Biologicals 60 (2019) 15–23

Contents lists available at ScienceDirect

Biologicals

journal homepage: www.elsevier.com/locate/biologicals

Review

The INN global nomenclature of biological medicines: A continuous challenge

James S. Robertson[a], Wai-Keung Chui[b], Armando A. Genazzani[c], Sarel F. Malan[d], Ana López de la Rica Manjavacas[e], Gilles Mignot[f], Robin Thorpe[a], Raffaella Balocco[g], Menico Rizzi[c,*]

[a] Independent Expert, United Kingdom
[b] Department of Pharmacy, National University of Singapore, Singapore
[c] Department of Pharmaceutical Sciences, University of Piemonte Orientale, Novara, Italy
[d] School of Pharmacy, University of the Western Cape, Bellville, South Africa
[e] Agencia Española de Medicamentos y Productos Sanitarios, Madrid, Spain
[f] Independent Expert, France
[g] INN Programme, TSN/RHT/MVP, World Health Organization, Geneva, Switzerland

ARTICLE INFO

Keywords:
International non-proprietary name
Biological medicines
Drug nomenclature

ABSTRACT

Medicines are assigned International Nonproprietary Names (INN) by the World Health Organization (WHO), pursuing the aim to increase patient safety. Following scientific developments in drug discovery and biotechnology, the number of biological medicines is constantly growing and a surge in INN applications for them has been observed. Pharmacologically active biological substances have a complex structure and mechanism of action posing new challenges in selecting names that appropriately reflect such properties. As a consequence, existing nomenclature naming schemes may need to be revised and new ones developed. This review reports on the recently implemented policies for naming fusion proteins, monoclonal antibodies, advanced therapy substances that cover gene and cell therapy, virus-based therapies as well as vaccines and vaccine-like substances. Different approaches, based on the use of a one-word versus a two-word naming scheme, have been developed for different categories of biological substances highlighting a major and still not completely resolved issue, i.e. how to assign a name that is both informative, short and euphonic.

1. Introduction

International Nonproprietary Names (INN) are assigned to active pharmaceutical substances by the World Health Organization (WHO) following the World Health Assembly Resolution WHA3.11 approved in 1953. The INN Programme for pharmaceutical substances was established in 1953 and the INN Expert Group was officially designated to select INN according to specific policies developed by the Group [1,2]. The main objective of INN is to provide names for pharmaceutical substances that are adopted globally, and are recognizable and distinct; they are deposited by WHO in the public domain and hence are public property. INN are intended to have broad usage covering drug regulation, prescribing, pharmacopoeias, pharmacovigilance, labelling, dispensing, teaching and scientific literature. One of the expected main benefits of INN is therefore, overall, to ensure patient safety.

INN typically begin with a fantasy prefix and terminate with a suffix that indicates the pharmacological relationships between substances. The suffix is selected as an official stem when a group of pharmacologically related substances is named with a common suffix, and can be composed by a simple suffix/stem and one or more infixes/substems [3,4]. In a few cases the stem can be placed in the beginning of the name. The purpose of the stem is to group medicines that have similar therapeutic or clinical actions, both to minimise the simultaneous use of similar medicines which may result in increased adverse responses and to facilitate the use of alternative medicines when use of one becomes ineffective. The stem is developed based on one, or a combination or exclusion, of three criteria: structure, mechanism of action, and clinical indication. INN generally consist of a single word but occasionally two or more words may be used.

The INN Programme follows the evolution of drug development and although in the beginning mainly classical chemical drugs were processed, INN have also been assigned to biological substances. For

* Corresponding author. Department of Pharmaceutical Sciences, University of Piemonte Orientale, Largo Donegani 2, 28100, Novara, Italy.
E-mail address: menico.rizzi@uniupo.it (M. Rizzi).

https://doi.org/10.1016/j.biologicals.2019.05.006
Received 21 February 2019; Received in revised form 7 May 2019; Accepted 14 May 2019
Available online 23 May 2019

instance, animal insulin preparations were given an INN in Recommended List 3 in 1959 and following this pioneering example, names were assigned to synthetic peptides and to antibiotics, hormones and other substances of biological origin. A major issue in naming biological substances derives from their high degree of structural micro-heterogeneity compared to chemical substances and according to a major policy of the INN Programme, *i.e.* that a name can be assigned only to a clearly chemically defined and homogeneous substance, several groups of biologicals are not assigned an INN, examples being natural blood products such as immunoglobulins fractioned from plasma, skin substitutes and vaccines. However, given the advances in science, well characterized biologicals albeit with some degree of micro-heterogeneity, are currently being named based on clear and specific policies. Thus in 1982, the name *insulin human* was selected for the recombinant DNA derived protein identical to human insulin and since then many recombinant proteins have been assigned INN including non-glycosylated and glycosylated proteins/peptides, fusion proteins, monoclonal antibodies (mAbs), pegylated proteins/peptides, cell therapies, gene therapies and virus-based therapies.

Biological substances represent the fastest-growing category in the pharmaceutical world and the scientific and technical developments that led in the past years to the impressive explosion of novel therapies based on biologicals continues now and promises to continue in the future. Therefore, more and different biological substances are expected to emerge for which INN will be assigned with the need for new naming policies to be developed. This review will focus on the latest implemented policies for naming fusion proteins, monoclonal antibodies and advanced therapy.

2. Nomenclature for proteins

Proteins are the most abundant class of biological macromolecules covering the widest spectra of biological functions and representing the largest class of biological medicinal products. In the INN system, protein substances can fall into many different categories, e.g. the large family of monoclonal antibodies (stem = *-mab*), the receptor group (stem = *-cept*), as well as several other groups depending on their mode of action. The policy used to assign INN to non-glycosylated proteins follows the general format for INN, i.e. a suffix/stem that identifies the pharmacological group, in some cases an infix giving more details of the group, and a fantasy prefix that identifies the specific protein, usually according to the amino acid sequence. Peptides, which biochemically can be considered as proteins and are also endowed with a variety of different modes of action, fall, in terms of INN, into a separate group featured by the *–tide* stem.

As the aim of this review is to report on the most recent developments in INN nomenclature schemes for biological medicines, any unchanged and well established policies to name a protein substance falling under a specific category (e.g. enzymes, growth factors and interleukins), will not be discussed and readers are directed to the INN Bioreview [5]. With regard to glycosylated proteins, the policy of using Greek letters to distinguish different glycoforms of the same protein is reiterated and will continue to be applied in a systematic manner starting with the letter *alfa*. The exception to this rule applies to the receptor and monoclonal antibody families, where only the second and subsequent glycosylated forms of the substance receive an identifier in the form of a Greek letter, beginning with *beta*. The necessity to differentiate glycosylated forms of a protein is considered of high relevance in the light of available evidence that suggests glycosylation can indeed have an impact on the pharmacological activity of the substance and not necessarily only on its stability [6].

