms as stems and hyponyms as sub-stems. Stems are underlined for brevity.

3 Results

1. Are isolated numbers, isolated characters, or hyphens present in INNs? (principle 6)

[Insert box on graphemes – end of doc]

The presence of isolated characters in an INN can endanger its interpretation. In word processed documents, the name may be split over two lines and the isolated character may be misinterpreted as a page number or footnote marker. In handwriting, a single letter is more easily misinterpreted than a word, because the reader cannot rely on other characters for context. Isolated numbers may be mistaken for part of the dosage instructions and result in wrong dosing.

In the multiword INNs database, seven INNs contained a hyphen, six of which hyphenate a Greek letter and alphanumeric code, e.g. peginterferon lambda-1a. There are no instances of a single isolated number or character, although INNs with a second alphanumeric word, such as ioflubenzamide ^{131}I, create a risk of misinterpretation due to similarity between an upper-case eye, a lower case el, and the number one, even when printed, which all look the same in, for example, the Bauhaus 93 and Gill Sans MT fonts [27].

2. Are prohibited graphs and digraphs present in INNs? (principle 7)

Principle 7 is in place to facilitate the translation and correct pronunciation of INNs. By prohibiting graphs and digraphs such as <h> and <th>, which correspond to phonemes not predictably used

in other languages, the principle facilitates translation of the name from Latin into the six official languages of the WHO (English, Spanish, French, Russian, Chinese, and Arabic), from which generic names in other languages are derived. There are wide variations in the pronunciation and writing systems in the world's languages, so a simplification of English phonology is necessary. Principle 7 also serves Principle 1a and 1b (names should be distinctive in sound and spelling), by promoting the use of primary graphs with a one-to-one correspondence with phonemes, such as <f> (not <ph>) for phoneme /f/, and <e> (not <ae> or <oe>) for /e/. In this way, redundancy is avoided by using a single grapheme for each phoneme, and name length is reduced by one letter. INNs are therefore required to have an internal shallow orthography, i.e. one in which the correspondences between graphemes and phonemes are close to one-to-one.

Principle 2, which is given priority by the WHO, stipulates that names must show pharmacological relationship, and thus there is a conflict between principles 2 and 7. Although <ph>, <th>, <oe>, <y>, <h>, and <k> are prohibited, these (di)graphs are present in the list of stems used to form INNs (e.g. -kacin, -methasone, -orphinol) and so inevitably will be used. Furthermore, the Greek letter names *theta* and *kappa* are used to distinguish between similar preparations, for example, the biosimilars epoetin alpha, beta, theta, and zeta. Given the lower priority of Principle 7, it must be assumed that it should be adhered to unless the prohibited (di)graph forms part of a recommended stem, such as in amikacin. The numbers of instances of prohibited (di)graphs in INNs and stems are given in Table 4.

Table 4: Instances of prohibited (di)graphs in INNs and stems

	ph	th	ae	Oe	y	h	k	Total
All INNs containing the (di)graph	157	257	1	8	577	561	106	1677
Stems containing the (di)graph	4	1	0	1	4	5	14	29

Many instances of 'h' are attributable to those in <ph> and <th>, although it does appear a further 171 times in the single-word INN database, either as an initial letter (e.g. hydrocortisone) or with 'chlor' (e.g. chlorpromazine). In total, the prohibited (di)graphs occurred 1,277 times in 1,036 INNs. Some INNs contained more than one, such as phthalylsulfathiazole (<ph>, <th>, <th>).

As shown by Fig. 1 below, the majority of words containing prohibited (di)graphs were designated in the early stages of the INN programme, and few words continue to be designated. For example, although -methasone is the recommended stem, recently designated INNs have used the stem -metasone (e.g. dexamethasone in 1962, betamethasone in 1965, beclometasone in 1970 and alclometasone in 1979).

Fig. 1: Number of (di)graphs contained in INNs published, by decade.

3. Word length statistics (principle 1c)

Fig. 2 shows the character count distribution in single-word INNs. The mean number of characters across the entire dataset was 10.54, with a standard deviation of 1.73 characters. Both median and mode were 10, and the interquartile range was 9 to 11.

Fig. 2: Mean character count in 7,111 INNs

There are nine outliers with more than 20 characters, and these were all recommended before 1962. All nine contain at least one of the prohibited (di)graphs, which partly explains their unusual length, and at least seven syllables: sulfachlorpyridazine (20 letters), succinylsulfathiazole (21), methyldihydromorphine (21), phthalylsulfathiazole (21), ethylmethylthiambutene (22), phthalylsulfamethizole (22), sulfamethoxypyridazine (22), diiodohydroxyquinoline (22), and phenoxymethylpenicillin (23). It is notable that these names more closely resemble chemical names than the other INNs do.

Fig. 3: Mean character count of newly designated INNs (This figure shows date of publication: medications may have been in prior use, but not published as INNs.)

As shown in Fig. 3, the average length of INNs dropped sharply in the 1970s, and has remained steady since that time. The sudden drop coincided with the decreased use of (di)graphs prohibited under principle 7 from the 11th list onwards.

