Challenges of Trustable AI and Added-Value on Health
B. Séroussi et al. (Eds.)

doi:10.3233/SHTI220478

Aggregations of Substance in Virtual Drug Models Based on ISO/CEN Standards for Identification of Medicinal Products (IDMP)

Robert VANDER STICHELE[a,b,1] and Dipak KALRA[b]
[a] *Ghent University, Belgium*
[b] *European Institute for Innovation through Health Data, Belgium*

Abstract. In this study representation of chemical substances in IDMP is reviewed, with an exploration of aggregation levels for substance used in the virtual drug data models of RxNorm, SNOMED-CT, ATC/INN, and the Belgian SAM database, for products with a single substance and combinations of substances. Active moiety and available solid states forms are explored for diclofenac, amoxicillin, carbamazepine, amlodipine, with regard to their representation in coding systems such as WHODrug, SMS, UNII, CAS, and SNOMED-CT. By counting the number of medicinal products in Belgium for amlodipine in each level of aggregation, concepts for grouper of substances and two levels of grouper of medicinal products are illustrated. Recommendations are made for the further development of IDMP and its link to international drug classifications.

Keywords. Pharmaceutical preparations, active ingredient, ontology, semantic interoperability, standardization, RxNorm, Anatomical Chemical Therapeutic Classification, SNOMED-CT, International Non-proprietary Name, Identification of medicinal Product (IDMP).

1. Introduction

Precise identification of active substances in pharmacotherapeutic drugs is not an easy task. Historically, the majority medicines contained chemical substances. But, over the years, more and more new products are biologicals and other more complex substances.[1] Medicinal drugs may be either substances as their free base or acid or in a solid state form e.g., salt or co-crystal. (. Drug developers take decisions when choosing a solid state form for several reasons, such as solubility or absorbability. but it is not always clear from the labeling what exactly was chosen. Regulatory authorities, when approving chemical drugs should precisely describe the substance which does not always occur when considered clinically irrelevant. Nevertheless, the distinction of the solid state form has implications for the expression of strength of the product. Therefore, the ISO/CEN standards for identification of medicinal products (IDMP) require precise

[1] Corresponding Author, Robert Vander Stichele, Unit of Medical Informatics, Department of Public Health, Ghent University, 3K3, Depintelaan, 185, 9000 Gent, Belgium; E-mail: robert.vanderstichele@ugent.be.

identification of the active moiety and its solid state [2]. There are many coding systems to represent active moieties and their solid state forms, such as the International Non-proprietary Nomenclature (INN) from the World Health Organisation, the Chemical Abstract Service (CAS) Registry; Pubchem; DrugBank; SNOMED-CT; WHODRUG; and the substance registries of the FDA and EMA. The EU, with the support of FDA in various aspects, is currently engaged in a major cleansing operation, with systematic revision of over 30.000 registered substances with therapeutic activities, starting 2020, with a scheduled endpoint in 2022. As the European Substance Management System (SMS) is part of the pharmacovigilance system, it contains information on terms used to describe active substances and excipients actually reported in Individual Case Study Reports (ICSR) in authorized medicinal products from European and other countries. This provides valuable input for the central registry about variations in naming of substances (spelling errors, synonyms) and existence of solid state forms, so that correct and comprehensive cleansing can be performed. The aim of this study is 1) to illustrate the complexity of representation of chemical substance with 4 substances (diclofenac, amoxicillin, carbamazepine, amlodipine), 2) to explore the aggregation levels for substance used in the virtual drug data models of RxNorm, SNOMED-CT, ATC/INN, and the Belgian SAM database; 3) to explore possible approaches for representation of combinations of substances in medicinal products.