2.1. Fusion proteins nomenclature scheme

A fusion protein in terms of the INN programme is defined as “a protein encoded from one nucleotide sequence which comprises two or more genes or portions of genes - and possibly linkers - which originally coded for separate proteins” [5]. Initially fusion proteins were named according to the primary mode of action, e.g. etaner*cept* comprises the TNF receptor fused to an IgG Fc moiety; however, the fast-growing family of fusion proteins consisting of multi-functional components has compelled the implementation of a new naming policy. A new scheme was developed from ideas and opinions collected from different stakeholders at the “Meeting on Biologicals” held in September 2016 in Geneva [7] and was based on the main issue of whether a one-word or two-word naming scheme would be more appropriate. It was reasoned that not all fusion proteins will deserve a new stem but only those in which at least two different components of the fusion protein are endowed with a pharmacological action relevant to the activity of the medicine and the main concept that emerged was that a fusion protein should be considered as a unique substance, with a new amino acid sequence, and that a one word naming scheme would be preferable. Starting from this major concept, the INN Expert Group developed a policy for naming fusion proteins that was implemented and used during the 64th INN Consultation, in April 2017.

The suffix *-fusp* was introduced as a robust stem in terms of linguistic requirements and that clearly identifies a fusion protein. In addition to the suffix *-fusp*, an infix syllable formed by one consonant and one vowel will be added in front of the suffix to indicate: (a) consonant - the pharmaceutical action, and (b) vowel - the targeting. The meanings of these infix letters are given in Table 1. It is emphasized that the *-fusp* naming scheme has not been designed to provide comprehensive information about the substance in the name but rather to indicate that it is a fusion protein and its type; indeed, the two letters infix preceding the *-fusp* suffix only indicate broad categories. To compensate for this paucity of information within the name, the description at the level of publication will provide extensive information about the precise content and action of the fusion protein. Another mandatory discriminating characteristic that must be satisfied to include a fusion protein in the *-fusp* family, is that each component must be endowed with a pharmacological activity; in a bifunctional fusion protein in which one component has a purely stabilizing function (e.g. to increase half-life), the substance will not be assigned the *-fusp* stem. For example, if the component is a stabilizing Fc fragment, the *ef-*prefix should be used, following policy for monoclonal antibodies, and not the *-fusp* suffix; however, in a multifunctional fusion protein that has more than one pharmacological action and also contains a stabilizing Fc fragment, both *ef-* and *-fusp* should be used.

Due to the expected abundancy of fusion proteins containing a monoclonal antibody as one of the components, the *–fusp* naming

Table 1
Nomenclature scheme for fusion protein: infix letters and their meaning.

Action		Targeting	
-b-	binding protein	*-a-*	antibody
-c-	encapsulation protein	*-e-*	receptor
-f-	hormone	*-i-*	antigen
-g-	antigen	*-o-*[b]	other
-k-	cytokine	*-u-*[c]	untargeted
-m-	membrane protein		
-n-	enzyme		
-p-	apoptosis		
-r-	receptor		
-t-	T-cell receptor		
-v-[a]	multiple actions/proteins		
-x-	toxin		

[a] *-v-* will be used when a multifunctional fusion protein has multiple and not related actions.

[b] *-o-* will be used when some other targeting mechanism (i.e. not antibody, receptor or antigen) is used in a bifunctional fusion protein or in a multifunctional fusion protein with multiple unrelated targeting.

[c] *-u-* will be used when a fusion protein has multiple actions and no targeting.