4. Are stems used to indicate pharmacological relationships? (principle 2a)

Pharmacological relationships between substances are demonstrated by the use of a common stem [28] attached to a prefix, infix, suffix, or 'freefix'. The use of stems and sub-stems creates a taxonomic conceptual system for INNs, and allows users to exploit this systematicity to improve retention, pronunciation, and recognition of the names. For example, montelukast comprises the stem -ast, and the sub-stem -lukast. Users can recognise montelukast as a medication used to treat asthma, or more specifically, a leukotriene receptor antagonist, and they can recognise that other names ending in –lukast have similar pharmacological actions to montelukast.

Stems and sub-stems may be suffixes (at the end of the name, such as granisetron and palonosetron), prefixes (at the beginning, such as artemether and arterolane), infixes (in the middle, such as mifepristone and ulipristal) and freefixes (which can appear anywhere, such as nabilone and nonabine). We encountered various types of taxonomy, which are outlined below. Stems analysed in the 1% sample are given in Table 5.

Table 5: Stems in 1% sample analysed semantically (stems, *sub-stems*)

INN	Actions and uses	freefix	prefix	infix	suffix
acridorex	Anorexic agent [29]				**-orex**
alverine	Relieves smooth muscle spasm in irritable bowel syndrome and dysmenorrhoea [2]				**-verine**
amezepine	Antidepressant [29]			*-ze-*	**-pine**
arterolane	Antimalarial; also known as RBx11160 [30]		**arte-**		
balaglitazone	Peroxisome proliferator-activated receptor (PPAR) gamma partial agonist, and potential treatment for diabetics [31]	**gli**			*-tazone*
benfosformin	Antihyperglycaemic agent [29]			*-fos-*	**-formin**
cebaracetam	Amide-type nootropic agent [23]				**-racetam**

cipamfylline	Phosphodiesterase (type IV) inhibitor. Selective inhibitor of tumour necrosis factor alfa production [29]				**-fylline**
clemastine	Allergic rhinitis and dermatoses [2]				**-astine**
conivaptan	Vasopressin receptor antagonist. Treatment of hyponatraemia [2]				**-vaptan**
disofenin	Imaging agent of liver and kidneys [29]				**-fenin**
doxaprost	Bronchodilator [29]	**prost**			
enolicam	Cyclooxygenase and lipoxygenase inhibitor . Anti-inflammatory, antirheumatic, antiarthritic agent [29]				**-icam**
fluperlapine	Antidepressant, neuroleptic agent. Never marketed [29]			*-a-*	**-pine**
indanidine	Partial alpha1-adrenoceptor agonist, alpha2-adrenoceptor antagonist; antihypertensive agent [29]				**-nidine**
indoprofen	Analgesic. Nonsteroidal anti-inflammatory agent [29]				**-profen**
inolimomab	Murine monoclonal antibody targeting interleukin-2 receptor-alpha chains [32]			*-lim-* *-o-*	**-mab**
ioxilan	Radiographic contract medium for urography or angiography [2]		**io-**		
lagatide	Antidiarrhoeal [23]				**-tide**
levopropylhexedrine	CNS stimulant; nasal decongestant; vasoconstrictor. Used as an anorectic in Germany. [29]				**-drine**
meprylcaine	Local anaesthetic used in dentistry [33]				**-caine**
micinicate	Vasodilator, spasmolytic agent [29]		**nico-**		*-nicate*
montelukast	Treatment and prophylaxis of asthma symptoms [2]			*-luk-*	**-ast**
nedocromil	Management of mild to moderate bronchial asthma; allergic conjunctivitis [2]				**-cromil**
nonabine	Cannabis analogue. Antiemetic agent [29]	**nab**			
palonosetron	$5HT_3$ receptor antagonist. Antiemetic used in cancer chemotherapy [2]				**-setron**
pelubiprofen	Analgesic. Non-steroidal anti-inflammatory drug [34]				**-profen**
peplomycin	Antineoplastic antibiotic [29]				**-mycin**

phenylephrine	Relief of nasal congestion associated with colds and hayfever [2]				**-frine**
pibutidine	Histamine-H2-receptor antagonist [23]				**-tidine**
pimelautide	Immunostimulant [29]				**-tide**
pranazepide	Allergic conjunctivitis [2]				**-azepide**
riodipine	Calcium channel blocker [23]				**-dipine**
salazosulfadimidine	Antibacterial agent [29]		**sal-** **sulfa-**	*-azo-*	
setiptiline	Tetracyclic antidepressant [35]				**-triptyline**
siplizumab	Humanized monoclonal antibody targeting CD2 receptors on T cells and natural killer cells [36]			*-li-* *-zu-*	**-mab**
sulfamerazine	Sulfonamide. Cream-coloured powder, darkens on exposure to light [29]		**sulfa-**		
tilmacoxib	COX-2 inhibitor [37]				**-coxib**
trifenagrel	Inhibitor of arachidonate and collagen-induced aggregation of platelets [38]	**grel**			
ubenimex	Competitive protease inhibitor [39]	**imex**			
ulipristal	Emergency contraception within 120 hours [2]			**-pris-**	
valaciclovir	Treatment of herpes zoster and ophthalmic zoster [2]	**vir**		*-ciclo-*	
volociximab	Chimeric monoclonal antibody targeting human α5β1 integrin [40]			*-ci-* *-xi-*	**-mab**