2. Methodology

Medicinal products authorized in Belgium were collected for amlodipine, amoxicilline, carbamazepine, and diclofenac. Only medicinal products containing single substances were included. The active moiety was identified, and an overview of relevant modifiers, if any. Fully identified (chemical) substances were attributed the role of Precise Active Ingredients (PAI) [2], namely an active moiety without modifier or the full list of one or more active moieties with modifiers [3]. The coding numbers were collected from WHODrug, SMS, UNII, CAS and SNOMED-CT. For concepts involving substance in virtual models in the of INN, ATC, RxNorm, SNOMED-CT, WHODrug, and SAM, the level of substance and product aggregation was defined. Finally, the approach to combinations of substances was analysed for ATC, SNOMED-CT, and SAM.

3. Results

3.1. Identification of the attribute of Precise Active Ingredient (PAI) for 4 substances

Substances data from SMS (EMA) and G-SRS (FDA)[4] were used, to list the possible modifiers for the selected substances amlodipine, amoxicillin, carbamazepine, and

[2] This term PAI is taken from the SNOMED Medicinal Product Specification V4.0 (http://snomed.org/mpm) and defined as : The substance that provides the therapeutic effect of the medicinal product, described using the fullest and most specific description of the substance as it is used in the product(s) being represented. This may include various Solid state forms , such as salts, and/or solvates.

[3] Acknowledgements: we thank Ursula Tschorn, Leonora Grandia, Julie James, Annet Rozema, Inti van Eck, Malin Fladvad for their contribution to table 1, as part of the UNICOM Action Program www.unicom.org

[4] https://www.fda.gov/industry/fda-data-standards-advisory-board/fdas-global-substance-registration-system

diclofenac. A choice has still to be made for a global codification and numbering system of substances, and a mechanism to attribute the role of PAI, which is essential for the production of a global identifier, the Pharmaceutical Product Identifier (PhPID), as proposed in the IDMP standard.[5] In Table 1, we listed the active moieties and the substances with the attribute of PAI for 4 substances, with the different coding numbers from WHODRUG, SMS, CAS, SNOMED-CT[6], as collected by the UNICOM Pilot Product List group. These concepts describe physical realities, with specific characteristics such as solubility, weight, etc.

From the SAM database, the medicinal products authorized and currently available on the market were selected. For amlodipine, 28 medicinal products packs from 8 marketing authorization holders (MAH) were selected, for carbamazepine 4 from 1 MAH, and for diclofenac 50 from 10 MAHs. For diclofenac, 4 salts were identified (no medicinal product was identified as the free acid as active moiety. as PAI); for amoxicillin two solid state forms; for carbamazepine there was no solid state described in a product; for amlodipine, three salt forms were identified: besilate, maleate, mesilate (only the 2 formers are on the market), although a lot of salts and salt hydrate forms are known for the amlodipine base.

Table 1. List of Moieties and Precise Active Ingredients

	WHODrug	SMS	UNII	CAS	SNOMED-CT
Active Moiety					
diclofenac	00372301001	100000092272	144O8QL0L1	15307-86-5	7034005 | Diclofenac (substance) |
amoxicillin (anhydrous, explicitly)	00249601145	100000091596	9EM05410Q9	26787-78-0	785686003 | Amoxicillin anhydrous (substance) |
carbamazepine	00052501001	100000092127	33CM23913M	298-46-4	387222003 | Carbamazepine (substance) |
amlodipine	00972401001	100000085259	1J444QC288	88150-42-9	386864001 | Amlodipine (substance) |
(Modified) substances with the attribute of Precise Active Ingredient					
diclofenac (iononized)	00372301001	100000092798	144O8QL0L1	15307-86-5	7034005 | Diclofenac (substance) |
dlclofenac sodium	00372302001	100000092272	QTG126297Q	15307-79-6	62039007 | Diclofenac sodium (substance) |
dlclofenac potassium	00372304001	100000092368	L4D5UA6CB4	15307-81-0	108515008 | Diclofenac potassium (substance) |
diclofenac diethylamine (syn. diclofenac diethylammonium)	00372303001	100000091074	6TGQ35Z71K	78213-16-8	426714006 | Diclofenac diethylammonium (substance) |
diclofenac epolamine	00372307001	100000085789	X5F8EKL9ZG	119623-66-4	425650004 | Diclofenac epolamine (substance) |
amoxicillin sodium	00249603001	100000090113	544Y3D6MYH	34642-77-8	427483001 | Amoxicillin sodium (substance) |
amoxicillin trihydrate	00249602001	100000092629	804826J2HU	61336-70-7	96068000 | Amoxicillin trihydrate (substance) |
carbamazepine	00052501001	100000092127	33CM23913M	298-46-4	387222003 | Carbamazepine (substance) |
amlodipine besilate	00972402001	100000090079	864V2Q084H	111470-99-6	84976003 | Amlodipine besilate (substance) |
amlodipine mesilate	00972404001	100000089571	291Y33EZHA	246852-12-0	not present
amlodipine benzoate	00972410001	Not existing	XD75TQ8A2P	1239916-29-0	789067004 | Amlodipine benzoate (substance) |
amlodipine maleate	00972403001	100000089370	CQ27G2BZJM	88150-47-4	421048000 | Amlodipine maleate (substance) |