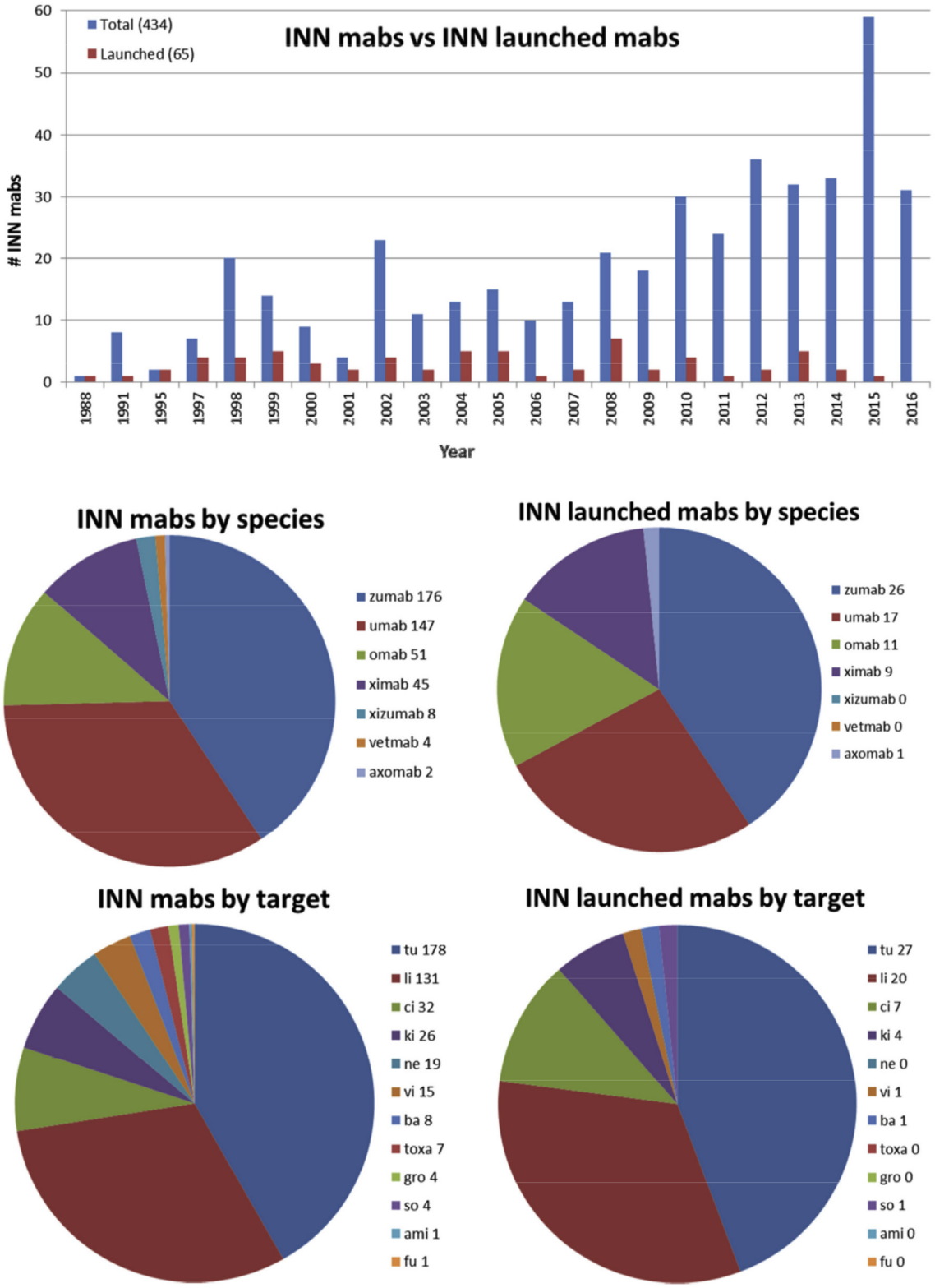

Fig. 1. Total versus launched INN mAbs in the period 1998–2016, global number. Generated based on WHO data (extracted from WHO Drug Information, Mednet and Integrity databases). Upper panel: all published mAbs (in INN publication) versus mAbs which received a marketing authorisation or already in the market. Lower panels: as in the upper panel divided by species or target according to infixes defined in the former mAbs nomenclature scheme.

scheme draws special attention to such cases. If both components of the fusion protein have a targeting action, and one of them is derived from a mAb or from a mAb fragment, when assigning the identifying infix letters, the "*-a-*" for *antibody* takes priority. For instance, for a fusion protein consisting of a receptor and an antibody, the infix will be *-ra-* (where *r* stands for *receptor* and *a* for *antibody*) and not *-be-* (where *b* stands for *binding protein* and *e* for *receptor*). Moreover, the infix letters will not distinguish between mAb or mAb fragments, and in all these cases the letter "*a*" will be selected; multiple mAb or mAb fragments will be named using the *-mab* nomenclature scheme, not the *-fusp*

scheme.

Although multi-component fusion proteins, i.e. those derived from three or more functional proteins, have not been assigned an INN so far, it is considered that applications for INN for such proteins will appear in the future; in such cases, the two infix letters used to represent the different action/targeting by class will continue to apply, e.g. if a fusion protein is composed of two mAbs and one receptor, the INN will end in *-rafusp*.

Finally, the policy currently in place for glycosylated proteins will be applied to fusion proteins, regardless of the nature of the different components, i.e. even in the presence of a receptor or a monoclonal antibody; therefore, the first glycosylated form of a fusion protein will receive a Greek letter starting from *alfa*.

The new scheme for naming fusion proteins responds to a logic that deviates from the general approach of the INN Programme. The scheme has been developed purposely to provide only a clear indication of the category the substance belongs to, without providing specific information on either the mechanism of action or the clinical indication. As examples, the following three fusion proteins have been named according to the new nomenclature scheme at the 64th Consultation: (1) vala-**n**-**a**-*fusp*, consists of the human enzyme alpha-L-iduronidase **(-n-** for enzyme) fused to a fragment of the immunoglobulin G1-kappa anti-human insulin receptor **(-a-** for antibody); (2) teben-**t-a**-*fusp*, consists of human T cell receptor beta chain (**-t-** for T-cell receptor) fused to a fragment of the immunoglobulin anti-human CD3E **(-a-** for antibody); (3) onfe-**k-a**-*fusp*, consists of human tumour necrosis factor superfamily member 2 (**-k-** for cytokine) fused to a fragment of the immunoglobulin anti-human fibronectin extra domain B (**-a-** for antibody).

The scheme tries to tackle the challenge posed by highly complex biological substances that act through their different components where the identification of the primary component responsible for the final pharmacological action is often difficult to assign unequivocally. The scheme aims to give simple, short and easy-to-pronounce names, which assures a large reservoir of possibilities for the unique names that will be required to cover the expected explosion in numbers of such substances in the future. To compensate for the limited information conveyed by these names, there will be additional information in the description and definition of the substance accompanying the publication of the INN and to which stakeholders will be encouraged to refer as a source of information.

A final remark concerning the proteins group of biological medicines concerns the fact that a new scheme has been developed for fusion proteins but not for conjugates, including conjugated monoclonal antibodies, which should still be named based on a one or two-word scheme. Although chemical conjugation is different from the genetic conjugation that creates fusion proteins, from a pharmacological point of view, if the conjugated protein acts as a whole with both components having a pharmacological action and no cleavage is required to deliver the active component, there is no real reason not to use a one-word naming scheme for these, for example, for a monoclonal antibody conjugated to a toxin. This issue is very complex and needs further and careful thought before implementing a one-word scheme; however, with the exception of monoclonal antibody conjugates, it should probably be implemented for all other protein conjugates making the naming policy for fusion and conjugated proteins consistent.

3. Nomenclature for monoclonal antibodies

Today, monoclonal antibodies (mAbs) are the largest class of biological medicinal products. They can vary considerably in structure and function and can be used therapeutically for a very wide range of clinical indications as well as for *in vivo* diagnosis. They can also be conjugated with chemical or biological toxins for enhanced cytotoxic potency. Initially, mAbs for clinical use were of rodent sequence and derived by hybridoma technology [8]. However, application of recombinant DNA technology has allowed production of mAbs of any sequence including human sequences, non-natural sequences (e.g. chimeric and humanized mAbs) and mutated sequences that show therapeutic benefit [9]. Moreover, together with intact immunoglobulin structures, mAb products consisting of fragments have also been developed. A remarkably large number of mAb products of all types are currently at various stages of clinical development and many have been approved for clinical use in various countries (Fig. 1).