In the sample of 43 INNs analysed semantically (in Table 5), 20 (47%) contained 'lone' stems, which have no taxonomic relation to any other stems or sub-stems, such as -astine in clemastine, -azepide in pranazepide, and -cromil in nedocromil. Others contained a stem belonging to a larger taxon, such as montelukast (-ast: antiasthmatics or antiallergics, not acting primarily as antihistaminics > -lukast: leukotriene receptor antagonist). The stem taxon (Table 3) contains a hyperonym (-ast), and multiple hyponyms or sub-stems (-lukast, -trodast).

Table 6: -ast stem taxon

sub-stem (hyponym)	stem (hyperonym)	Mode of action	Example INN	Therapeutic indication
(without sub-stem)	-ast	antiasthmatics or antiallergics, not acting primarily as antihistaminics	zaprinast	A selective PDE inhibitor, precursor to PDE-5 inhibitors such as sildenafil (Viagra) [41]
-lukast		leukotriene receptor	montelukast	Treatment for reversible

		antagonists		bronchoconstriction [2]
-milast		phosphodiesterase type IV (PDE IV) inhibitors	cilomilast	Treatment for chronic obstructive airways disease and psoriasis [2]
-trodast		thromboxane A-2 receptor antagonists, antiasthmatics	seratrodast	Long-term management of asthma [42]
-zolast		leukotriene biosynthesis inhibitors	quazolast	Mediator release inhibitor [43]

An archetypal taxonomic system entails a clear tree hierarchy of concepts, based on hyperonyms and branching out to hyponyms. In the -ast taxon, and in many other INN taxa, the stem is a suffix, attaching to the end of the name, with its sub-stem as an infix. This right-to-left display of taxonomy in montelukast is a predictable approach for the user, as they can first categorise the substance under its main stem, antiasthmatics (-ast), and then sub-categorise it as a leukotriene receptor antagonist (-lukast).

Three names in the subset 'mab', shown in Table 7, are of monoclonal antibodies: inolimomab, siplizumab and volociximab. As a newer branch of biochemistry, they adhere to a much stricter nomenclature and a more systematic approach.

Table 7: -mab stem taxon

sub-stem A (target)	sub-stem B (source)	stem	Definition	Example INNs
		-mab	Monoclonal antibody	
	-a-		rat	Not yet designated
	-axo-		rat/mouse	catumaxomab
	-e-		hamster	Not yet designated
	-i-		primate	Not yet designated
	-o-		mouse	solitomab
	-u-		human	namilumab
	-xi-		chimeric	pagibaximab
	-xizu-		chimeric/humanised	otelixizumab
	-zu-		humanised	natalizumab
-b(a)-			bacterial	tefibazumab
-c(i)-			cardiovascular	volociximab
-f(u)-			fungal	Not yet designated
-k(i)-			interleukin	lebrikizumab
-l(i)-			immunomodulating	infliximab
-n(e)-			neural	atinumab
-s(o)-			bone	romosozumab
-tox(a)-			toxin	urtoxazumab

With the exception of the first medication in this class, muromonab-CD3, names for monoclonal antibodies comprise a random prefix, followed by two infixes and a stem referring in a specified order to (a) the target class or disease class; (b) the source class on which the immunoglobulin

sequence is based; and (c) the hyperonym -mab [23]. There are currently eight infixes to denote the source class and nine infixes to denote the target class. These may be combined freely with each other but the order in which they appear in the word is fixed. As an example, trastuzumab, which is a humanised monoclonal antibody directed against human epidermal growth factor receptor type 2 (HER2), can be decomposed as tras-tu-zu-mab, in which -tu- indicates that it targets tumour cells, and -zu- indicates that it is humanised (derived from a non-human antibody, which is then engineered to be more homologous with human antibodies). This is not an example of a three-tiered taxonomy, but rather two mutually independent parameters of classification under a single hyperonym, as shown in Table 7.

However, there is no simple rule governing the position of stems and sub-stems in INNs. There are many other ways in which stems are ordered in the name, because stems and sub-stems may be prefixes, suffixes, infixes or freefixes. The antiviral taxon, with hyperonymic stem vir, is below in Table 8. Vir is a freefix, and can appear anywhere in the name. Sub-stems of vir, including -viroc and -ciclovir, show that vir may be used as either a suffix or an infix, and so it is difficult for a user to immediately categorise a name under its hyperonym as an antiviral. A user may mistakenly categorize maraviroc under the stem *-oc, which does not exist. Other pharmacologically unrelated medications that happen to include the word part vir, such as virginiamycin and viridofulvin, may be misinterpreted as antivirals. In two other names in the sample, the sub-stem is a suffix and thus the name cannot be immediately recognised by its stem (such as: micinicate, with infix stem -nic(o)-; balaglitazone, with freefix stem gli).