[5] Once the cleansing results/advice is implemented, the EU will offer the US the match between the UNII code and the SMSID (formerly known as EUTCT code) to be loaded into the US Substance database (GSRS). The SMSID will also be fed into the European SPOR (Substance, Product, Organisation, Reference) multilingual terminological system to support IDMP implementation.

[6] https://who-umc.org/whodrug/whodrug-global/ ; https://www.ema.europa.eu/en/human-regulatory/research-development/data-medicines-iso-idmp-standards/spor-master-data/substance-product-data-management-services. . https://www.cas.org/ . https://www.fda.gov/industry/structured-product-labeling-resources/uniis-preferred-substance-names-and-their-identified-synonyms

[7] Acknowledgements: we thank Ursula Tschorn, Leonora Grandia, Julie James, Annet Rozema, Inti van Eck for their contribution to table 1, as part of the UNICOM action Program www.unicom.org

3.2. Analysis of aggregation of precise active ingredients and products in virtual drug data models

To proceed to aggregation, a distinction must be made between a grouping of substances and a grouping of products. A grouping of substances is allocating a collection of all substances that share the same moiety. When the substance has no solid state form (e.g. Carbamazepine, this substance will be identified as PAI and its group will have only one member. When the substance has one or more solid state forms, the group will contain all possible salts, hydrates or salt-hydrate substances.
A grouping of products with the same substance is a collection of all medicinal products with the same substance. In a first level of aggregation, the low level group of products can collect all medicinal products with the same substance attributed as PAI, as foreseen in the IDMP standard for the Pharmaceutical Product Identifier Level 1 (coding system yet to be established within IDMP). This group concept is very close to the SNOMED-CT abstraction of Medicinal Product Precisely (MP precisely). In a second high level of aggregation for the group of products, all products with the same active moiety (or with the same group of substances) can be collated in a "virtual therapeutic moiety", as implemented in the Belgian SAM database and in the SNOMED-CT concept "Medicinal Product Only (MP only)[7].

It is important to note that the same term "amlodipine" is used to describe the active moiety, the allocator of amlodipine solid state forms, and the group of products by the same active moiety of substances, the latter two to be represented by a different coding system than for the active moiety. In Table 2, this concept of grouping illustrated by the number of Medicinal Products (packs) on the Belgian Market.