3.1. mAbs nomenclature scheme

INNs for mAbs are clearly needed and an INN was given many years ago (in 1989) to the first mAb approved for clinical use in the USA and EU. This was for a mouse sequence mAb against the CD3 antigen and was named *muromonab-CD3*. However, it was recognized quickly that this nomenclature system was not appropriate for future mAbs for a number of reasons, e.g. it breaks the rule of not having numerals in INN, and in the early 1990s a new system for INNs for mAbs was devised [10]. Monoclonal antibodies are a very heterogeneous class of biological medicines and, taking this into account, the naming convention adopted tried to reflect the source origin in the name, as well as the intended therapeutic target. Since then mAbs have been allocated an INN using a consistent although evolving nomenclature scheme. Up until the 64th INN Consultation in April 2017 [11] an INN for a mAb comprised a fantasy prefix, which contributes to a euphonious and distinctive name, followed by a substem A (also known as infix A), which indicates the target (molecule, cell or organ) class of the mAb, followed by a second substem, substem B (or infix B), which indicates the species on which the immunoglobulin sequence of the mAb is based, and finally by the stem *–mab* (Table 2). This stem is used for all molecules that contain an immunoglobulin variable domain that binds to a defined target, except those named using the *–fusp* stem. This includes intact immunoglobulins of all classes and species, fragments of mAbs and smaller molecules such as single-chain variable fragments (scFv). This nomenclature scheme has been used to allocate INN to over 500 mAbs. Recently, however, concerns have been raised regarding the above nomenclature scheme for mAbs [12]. Firstly, safe INN have to be distinct, not too long, pronounceable, euphonious in all United Nations languages (English, French, Spanish, Arabic, Chinese, Russian) and clear in script and print (INN are also translated in Latin and published in alphabetical order accordingly). INN applications for mAbs has increased greatly over the years, with the outcome that identifying new INN for mAbs that are distinct, free of conflict and not too long has become exceedingly difficult and a concern for both the WHO INN Programme and National nomenclature bodies. Secondly, sub-stem B, indicating the origin and similarity of a mAb to human species, was being interpreted as an indicator of potential undesirable immunogenicity of a mAb. However, there is limited scientific evidence for this and an association between immunogenicity and the source, including sub-stem B, has never been confirmed by data from large scale clinical trials. Furthermore, the USAN Programme has alerted the INN Programme that a particular antibody candidate might be claimed as humanized or chimeric as part of efforts to gather information that will be used to determine which antibody candidates should be advanced in development. This is a major concern as such developments are not oriented towards the more effective and safer mAbs, but towards obtaining the ‘best-selling’ name.

In view of these concerns, the INN Expert Group decided to revise the mAb INN nomenclature scheme and following a broad consultative process during INN Consultations and through *ad hoc* meetings with third parties and national nomenclature bodies, the Expert Group, at the 64th INN Consultation in April 2017, recommended to discontinue substem B, the ‘source’ infix. This will ease the difficulty in creating new INN for mAbs free of conflicts and not liable to be confused with names already in use and could also allow a longer fantasy prefix, which should lead to greater diversity in possible INN for mAbs. In some cases, it may be necessary to alter the target infix to avoid

Table 2
Former Nomenclature scheme for mAbs.

Prefix:	Substem A: target class	Substem B: species	Stem:
random	**Up to Proposed INN List 102:**		*-mab*
	-ba(c)-bacterial	*-a*-rat	
	-ci(r)-cardiovascular	*-axo-* rat-murine hybrid (pre-sub-stem)	
	-fung-fungal	*-e*-hamster	
	-ki(n)-interleukin (pre-sub-stem)	*-i*-primate	
	-le(s)-inflammatory lesions	*-o*-mouse	
	-li(m)-immunomodulator	*-u*-human	
	-os-bone	*-xi*-chimeric	
	-vi(r)-viral	*-zu*-humanized	
	tumours:		
	-co(l)-colon		
	-go(t)-testis		
	-go(v)-ovary		
	-ma(r)-mammary		
	-me(l)-melanoma		
	-pr(o)-prostate		
	-tu(m)-miscellaneous		
	From Proposed INN List 103 up to Proposed INN List 117:		
	-b(a)-bacterial	*-a*-rat	
	-am(i)-serum amyloid protein (SAP)/amyloidosis (pre-substem)	*-axo*-rat-mouse (pre-substem)	
	-c(i)-cardiovascular	*-e*-hamster	
	-f(u)-fungal	*-i*-primate	
	-gr(o)-skeletal muscle mass related growth factors and receptors (pre-substem)	*-o*-mouse	
		-u-human	
		-vet-veterinary use (pre-substem)	
		-xi-chimeric	
	-k(i)-interleukin	*-xizu*-chimeric-humanized	
	-l(i)-immunomodulating	*-zu*-humanized	
	-n(e)-neural		
	-s(o)-bone		
	-tox(a)-toxin		
	-t(u)-tumour		
	-v(i)-viral		

Table 3
Revised Nomenclature scheme for mAbs adopted at the 64th INN Consultation in April 2017.

Prefix:	Infix: target class		Stem:
random	*-ami-*	serum amyloid protein (SAP)/amyloidosis *(pre-substem)*	*-mab*
	-ba-	bacterial	
	-ci-	cardiovascular	
	-de-	metabolic or endocrine pathways and related targets	
	-fung-	fungal	
	-gros-	skeletal muscle mass related growth factors and receptors *(pre-substem)*	
	-ki-	interleukin	
	-li-	immunomodulating	
	-ne-	neural	
	-os-	bone	
	-ta-	tumour	
	-toxa-	toxin	
	-vet-	veterinary use *(pre-stem)*	
	-vi-	viral	

confusion between the old and new INN nomenclature schemes, for example -t(u)- (for tumour targets) is no longer used and is replaced by -ta-, as the 'u' may be misinterpreted as indicating an antibody of human origin [13]. Table 2 shows the naming scheme used for INN up to Proposed List 117; Table 3 shows the new naming scheme for mAbs, with first usage of this new scheme occurring during the 64th INN Consultation. However, as INN assigned pre-the 64th Consultation are not incorrect from the scientific viewpoint they will not need to be changed to names formed using the new system and they should remain as they are. Yet, it should be said that the concept of target, even in the new scheme, could possibly be reviewed and redefined to tackle the problem of the reposition of a marketed mAb in a therapeutic area different from the one originally reflected when the INN was assigned. As an example, *rituximab* has the substem *-tu-* for tumour while it is now also used for the treatment of rheumatoid arthritis.

For bispecific monoclonal antibodies, INN are assigned based on the primary target indicated by the applicant. Conjugated mAbs will continue to be assigned a two-word name although it is recognized that potential risks of drug name confusion leading to medication errors exist between the antibody and the antibody-drug conjugate as well as between antibody-drug conjugates with either the same antibody or the same drug/toxin.

Indeed, in 2013, following reports of a small number of patients inadvertently receiving *trastuzumab emtansine* instead of *trastuzumab*, regulators and safety authorities including Medsafe, the FDA, and Health Canada alerted health professionals of potential risks associated with confusion between these two look-alike sound-alike INNs, and advised prescribers to always use the respective brand names, Kadcyla™ and Herceptin™, to help reduce medication errors [14–16]. Moreover, the FDA also decided to modify the approved INN of the conjugated antibody to *ado-trastuzumab emtansine* to make it more distinguishable from the naked antibody. An analysis of the topic, based on available data, has been published by WHO [17]. The major conclusion was that under reporting of medication errors and adverse effects made it impossible to determine effectively the level of occurrence of these events and the true incidence of errors between *trastuzumab* and *trastuzumab emtansine* remains unknown. Moreover, the use of the prefix '*ado*' was considered not to resolve unequivocally the potential conflicts and, at the same time, posed challenges for new conjugates or monoclonal antibodies. Overall, therefore, the INN Expert Group took the decision to maintain a two-word INN for conjugated antibodies given that continuous implementation by health care providers of standard risk mitigation strategies will also be key in preventing medication errors.