Table 8: -vir stem taxon

Sub-stem	Stem	Mode of action	Example INN	Therapeutic indication
	vir	Antivirals	efavirenz	Antiretroviral combination therapy in treatment of HIV-1
-amivir		Neuraminidase inhibitors	zanamivir	Treatment of influenza [2]
-asvir		Antivirals, hepatitis C virus (HCV) NS5A inhibitors	daclatasvir	Combination therapy for chronic hepatitis C [2]
-buvir		RNA polymerase (NS5B) inhibitors	dasabuvir	Combination therapy for chronic hepatitis C [2]
-cavir		Carbocyclic nucleosides	abacavir	Antiretroviral combination therapy in treatment of HIV [2]
-ciclovir		Bicyclic heterocycle compounds	valaciclovir	Treatment of herpes zoster and ophthalmic zoster [2]
-fovir		Phosphonic acid derivatives	adefovir	Treatment of chronic hepatitis B [2]
-gosivir		Glucoside inhibitors	celgosivir	Potential treatment of dengue (DENV) virus [44]
-navir		HIV protease inhibitors	saquinavir	Treatment of HIV-1 infected adults [2]
-previr		Hepatitis Virus C (HVC) protease inhibitors	telaprevir	Treatment of genotype 1 chronic hepatitis C [2]
-virine		Non-nucleoside Reverse Transcriptase Inhibitors (NNRTI)	etravirine	Treatment of HIV-1 in antiretroviral-experienced adults and children over 6 [2]
-viroc		CCR5 (Chemokine CC motif receptor 5) receptor antagonists	maraviroc	Treatment of detectable CCR5-tropic HIV-1 [2]

In many cases the orthographic form, instead of being exploited to facilitate interpretation, seems to obfuscate the semantics of the names. While pharmacological relationship must be shown by using a common stem, some stems are distinguishable by only a single letter, such as -fenin (diagnostic aids; (phenylcarbamoyl)methyl iminodiacetic acid derivatives, e.g. lidofenin and disofenin) and -fenine (analgesics, glafenine derivatives, e.g. florifenine and glafenine) and many sub-stems in the -mab taxon comprise a single letter (cf. Table 7). In other cases, variation in the spelling of a stem (allomorphy) does not indicate a change in meaning (such as indoprofen and diprofene, or setiptiline and noxiptiline).

The use of -ine as the ending of INNs (the most frequent ending, 1502 of 7,111 names in the database) can falsely suggest a pharmacological relationship: for example, riodipine in the sample contains the stem -dipine (calcium channel blockers, 1,4-dihydropyridine derivatives), but could be mistakenly interpreted under -pine (tricyclic compounds). Homophonic yet distinct stems can also be misleading, such as -micin (gentamicin and netilmicin) and -mycin (erythromycin and kanamycin), denoting antibacterials. This example also reveals inconsistencies in the semantic motivation of stems. Here, the stem distinguishes the genus from which it is derived, -micins from Micromonospora and -mycins from Streptomyces, and not the pharmacological group, thereby blurring the boundaries between meanings of names containing these stems.

As in the INNs for monoclonal antibodies, meaning may depend on the order of stem concatenation. For example, when -fos is used as a suffix, it is a hyperonym denoting "insecticides, anthelminthics, pesticides etc., phosphorus derivatives", e.g. uredofos, but when it appears as an infix or a prefix, it is the hyponym denoting "various pharmacological categories belonging to fos, other than those above", e.g. benfosformin [23]. In other cases, position in the word does not indicate semantic difference, such as grel and vir, which have the same meaning regardless of whether they are used as an infix or a suffix.

5. Are there patterns of similarity between INNs? (principle 1d)

Of 504,881 total similarity measures, 1,463 had a Levenshtein edit distance of between 1 and 4, i.e. no more than 4 characters or deletions distinguished the names. 33% (478 pairs) of these shared a stem, and 88 (6%) also a sub-stem. These included prefixal stems such as arte- in arterolane-arteflene, with an LED of 3, and the prefixes salazo- and sulfa- present together in salazosulfadimidine-salazosulfamide, with an LED of 4. There were also pairs that had the same final letters but did not share a stem indicating pharmacological relationship, such as lagatide-giractide, in which the former has the stem -tide and the latter has the unrelated stem -actide. Similarity statistics are given in Table 6. The table indicates the strong influence of stems on similarity; the more similarity between two names, the more likely it is that they will share a stem and/or a sub-stem.

Table 6: Similarity statistics

Levenshtein Edit Distance	Frequency	Examples	Percentage sharing a stem	Percentage sharing both a stem and a sub-stem
1	2	alverine-salverine; amezepine-mezepine	100%	0%
2	19	peplomycin-peliomycin; clortermine-	84.2%	26.3%

		clormercaine		
3	240	inolimomab-solitomab; pelubiprofen-flurbiprofen	62.9%	13.3%
4	1202	pibutidine-sufotidine; meclofenoxate-metofenazate	25.7%	4.2%

The sample group contained four monoclonal antibody substances, represented by the -mab stem family (icrumab, inolimomab, siplizumab and volociximab). Of the pairs with an LED of <5, names for monoclonal antibodies occurred only in pairs, and did not display a high degree of similarity with names with other stems. Pairs of monoclonal antibodies presented a high proportion of all stem-based similarities (3 with LED=2; 27 with LED=3; 58 with LED=4; total 88). For example, siplizumab scored an LED of 2 with both ruplizumab and teplizumab, as they are only distinguishable by the first two letters. They share the stem -mab, and both sub-stems -zu- and -li-. In words such as these, when seven letters are predetermined by the norms of the designation process, the random prefix is responsible for the essential role of distinguishing the name from its co-hyponyms.