Table 2. Distribution over medicinal products in Belgium over combinations of substance, dose form, and combinations for amlodipine (ATC C08CA01)

				Belgium	
Virtual Medicinal Product Group (INN prescription)				**MPs**	**MMPs**
Amlodipine oral 5mg					
	Pharmaceutical Product (PHPID Level IV)				
	amlodipine besilate	capsule, hard	5 mg	1	2
	amlodipine besilate	tablet	5 mg	6	9
	amlodipine mesilate	tablet	5 mg	0	0
	amlodpine maleaat	tablet	5 mg	1	2
Amlodipine oral 10 g					
	amlodipine besilate	capsule, hard	10 mg	1	2
	amlodipine besilate	tablet	10 mg	7	9
	amlodipine mesilate	tablet	10 mg	0	0
	amlodipine maleate	tablet	10 mg	1	2
	amlodipine maleate	coated tablet	10 mg	1	2

INN International Non-Proprietary Name Nomenclature **PhPID** Pharmaceutical Product Ide
MP : Medicinal Product **MPP** Medicinal Product Pack **MAH** Marketing Authorization Holder

[7]One could argue that the INN nomenclature could act as a terminology for the allocator of the active substance, and in fact it is used in the ATC classification as such. However, this is not an explicit function of the INN nomenclature, and there are only internal, not very robust coding numbers available. In the US based RxNorm system of the National Library of Medicine, little attention is paid to the substance solid state forms in medicinal products. Only high level grouping of substances is used at the virtual level (Semantic Clinical Drug concept). The IDMP model does not provide sophisticated levels of higher aggregation of Medicinal Products by substance, except for PhPID-Level 1.

3.3. Analysis of approaches for combinations of medicinal products

There are basically 3 approaches to deal with combinations of substances in Medicinal Products. The approach of the ATC Classification is to have separate fifth level classes for products with a single substance and for products with combinations (sometimes specified and sometimes not or designated by drug class names). In the Belgian SAM, combinations of substances are entered in the database at the same level of the single products, are limited to 3 substances (labeled a+b+c plus in case of more than 3 active ingredients, and linked to a separate database of single constituting substances with the attribute of PAI. SNOMED-CT has a mixed approach with MP Only (all products containing amlodipine as a single or in a specified combination) as a subclass of MP (all products containing amlodipine). It is still unclear which approach will be adopted in IDMP for handling drug combinations.

4. Discussion and Conclusions

With this work a proposal is made for the consistent use of substance terms with the attribute of precise active ingredient(s), either the active substance or its solid state form to build the PhPIDs of medicinal products. Several coding systems exist to describe these physical realities, and a choice will have to be made between one of these coding systems, or a new global system created. In addition, a terminology will have to be created with a coding system for the concept of grouping of substance, preferably as an extension of the INN Nomenclature. For the low level aggregation of medicinal products based on PAI, IDMP will have to provide a coding system for PhPID-Level 1. For the higher level aggregation of medicinal products based on the grouping of substances, also a coding system will need to be determined.[3] This higher level of aggregation is needed to operationalize the concept of INN prescriptions, and to make a precise link with the ATC classification. This could enhance the possibilities for international Computerized Decision Support Systems and Drug Information Resources to link more easily to the different national medicinal drug dictionaries. Precise identification of the substance in chemical substances as either the moiety or the substance including its solid state form is crucial to determining the precise active ingredient, as foreseen in the IDMP standards.

References

[1] Peryea T, Southall N, Miller M, Katzel D, Anderson N, Neyra J, Stemann S, Nguyễn ĐT, Amugoda D, Newatia A, Ghazzaoui R, Johanson E, Diederik H, Callahan L, Switzer F. Global Substance Registration System: consistent scientific descriptions for substances related to health. Nucleic Acids Res. 2021 Jan 8;49(D1):D1179-D1185.

[2] ISO 11238:2018 Health informatics -- Identification of medicinal products -- Data elements and structures for the unique identification and exchange of regulated information on substances. Snomed-CT Medicinal Product Model Specification V4.0.

[3] Van Bever E, Wirtz VJ, Azermai M, De Loof G, Christiaens T, Nicolas L, Van Bortel L, Vander Stichele R. Operational rules for the implementation of INN prescribing. Int J Med Inform. 2014 Jan;83(1):47-56.