4. Nomenclature for advanced therapy substances

The advent of advanced therapeutic medicinal products, such as cell and gene therapies, has posed several dilemmas to the INN Programme. Indeed, the INN Expert Group was required to consider whether these therapies should be identified with an INN, whether there were sufficient benefits for stakeholders (e.g. regulatory agencies, prescribers, manufacturers and patients) and whether a rigid naming scheme was required. These issues were extensively discussed through the years in the INN Expert Group, interacting with multiple experts and stakeholders, before agreeing to provide these therapies with an INN and devising a naming scheme.

Indeed, the first and foremost challenge in attributing an INN to advanced therapy substances, and cell therapies in particular, is the difficulty in unequivocally characterizing these products. This issue translates into difficulties in the ability of the Programme to issue a name to a single, imprecisely characterized unique product. Furthermore, at present it is difficult to envisage follow-on generic products for some advanced therapy substances, for example cell therapies, and therefore an INN may be seen as identifying a unique manufacturing process, thereby denying one of the principal strongpoints of INN, i.e. independence from production. While this is currently the situation, with technological advances this may not hold true in the future.

On the other hand, the INN Programme recognized that a non-proprietary name for a medicinal substance allows the divulgence of scientific data concerning the substance without reference to brand names that may differ from country to country and denote commercial partialities to specific drug products, facilitates trans-border trade as well as allowing these substances to fit into mature schemes for drug

evaluation, prescription and use. Lastly, the INN Programme had to consider that different regulatory environments had already put in place naming schemes for these therapies, and that the proliferation of different non-proprietary names referring to the same compounds would have been of hindrance to scientific disclosure, would have created confusion and may have led to prescription errors.

In conclusion, weighing the pros and cons, the INN Expert Group decided that the benefits in having an INN for advanced medicinal active substances outweighed any possible down-side. Also, given that different legislating agencies provide alternative definitions for what comprises advanced therapy medicinal products, the INN Expert Group decided to use the term 'advanced therapies' to cover substances for gene therapy, cell therapy, cell-based gene therapy and virus-based therapy [5,18]. These can be classified and regulated differently in different regulatory jurisdictions with no contradiction.

An overarching nomenclature guideline for advanced therapy active substances was recently issued by the INN Programme Expert Group, which provides rules for naming and defining advanced therapies, ranging from those that can be defined by a specific DNA/RNA sequence to less defined cell therapies [18]. This new guidance stems from previous guidelines that defined nomenclature schemes for single product types, the first of which was adopted in 2005 for gene therapy active substances. The current general scheme for naming these substances is depicted in Tables 4–7. The scheme attempts to clarify how specific advanced therapy substances should be named and provides for suffixes and infixes to be used in creating the INN; these are not necessarily all encompassing and will be supplemented with more suffixes and infixes if necessary. It should be noted that the adoption of an INN leads to a definition of the substance which is published in the INN list that is accessible via the hyperlinks in the INN website. This holds true for advanced therapies and it is the ambition of the INN Expert Group to provide minimal descriptive infixes within an adopted INN but then link it to a comprehensive description of the drug substance, thereby making names less scientifically descriptive of the drug substance but shorter and more euphonic.

4.1. Gene therapies nomenclature scheme

4.1.1. Viral, bacterial and plasmid-based gene therapy

The INN nomenclature scheme for gene therapies first adopted in 2005 was updated in 2016. The scheme covers substances based upon recombinant nucleic acid sequences involving viral and bacterial vectors and plasmid DNA. It involves a two-word system where the first word indicates the relevant gene component while the second word identifies the vector component.

The suffix for the first word is *-gene* with a preceding infix identifying the gene of interest using, when available, existing infixes for biological products, e.g. *-ermin-* for growth factors and *-stim-* for colony stimulating factors (Table 4). Suffixes proposed for the second word are *-vec* for non-replicating viral vectors, *-repvec* for replicating viral vectors, *-bac* for bacterial vectors and *-plasmid* for plasmid vectors (Table 5). In most cases, conditionally replicative viral vectors will use the suffix *-repvec*. The preceding infix identifies in more detail the nature of the vector. Thus, several infixes are proposed for different viral vectors such as *-adeno-* for adenovirus, *-erpa-* for herpes virus, *-lenti-* for lentivirus and *-parvo-* for parvovirus. The full list of proposed infixes for viral vectors and for the infixes for the genes of interest is shown in Tables 4 and 5

Both Word 1 and Word 2 will have fantasy prefixes that are random to contribute to a euphonious and distinctive name. Given that the scheme yields long two-word names, the INN Expert Group has given consideration to revising it. Yet, given that the scheme had been running for a number of years and that therapies will necessarily be used in very specialized settings, it was decided to give continuity to the scheme and not to change it.

Table 4
Nomenclature scheme for the GENE component (WORD 1) of: • Viral, bacterial and plasmid vectors • cell vectors.

Prefix	Infix		Suffix
Random, to contribute to euphonious and distinctive name	to identify the gene using, when available, existing infixes for biological products or using similar infix as for the protein for which the gene codes, e.g.:		-(a vowel) *gene* e.g. *-(o) gene*
	-cima-	cytosine deaminase	
	-ermin-	growth factor	
	-kin-	interleukin	
	-lim-	immunomodulator	
	-lip-	human lipoprotein lipase	
	-mul-	multiple gene	
	-stim-	colony stimulating factor	
	-tima-	thymidine kinase	
	-tusu-	tumour suppression	

Table 5
Nomenclature scheme for the VECTOR component (WORD 2) of viral, bacterial and plasmid based gene therapies.

Prefix	Infix		Suffix
Random, to contribute to euphonious and distinctive name	to identify the viral vector type, e.g.:		*-vec* (non-replicating viral vector) *-repvec* (replicating viral vector)
	-adeno-	adenovirus	
	-cana-	canarypox virus	
	-foli-	fowlpox virus	
	-erpa-	herpes virus	
	-lenti-	lentivirus	
	-morbilli-	Paramyxoviridae morbillivirus	
	-parvo-	adeno-associated virus (Parvoviridae dependovirus)	
	-retro-	other retrovirus	
	-vaci-	vaccinia virus	
	to identify the bacterial vector type, e.g.:		*-bac* (bacteria vector)
	-lis-	*Listeria monocytogenes*	*-plasmid* (plasmid vector)

In the case of substances for gene therapy based on non-plasmid DNA, there is no need for a second word in the name.