4 Discussion

The WHO naming principles analysed in this paper are not strictly observed. However, only the first two are prioritised, and these relate to the fourth and fifth objectives in this paper. The extremely long (20 char+) names and those containing prohibited (di)graphs mostly occurred in lists published before 1960, and there is a clear trend towards stricter adherence to WHO principles at later times. However, these earlier INNs are still in use, and it is difficult to amend names after publication in a recommended list in the public domain.

Another problem is that not all the principles are delimited and quantifiable, meaning they are difficult to follow and almost impossible to regulate. For example, principle 1 ("International Nonproprietary Names (INNs) should be distinctive in sound and spelling. They should not be inconveniently long and should not be liable to confusion with names in common use" [23] does not provide quantifiable criteria, such as a character limit, or a maximum degree of similarity.

4.1 Formal Properties of Medication Names

In general, medication name designation complied with the WHO principles relating to formal properties of nomenclature.

Some INNs contained hyphens, but no isolated letters and numbers were found. 1,277 INNs contain prohibited (di)graphs such as <ph> rather than <f>. Words without a one-to-one grapheme-phoneme correspondence, such as *thorough* (eight graphs, four phonemes – /θərə/), take longer to be recognised in reading. Words containing graphs (letters) without direct correspondence to their phonemes have been found to take longer to be recognised in reading [45]. For example, *fooph (3 phonemes and 5 graphs) will take longer to recognise than *fruls (5 phonemes and 5 graphs), because the reader first interprets the <p> in *fooph as the phoneme /p/, but then on meeting <h> is forced to reinterpret as the phoneme /f/. Recognition time is further increased when non-correspondences occur earlier in the word, as the reader would not be

able to infer meaning from the context of earlier letters. When analysed into single graphs, the digraphs <ph>, <th>, <ae>, and <oe> do not have direct correspondence with the primary phonemes of each graph and may increase recognition time and reduce usability.

Word lengths of INNs are relatively stable diachronically, with an overall mean character count of 10.54, but there are a few INNs with more than 20 characters. Long INNs, such as phenoxymethylpenicillin (23 characters), are problematic to fit on packaging; the NPSA in the UK, for example, has recommended a minimum font size of 16 points for the generic name [46]. Long INNs risk being hyphenated and running to multiple lines when printed, reducing legibility and increasing the risk of misunderstanding or confusion with other names. Recognition time will be increased by high character count, low frequency of the words and perceived 'nonwords' [47,48], and this effect may be magnified by variation in the prescribing frequency of INNs [49].

4.2 Potential for LASA errors

We have identified a tension between WHO principles stipulating the use of stems to indicate pharmacological relationships and those aiming to reduce similarity in nomenclature.

Stems indicating pharmacological relationships

There is no single way for a user to predict meaning from an INN, although we found consistency within some taxa, e.g. monoclonal antibodies and antivirals. In some cases, a single letter will be used to distinguish between unrelated pharmacological groups, whereas in other cases there is simply wide spelling variation that does not contribute to meaning. The burden on users is high: they need to understand the meanings of stems and the layout of the taxonomy, and also to learn when to ignore spelling variations and when to take note of a single letter distinguishing meaning. They must understand that meaning may be motivated at the supramorphemic level by the class of affix and the concatenation of stems and sub-stems. Consequently, most clinicians make little use of pharmacological nomenclature in routine practice, relying instead on the appearance and sounds of whole words, memorised during experiential learning in clinical environments and on pharmaceutical company presentations, and preferring to use simpler brand names.

Patterns of similarity between INNs

Pharmacologically related substances whose names show their relationship by the use of a common stem, and those that were unrelated but erroneously shared a stem, have a higher level of similarity and are thus more likely to be confused, such as:

1. arterolane-arteflene (arte-: antimalarial agents, artemisinin related compounds);
2. salazosulfadimidine-salazosulfamide (sal-: analgesic anti-inflammatories; salazo-: phenylazosalicyclic acid derivatives antibacterial) (sulfa-: anti-infectives, sulfonamides)
3. siplizumab-teplizumab (-mab: monoclonal antibodies; -li- and -zu-: humanised, targeting the immune system)

This suggests that the use of the stem system may actually increase the risk of confusion and thus endanger patients. However, without it users of pharmaceutical names would need to learn the meanings of all medicines by rote, without the benefit of common affixes. Thus, two primary objectives of the WHO – usability and taxonomy – are in competition with each other, and this is compounded by a messy underlying taxonomy.

We found significant levels of similarity between pharmacologically related INNs. Hence, we have identified dissonance between sub-principles 1d (mitigating the risk of confusion) and 2a (stipulating the display of pharmacological relationship). These sub-principles compete in the

pursuit of the primary goals, namely, reduction of the risk of confusion with other medication names and accurate perception of the meanings of INNs. This conflict can only be mitigated when the stem system is predictably structured, to avoid increasing the risk of confusion.

5 Limitations

We have looked at formal and semantic properties of International Nonproprietary Names within a selection of the WHO naming principles. We have examined orthographic form, but not phonetic form, and used only one similarity measurement method. We analysed only INNs in English, but studies are underway, adapting these methods to the analysis of translated forms (for example, in Latin, Spanish, French, Russian, Chinese, and Arabic), and evaluating their compliance with WHO naming principles [23]. Further work is needed to explore the clinical implications of this work.

6 Conclusions

INNs carry out different functions depending on the user, and so they must be understandable at multiple levels. For patients and non-professionals they must be recognisable and pronounceable, and simultaneously their meaning and pharmacological relationships with other names must be understood by health professionals if medication errors are to be avoided.

The pharmaceutical nomenclature and its peripheral systems of nomenclature (such as pathology, anatomy, nosology, etc.) are extensive and complex. It is inevitable that a taxonomy developed over a number of decades will contain some broken links and general inconsistencies, but these should not work to the detriment of the overall aim of the system.

Like two sides of a coin, the formal and semantic aspects of language are inextricably linked, and it is impossible to speak of formal motivation without referring to semantic motivation. Formal aspects of INNs are motivated by the semantics they represent, and while the formal realisation of INNs is, at times, conducive to conveying their meaning, it can also misrepresent meaning and increase the risk of confusion.

Findings on word length (Question 3) were closely aligned with a similar study on USANs [22], and have highlighted certain (di)graphs that are prohibited but nevertheless still in use (Question 2). Results for Questions 4 and 5, regarding the use of stems and similarity, have exposed a tension in the INN nomenclature, and highlighted the need for further research into the exact interplay between these naming principles and their implementation.

7 References

[1] European Union, EU. Council Directive. Official Journal of the European Communities No L, 1992: 31 March: 113/8-12(92/27/EEC); 1992.

[2] Electronic Medicines Compedium, eMC. Home page [cited 22 Jul 2015]. Available from: http://www.medicines.org.uk/emc/.

[3] EU. Council Directive. Official Journal of the European Communities No L 311, 2001: 28 November: (01/83/EC); 2001.

[4] Oxford English Dictionary, OED. Home page [cited 22 Jul 2015]. Available from: http://www.oed.com/.

[5] Aronson, JK. Medication errors: definitions and classification. Br J Pharmacol. 2009a;67(6):599–604.

[6] Jordan, S, Kyriacos, U. Medicines' management: a public health problem on nursing's agenda. J Nurs Manag. 2014;22(3):271–5.

[7] Ostini, R, Roughead, EE, Kirkpatrick, CMJ, Monteith, GR & Tett, SE. Quality Use of Medicines - medication safety issues in naming; look-alike, sound-alike medicine names. Int J Pharm Pract. 2012;20(6):349–57.

[8] Kaushal, R, Bates, DW, Landrigan, C, McKenna, KJ, Clapp, MD, Federico, F et al. Medication Errors and Adverse Drug Events in Pediatric Inpatients. JAMA. 2001;285(16):2114–2120.

[9] Aronson, JK. Medication errors: what they are, how they happen, and how to avoid them. QJM: Monthly Journal of the Association of Physicians. 2009b;102(8):513–21.

[10] Runciman, WB. Adverse drug events and medication errors in Australia. Int J Qual Health Care. 2003;15(90001):49–59.

[11] Moorman SM, Carr D. Spouses' effectiveness as end-of-life health care surrogates: accuracy, uncertainty, and errors of overtreatment or undertreatment. Gerontologist 2008;48(6):811–9.

[12] Nute, C. Reducing medication errors. Nurs Stand. 2014;29(12):45.

[13] Emmerton, LM, Rizk, MFS. Look-alike and sound-alike medicines: risks and "solutions". Int J Clin Pharm. 2012;34(1):4–8.

[14] Lambert, BL, Lin, S-J, Tan, H. Designing Safe Drug Names. Drug Saf. 2005;28(6):495–512.

[15] Ghaleb, MA, Barber, N, Franklin, BD, Wong, ICK. The incidence and nature of prescribing and medication administration errors in paediatric inpatients. Arch Dis Child. 2010;95(2):113–8.

[16] Kovacic, L, Chambers, C. Look-alike, sound-alike drugs in oncology. J Oncol Pharm Pract. 2011;17(2):104.

[17] Galanter WL, Bryson ML, Falck S, Rosenfield R, Laragh M, et al. Indication Alerts Intercept Drug Name Confusion Errors during Computerized Entry. PLoS ONE 2014;9(7):e101977.