4.1.2. Cell-based gene therapy

Initially, the INN Experts took the approach that where a viral vector is to be used in *ex vivo* genetic modification of autologous cells, this was gene therapy and not cell therapy, and an INN was provided to the viral vector according to the established nomenclature scheme for viral vectors. This approach was in contrast to that of the USAN which considered such therapies to be cell therapies. In 2016, a harmonized approach between INN and USAN was agreed whereby cells (autologous and allogeneic) being developed as therapeutic substances and that had been genetically modified *in vitro* or *ex vivo*, would be classed as cell therapies and a naming scheme for them distinct from that for non-genetically modified cells was agreed. This harmonized scheme for cell-based gene therapies comprises a two-word scheme in which the first word would identify the relevant gene of interest (following the same approach as for viral, bacterial and plasmid-based gene therapies) and the second word would identify the cell component. This therefore identifies this therapeutic approach via a unique scheme but formally links these therapies to both cell and gene therapy.

The first word is developed in exactly the same way as for gene therapies (Table 4). The suffix would be *-gene*, whilst the gene would be identified making use of the existing infixes for biological products or using a similar infix as for the protein for which the gene encodes. The derivation of the second word, defining the cell component, would

Table 6
Nomenclature scheme for the CELL component of: • cell therapies – SINGLE WORD •cell-based gene therapies – WORD 2.

Prefix	Infix 1: manipulation/s[a]	Infix2: cell type	Suffix
All cell therapies; allogeneic cell-based gene therapies: Random, to contribute to euphonious and distinctive name **Autologous cell-based gene therapies:** *Auto-*	to specify, if appropriate, which manipulation the cells have undergone, using, when available, existing infixes for manipulation[b], e.g.: *-fus* - fusion to a cell	to identify the primary cell type[c] using, when available, existing infixes for cell types[d]	*-cel* (cell)

Note: Information concerning manipulation and/or modification, and the type of the cell-based therapy (i.e. allogeneic, autologous and xenogeneic), will be specified in the description of the product.
[e]The cell type infix *-leu-* is used to describe hematologic cell preparations that do not fit a particular or specific cell type. Such cell preparations may be comprised of a mixture of the various blood cell elements, a subset of blood elements such as T- B- or NK-cells, or antigen presenting cells (APCs) that do not fit the definition of dendritic cells fall into this category.
[a] There may be more than one manipulation infix in the same INN, but this should be avoided if possible to avoid overly long names.
[b] In the case of manipulation such as cell expansion and cell activation (with cytokines/drug, etc.), there is no need for an infix; this kind of manipulation will be specified in the description.
[c] Residual cells not expected to contribute to the intended function, are not named.
[d] ***-co(n)-*** for *chondrocytes*; ***-cor-*** for *umbilical cord cells*; ***-defitem-*** for *differentiated stem cells* (not filling in any existing category); ***-den-*** for *dendritic cells*; ***-end(o)-*** for *endothelial cells*; ***-ep(a)-*** for *hepatocytes*; ***-fi(b)-*** for *fibroblasts*; ***-isle-*** for *islet cells*; ***-ker(a)-*** for *keratinocytes*; ***-leu-*** for *lymphocytes/monocytes/APC (white cells) (*; ***-mestro-*** for *mesenchymal stromal cells (msc)*;***-mio(b)-*** for *myoblasts*; ***-ova-*** for *ovary cells*; ***-pla(c)-*** for *placenta cells*; ***-ren-*** for *renal tubular cells*; ***-ret-*** for *retinal epithelial cells*; ***-tem-*** for *stem cells*; ***-tesi-*** for *testis cells*; ***-tu-*** for *tumour cells*; ***-ur-*** for *urothelial cells*.

Table 7
Nomenclature scheme for virus-based therapies.

Prefix	Infix 1: virus type		Infix 2:	Suffix
Random, to contribute to euphonious and distinctive name	*-adeno-*	adenovirus	*-tu-* for tumoricidal	*-rev* (therapeutic virus)
	-cana-	canarypox virus		
	-foli-	fowlpox virus		
	-erpa-	herpes virus		
	-lenti-	lentivirus		
	-morbilli-	Paramyxoviridae morbillivirus		
	-parvo-	adeno-associated virus (Parvoviridae dependovirus)		
	-retro-	other retrovirus		
	-vaci-	vaccinia virus		

follow the same procedure as for cell therapies (see below), viz. the suffix would be *-cel*, infix 1 would specify which manipulation the cells had undergone, if any, using when available existing infixes, e.g. *-fus-* for fusion, and infix 2 would identify the primary cell type, e.g. *-tem-* for stem cells (Table 6). As with cell therapies, additional information concerning manipulation or modification and the type of cell therapy, i.e. autologous, allogeneic or xenogeneic, would be specified in the Description/Definition. A further modification of this scheme was made in 2018 specifically for autologous cell-based gene therapies. Where identical autologous cells (to the extent that autologous cells can be defined as 'identical') are transduced with distinct gene constructs, the nature of and the need for a unique fantasy prefix for the second word was questioned. Given the nature of autologous cells, a consensus was reached between experts from INN and the US FDA, and ratified by the INN Expert Group at the 67th INN Consultation, to omit the fantasy prefix from the second, cell-component, word and replace it with *auto-*. The requirement for a fantasy prefix on the first word would remain and contribute to a unique INN. For INN for all other types of gene therapy substance, including allogeneic cell-based gene therapies, both words will continue to begin with a random fantasy prefix to create a euphonious and distinctive name.

4.2. Cell therapies nomenclature scheme

Prior to the formal adoption of a nomenclature scheme for cell therapies, discussions were held between the INN and USAN (United States Approved Name Council). USAN had already established a cell therapy nomenclature scheme and talks were held in order to agree a harmonized scheme acceptable to both the INN and USAN. This INN-USAN-harmonized nomenclature scheme for substances for cell therapies was formally approved by the members of the INN Expert Group in 2016. The scheme names all non-genetically modified substances for cell therapies, with the exception of minimally manipulated hematopoietic elements and combinations of substances, which are not named.

INN for cell therapies comprises a single word beginning with a random fantasy prefix and ending with the suffix *-cel* (Table 6). Between the fantasy prefix and the *-cel* suffix can be two infixes. Infix 1 will specify if appropriate any manipulation the cells have undergone, again using when available existing infixes, e.g. *-fus-* for fusion. In the case of manipulation such as cell expansion and cell activation (with cytokines/drug, etc.), there is no need for an infix and this kind of manipulation will be specified in the description. There may be more than one manipulation infix in the same INN, but this should be avoided where possible to avoid overly long names. Indeed, it is the desire of the INN Expert Group to avoid overly long and complicated names for cell therapies (and other substances) and it should be noted that all information concerning manipulation and/or modification, and the type of the cell-based therapy (i.e. allogeneic, autologous and xenogeneic), will be specified in the description of the product.