[18] Medicines and Healthcare Products Regulatory Agency, MHPRA. Mercaptamine and mercaptopurine: confusion between drug names. 2002 [cited 22 Jul 2015]. Available from: https://www.gov.uk/drug-safety-update/mercaptamine-and-mercaptopurine-confusion-between-drug-names.

[19] Tuohy, N, Paparella, S. Look-alike and sound-alike drugs: errors just waiting to happen. J Emerg Nurs. 2005;31(6):569–71.

[20] National Reporting and Learning Service, NRLS. Safety in Doses: Improving the use of medicines in the NHS [cited 22 Jul 2015]. NPSA: London; 2009. Available from http://www.nrls.npsa.nhs.uk/resources/?entryid45=61625.

[21] Sauberan, JB, Dean, LM, Fiedelak, J, Abraham, JA. Origins of and solutions for neonatal medication-dispensing errors. Am J Health Sys Pharm. 2010;67:49–57.

[22] Lambert, BL, Chang, K-Y, Lin, S-J. Descriptive analysis of the drug name lexicon. Drug Inf J. 2001;35(1):163.

[23] World Health Organization, WHO. WHO Stembook 2011: The use of stems in the selection of International Nonproprietary Names (INN) for pharmaceutical substances. 2011 [cited 22 Jul 2015]. Available from: http://www.who.int/medicines/services/inn/StemBook_2011_Final.pdf.

[24] Bryan, R. Taxonomy and Transparency in International Pharmaceutical Nomenclature. In: ten-Hacken, P, Panacova, R. editors. Word Formation and Transparency in Medical English. Cambridge: Cambridge Scholars Press; 2015, in press.

[25] Sanders, C. The Cambridge Companion to Saussure. Cambridge: Cambridge University Press; 2004.

[26] Lambert, BL, Lin, S-J, Chang, K-Y, Gandhi, SK. Similarity as a Risk Factor in Drug-Name Confusion Errors. Med Care. 1999;37(12):1214.

[27] Garfield, S. Just My Type: A book about fonts. London: Profile Books; 2010.

[28] WHO. Guidance on the Use of International Nonproprietary Names (INNs) for Pharmaceutical Substances. 1997 [cited 22 Jul 2015]. Available from: http://whqlibdoc.who.int/hq/1997/who_pharm_s_nom_1570.pdf.

[29] Ganellin, CR, Triggle, DJ. Dictionary of Pharmacological Agents. London: Chapman & Hall; 1996.

[30] Uhlemann, A-C, Wittlin, S, Hugues, M, Bustamente, LY, Krishna, S. Mechanism of Antimalarial Action of the Synthetic Trioxolane RBX11160 (OZ277). Antimicrob Agents Chemother. 2007;51(2):667–672.

[31] Youssef, JA, Badr, MZ. Peroxisome Proliferator-Activated Receptors. New York: Humana Press; 2013.
[32] Bay, JO, Dhédin, N, Goerner, M, Vannier, JP, Marie-Cardine, A, Stamatoullas, A, et al. Inolimomab in steroid-refractory acute graft-versus-host disease following allogeneic hematopoietic stem cell transplantation: retrospective analysis and comparison with other interleukin-2 receptor antibodies. Clin Transplant. 2005;80(6):782–788.
[33] Morton, I, Morton, I, Hall, JM. Concise Dictionary of Pharmacological Agents. The Netherlands: Kluwer Academic Publishers; 1999.
[34] Nuutinen, L, Raj, PP. An overview of current and investigational non-narcotic drugs for treatment of acute and chronic pain. Curr Rev Pain. 1998;2(3):187–192.
[35] Miyamoto, S. Setiptiline Maleate. In: Stolerman, IP, editor. Encyclopedia of Psychopharmacology. Springer: Berlin; 2010.
[36] Langley, RG, Papp, K, Bissonnette, R, Toth, D, Matheson, R, Hultquist, M, et al. Safety profile of intravenous and subcutaneous siplizumab, an anti-CD2 monoclonal antibody, for the treatment of plaque psoriasis: results of two randomized, double-blind, placebo-controlled studies. Int J Dermatol. 2010;49(7):818–828.
[37] Yamamoto, H, Kondo, M, Nakamori, S, Nagano, H, Wakasa, KI, Sugita, Y, et al. JTE-522, a cyclooxygenase-2 inhibitor, is an effective chemopreventive agent against rat experimental liver fibrosis. Gastroenterology 2003;125(2):556–571.
[38] Mukhopadhyay, C, Tapaswi, PK, Drew, MGB. Room temperature synthesis of tri-, tetrasubstituted imidazoles and bis-analogues by mercaptopropylsilica (MPS) in aqueous methanol: application to the synthesis of the drug trifenagrel. Tetrahedron Lett. 2010;51(30):3944–3950.
[39] Sekine, K, Fujii, H, Abe, F. Induction of apoptosis by Bestatin (ubenimex) in human leukemic cell lines. Leukemia. 1999;13(5):729–34.
[40] Bell-McGuinn, KM, Matthews, CM, Ho, SN, Barve, M, Gilbert, L, Penson, RT, et al. A phase II, single-arm study of the anti-α5β1 integrin antibody volociximab as monotherapy in patients with platinum-resistant advanced epithelial ovarian or primary peritoneal cancer. Gynecol Oncol. 2011;121(2):273–279.
[41] Choi, SH, Choi, DH, Song, KS, Shin, KH, Chun, BG. Zaprinast, an inhibitor of cGMP-selective phosphodiesterases, enhances the secretion of TNF-α and IL-1β and the expression of iNOS and MHC class II molecules in rat microglial cells. J Neurosci Res. 2002;67(3):411–421.
[42] Dogné, JM, de Leval, X, Benoit, P, Delarge, J, Masereel, B. Thromboxane A2 inhibition: therapeutic potential in bronchial asthma. Am J Respir Med. 2002;1(1):11–17.
[43] Dave, NK, McMahon, SC, Grubbe, RE, Bewtra, AK, Hopp, RJ, Nair, NM, et al. A controlled, double-blind study of the effect of quazolast on nasal challenge with ragweed antigen. Ann Allergy. 1990;65(4):298–302.
[44] Low, JG, Sung, C, Wijaya, L, Wei, Y, Rathore, AP, Watanabe, S, et al. Efficacy and safety of celgosivir in patients with dengue fever (CELADEN): a phase 1b, randomised, double-blind, placebo-controlled, proof-of-concept trial. Lancet Infectious Dis. 2014;14(8):706–715.
[45] Rastle, K, Coltheart, M. Whammies and double whammies: The effect of length on nonword reading. Psychon Bull Rev. 1998;5(2):277–282.
[46] National Patient Safety Agency, NPSA. A guide to the graphic design of medication packaging. 2006 [cited 22 Jul 2015]. Available from: http://www.hhc.rca.ac.uk/CMS/files/NPSA-Design-for-patient-safety-.pdf.
[47] New, B, Ferrand, L, Pallier, C, Brysbaert, M. Reexamining the word length effect in visual word recognition: New evidence from the English Lexicon Project. Psychon Bull Rev. 2006);13(1):45–52.
[48] Coltheart, M, Rastle, K, Perry, C, Langdon, R, Ziegler, J. DRC: a dual route cascaded model of visual word recognition and reading aloud. Psychol Rev. 2001;108(1):204.
[49] Lambert, BL, Chang, K-Y, Gupta, P. Effects of frequency and similarity neighborhoods on pharmacists' visual perception of drug names. Soc Sci Med. 2003;57(10):1939–55.