Infix 2 identifies the primary cell type using when available existing infixes for cell types, e.g. *-den-* for dendritic cell and *-tem-* for stem cells. A full list of proposed infixes for cell types is provided. Residual cells that are not expected to contribute to the intended function of the cell therapy drug substance will inevitably be present but are not named.

Before this scheme came into effect, some cell therapies were named by the INN program or by USAN using different approaches. Given that an adopted INN can be modified only under exceptional circumstances, cell therapy INN selected before the adoption of the present nomenclature scheme may have followed different rules, but will remain unchanged.

4.3. Virus-based therapies

To date, all proposed and recommended INN for virus-based therapies concern so-called oncolytic viruses, viruses that will specifically, or at least preferentially, infect and destroy tumour cells with minimal effect on normal tissues. Most oncolytic viruses have been modified to enhance their selective replication on cancer cells. The nomenclature scheme for virus-based therapeutic substances involves one word, beginning with a random fantasy prefix and ending with the

suffix *-rev*, for therapeutic virus (Table 7). The word would also contain two infixes: infix 1 would define the basic virus type, using the same infixes as for the second word in the viral vectors nomenclature scheme, e.g. *-adeno-* for adenoviruses and *-lenti-* for lentiviruses (Table 7). Currently, only one infix 2 is proposed, *-tu-* for tumouricidal. Alternative infixes may be proposed as and when necessary.

Examples of oncolytic viruses with INN include *tasadenoturev*, a modified, conditionally replicating adenovirus, and *canerpaturev*, a replication-competent, spontaneously occurring mutant of herpes simplex virus type 1. Initially, a few oncolytic viruses were named with the slightly longer infix 2 *-tuci-*, e.g. *enadenotucirev*, a live adenovirus selected for optimal cancer killing properties.

Oncolytic viruses that have been engineered to express a protein that augments the anti-cancer effect, e.g. GM-CSF, are being given a two-word name according to the gene therapy nomenclature scheme; thus, *pexastimogene devacirepvec* is a genetically engineered vaccinia poxvirus expressing GM-CSF being developed as a broad spectrum anti-tumour virus.

5. Vaccines and vaccine-like substances

5.1. Vaccines

Historically, vaccines, which were difficult to characterize, did not receive an INN but received a descriptive name from a separate committee of the World Health Organization, the Expert Committee on Biological Standardization (ECBS). As this procedure has shown no weaknesses since it was adopted, the INN Expert Group reiterated that it has no intention to issue INNs to such vaccines. In addition, given that a vaccine is usually a drug product, characterized by the presence of adjuvants and/or by the presence of multiple active substances, it is unlikely that an INN could be used for a vaccine, although it may reside in its description.

Traditional vaccines consist of whole killed pathogens, live attenuated pathogens, subunits (antigens) derived from pathogens, or inactivated pathogenic toxins. With the advent of recombinant DNA technology novel approaches for the development of vaccines against infectious diseases were developed including recombinant DNA expressed protein antigens, recombinant DNA derived virus-like particles, recombinant live vectors expressing heterologous antigens, and DNA/RNA vaccines. During the INN Consultation in 1993, it was agreed that recombinant protein vaccines may fulfil the requirements of being defined and homogeneous substances and so could be assigned INN, and this position remains [1]. Several recombinant protein vaccines are now on the market, with many under development, but there have been no requests for INN for the recombinant proteins used as active substances in them; nevertheless, these could be assigned INN upon request. Defined recombinant nucleic acids used as active substances in vaccines, whether of biological or synthetic origin, could similarly be assigned INN. Indeed, the active substance of a prophylactic RNA vaccine against rabies virus based upon an mRNA molecule encoding the rabies surface glycoprotein was assigned the INN *nadorameran*.

The advanced therapy nomenclature scheme for gene therapy substances infers that it is possible to adopt an INN in cases in which a complex drug substance is defined by a DNA/RNA sequence. This implies that viruses and bacteria, live or non-replicating, and genetically modified or not, that comprise the active substance of a prophylactic vaccine could be assigned INN. Again, though, there have been no requests for INN for such vaccines, and there is no intention to assign them INN, with common names continuing to be provided by ECBS. However, several recent INN applications have fallen into a grey area of 'vaccine-like' substances.

5.2. Vaccine-like substances – cancer vaccines

In addition to vaccines against infectious disease, the term vaccine is also being applied to other medicinal substances such as 'cancer vaccines' typically containing a tumour antigen with the intention of stimulating the immune system to attack and destroy the tumour, and several 'vaccine-like' substances for anti-cancer immunotherapy have been assigned INN.

Thus, a recombinant fusion protein comprising the BCG heat-shock protein HSP 65 fused to transcription factor E7 of human papilloma virus (HPV) 16 for treatment of cervical cancer was assigned the INN *verpasep caltespen* in 2005. More recently, a genetically modified, live attenuated strain of *Listeria monocytogenes* developed for immune stimulation against HPV 16 E7 protein-expressing cells was assigned the INN *axalimogene filolisbac*, while a therapeutic DNA vaccine expressing the E6 and E7 antigens of human papillomavirus types 16 and 18 for treatment of HPV induced cancers was assigned the INN *tirvalimogene teraplasmid*. In these latter two cases two word names were assigned according to the nomenclature scheme for gene therapy substances.

5.3. Peptides (vaccines)

Another approach in vaccine technology is the development of peptide vaccines (epitopes involved in immune response formation). Since these peptides are chemically well-defined, they fall within the INN naming system. Peptides in general are given INN ending with the stem *-tide* while immunomodulating peptide 'vaccines' are included in the stem *-motide*. However, the peptides so far named with the *-motide* stem have immunomodulatory activity and are not true vaccines containing microbial-derived antigens that stimulate an adaptive immune response, i.e. they are more frequently referred to as cancer vaccines. Some synthetic peptides comprise all or part of a tumour antigen.

6. Conclusions

The science of drug nomenclature is a continuous evolving field as it must cope with the constant development of novel therapeutics both in terms of structure and/or mode of action. Therefore, existing policies must be constantly adapted and new ones developed, making the creation of INN a stimulating process. Although this is certainly true for all classes of drugs, there is little doubt that nowadays it is particularly challenging for biological substances that are increasingly becoming the predominant products, as is indicated by the observation that they now represent more than 50% of all new applications handled at recent INN Consultations. Amongst them, monoclonal antibodies are certainly the major class but both advanced therapies and fusion proteins are also constantly growing in numbers. Two major challenges are posed by biologicals in assigning INNs: (i) the need to develop naming schemes that produce INNs capable of conveying the information on the structure and mode of action of very complex biological substances and (ii) the boundaries provided by linguistic considerations, such as the length of a name, the maximum number of syllables that it can contain to remain clearly pronounceable, and the issue of a one-word versus two-word naming scheme. All this is focused on avoiding or minimizing errors in prescription, to guarantee the efficient traceability for pharmacovigilance and, last but not least, to accommodate the large number of substances belonging to the same category while still giving each of them a unique and distinguishable name.

A possible strategy to overcome such challenges might be to develop a scheme that consists of more than one and possibly long words. This would satisfy the desire to increase the information carried by the INN, but at the cost of increasing the risk of errors in prescriptions, creating names that are unacceptably difficult to pronounce and, perhaps even more relevant, limiting in the future the identification of names for similar substances that are sufficiently different from one another, a situation that occurred with monoclonal antibodies. A two-word naming strategy is followed in the naming scheme for several advanced therapies. A second approach, the one chosen for fusion proteins and to an extent also in the new scheme for naming monoclonal antibodies, is

to have simple, short and easy to pronounce names that would provide a good reservoir of unique variants covering the expected upsurge of such substances in the future, but carrying limited pharmacological information. The compensation for the lack of information carried by shorter INN would reside in a detailed description and definition of the substance reported in the full publication that accompanies each recommended INN, to which all stakeholders should refer and use more extensively. As presented in this review, the new policies developed for advanced therapies and for fusion proteins provide examples of different strategies that have been developed to give an adequate response to the challenge of complexity featuring the biological substances. Which type of approach is more appropriate and responds better to the needs of all stakeholders in the INN arena, will be a subject of continuous discussion and will require time for the analysis and evaluation of their implementation. Certainly, further adaptations and new schemes will be required to keep pace with the wealth of novelty produced by drug developers.

Disclosure statement

The authors declare no conflicts of interest and are responsible for the content and writing of this article. The authors are members of the INN Expert Group and one is a WHO staff member of the World Health Organization. The authors alone are responsible for the views expressed in this publication and they do not necessarily represent the decisions or policies of the World Health Organization.

Acknowledgments

The Authors would like to thank Dr Sofia De Sousa Guimaraes Koch, Dr Sophie Andree Jeanne Lasseur and Dr Antonio Romeo (INN Programme, WHO) for their essential assistance and technical contribution.

References

[1] Guidance on the use of international non-proprietary names (INNs) for pharmaceutical substances. 2017www.who.int/medicines/services/inn/FINAL_WHO_PHARM_S_NOM_1570_web.pdf?ua=1.

[2] General principles for guidance in devising international nonproprietary names for pharmaceutical substances. 13th consultation on nonproprietary names for pharmaceutical substances. 2013. (PHARM S/NOM 928). www.who.int/medicines/services/inn/GeneralprinciplesEn.pdf.

[3] The use of stems in the selection of international nonproprietary names (INN) for pharmaceutical substances 2018 (StemBook 2018). 2018www.who.int/medicines/services/inn/StemBook_2018.pdf?ua=1.

[4] Addendum to the use of stems in the selection of International Nonproprietary Names (INN) for pharmaceutical substances. 2019. "WHO/EMP/RHT/TSN/2018.1/INN Working Document 19.450 www.who.int/medicines/services/inn/Addendum_StemBook2018_201902.pdf?ua=1.

[5] International Nonproprietary Names (INN) for biological and biotechnological substances (a review). 2019. WHO/EMP/RHT/TSN/2019.1 www.who.int/medicines/services/inn/BioReview2019.pdf?ua=1.

[6] Schiestl M, Stangler T, Torella C, Cepeljnik T, Toll H, Grau R. Acceptable changes in quality attributes of glycosylated biopharmaceuticals. Nat Biotechnol 2011;29:310–2.

[7] 63rd consultation on international non-proprietary names for pharmaceutical substances. Executive summary. 2017. INN Working Doc. 17.408 www.who.int/medicines/services/inn/63rd_Executive_Summary.pdf.

[8] Kohler G, Milstein C. Continuous cultures of fused cells secreting antibody of predefined specificity. Nature 1975;256:495–7.

[9] Liu JK. The history of monoclonal antibody development - progress, remaining challenges and future innovations. Ann. Med. Surg. (Lond.) 2014;3:113–6.

[10] international non-proprietary names (INN) working group meeting on nomenclature for monoclonal antibodies (mAb). 2008. INN Working Doc. 08.24 www.who.int/medicines/services/inn/ApprovedFinalWHOINNWGroupMeetingNMAreport.pdf?u=1.

[11] 64th consultation on international non-proprietary names for pharmaceutical substances. Executive summary. 2017. INN Working Doc. 17.419 www.who.int/medicines/services/inn/64th_Executive_Summary.pdf?ua=1.

[12] Jones TD, et al. The INNs and outs of antibody non-proprietary names. mAbs 2016;8:1–9.

[13] Revised monoclonal antibody (mAb) nomenclature scheme. 2017. INN Working Doc. 17.416 www.who.int/medicines/services/inn/Revised_mAb_nomenclature_scheme.pdf?ua=1.

[14] Medsafe. Preventing medication errors: Kadcyla and Herceptin. Prescriber update 2013;34:43www.medsafe.govt.nz/profs/PUArticles/December2013PreventMedErrorsKadcylaAndHerceptin.htm.

[15] U.S. FDA. Kadcyla (ado-trastuzumab emtansine): drug safety communication – potential medication errors resulting from name confusion. 2013www.fda.gov/Safety/MedWatch/SafetyInformation/SafetyAlertsforHumanMedicalProducts/ucm350817.htm.

[16] Health Canada. Kadcyla (trastuzumab emtansine) and Herceptin (trastuzumab) – potential risk for medication error due to name confusion. Dear Healthcare Professional Letter; 2013www.healthycanadians.gc.ca/recall-alert-rappel-avis/hc-sc/2013/36173a-eng.php.

[17] Pepper S, Malal S, Mignot G, Balocco R. Look-alike sound-alike drug name confusion: trastuzumab emtansine. WHO Drug Inf 2016;30:29–34.

[18] Nomenclature schemes for advanced therapies (substances for gene therapies, substances for cell therapies, substances for cell-based gene therapies and virus-based therapies) INN Working Doc. 17417. 2017www.who.int/medicines/services/inn/Nomenclature_schemes_advanced_therapies_201707.pdf?ua=1.