BOX

Some key concepts in linguistics need to be explained. Graphemes are the letters and symbols in the writing system of a language that correspond to the phonemes in the spoken system of that language. Graphemes should be seen as abstract concepts, which are realised by graphs (letters) and allographs (variant forms of those letters) in the actual typed or handwritten language. For example, the phoneme /f/ corresponds to the grapheme <f>, which is realised by the primary graph <f>, but also by allographs <ff> (*puff*), <ph> (*photo*), <gh> (*rough*) [4]. Single phonemes may be represented by two graphs, or a 'digraph', such as <ph>. Languages without a one-to-one grapheme-phoneme correspondence are said to be 'deep orthographies', and English is a prime example: it has been famously pointed out that in English, fish could equally be written *ghoti*, using the 'gh' in *rough*, the 'o' in women, and the 'ti' in *nation*.

Appendix 1: WHO naming principles for designation of INNs [23]

Principle 1 (guiding principle)	International Nonproprietary Names (INN) should be distinctive in sound [a] and spelling [b]. They should not be inconveniently long [c] and should not be liable to confusion with names in common use [d].
Principle 2 (guiding principle)	The INN for a substance belonging to a group of pharmacologically related substances should, where appropriate, show this relationship [a]. Names that are likely to convey to a patient an anatomical, physiological, pathological or therapeutic suggestion should be avoided [b].
Principle 3	In devising the INN of the first substance in a new pharmacological group, consideration should be given to the possibility of devising suitable INNs for related substances, belonging to the new group.
Principle 4	In devising INNs for acids, one-word names are preferred; their salts should be named without modifying the acid name, e.g. "oxacillin" and "oxacillin sodium", "ibufenac" and "ibufenac sodium"
Principle 5	INNs for substances which are used as salts should in general apply to the active base or the active acid. Names for different salts or esters of the same active substance should differ only in respect of the name of the inactive acid or the inactive base.
Principle 6	The use of an isolated letter or number should be avoided; hyphenated construction is also undesirable.
Principle 7	To facilitate the translation and pronunciation of INN, "f" should be used instead of "ph", "t" instead of "th", "e" instead of "ae" or "oe", and "i" instead of "y"; the use of the letters "h" and "k" should be avoided. When devising an INN it is important to be aware of possible language problems. Since the name is used worldwide, not only should certain letters be avoided, but experts need to be aware of unsuitable connotations in the major languages spoken in the world.
Principle 8	Provided that the names suggested are in accordance with these principles, names proposed by the person discovering or first developing and marketing a pharmaceutical preparation, or names already officially in use in any country, should receive preferential consideration.

Manuscript

Principle 9	Group relationship in INNs (see Guiding Principle 2) should if possible be shown by using a common stem.