Evaluating the Relevance of Virtual Drugs

Yvon K. Awuklu, Vianney Jouhet, Sebastien Cossin, Frantz Thiessard, Romain Griffier, Fleur Mougin

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

doi:10.3233/SHTI220467

Evaluating the Relevance of Virtual Drugs

Yvon K. AWUKLU[a,1], Vianney JOUHET[a,b], Sébastien COSSIN[b], Frantz THIESSARD[a,b], Romain GRIFFIER[a,b] and Fleur MOUGIN[b]
[a] *Bordeaux Univ. Hospital, Public Health Unit, Medical Information Department, F-33000 Bordeaux, France*
[b] *Univ. of Bordeaux, Inserm UMR 1219, Bordeaux Population Health Research Center, team ERIAS, F-33000 Bordeaux, France*

Abstract. Information about drugs is numerous and varied, and many drugs can share the same information. Grouping drugs that have common characteristics can be useful to avoid redundancy and facilitate interoperability. Our work focused on the evaluation of the relevance of classes allowing this type of grouping: the "Virtual Drug". Thus, in this paper, we describe the process of creating this class from the data of the French Public Drug Database, which is then evaluated against the codes of the Anatomical Therapeutic Chemical classification associated with the drugs. Our evaluation showed that 99.55% of the "Virtual Drug" classes have a good intra-class consistency.

Keywords. drug, virtual drug, semantic interoperability

1. Introduction

Information about drugs is numerous and varied, and many drugs like generic medicines share the same information. Therefore, a collection of drugs sharing common characteristics can be useful to avoid describing the same information several times, but also to contribute to the interoperability of the drug circuit and thus facilitate the international sharing of drug information [1].

Efforts are being made to aggregate information on drugs. Indeed, the SNOMED CT (Systematized NOmenclature of MEDicine Clinical Terms) and RxNorm, two major terminologies in the biomedical field, propose a representation of classes grouping drugs [2,3]. These are respectively the "Clinical Drug" and the "Semantic Clinical Drug", which propose a representation of drugs according to a combination of "active substance - strength in active substance - pharmaceutical form" (e.g. Product containing precisely paracetamol 500 milligram/1 each conventional release oral tablet [SCTID = 322236009], acetaminophen 500 MG Oral Tablet [RxCUI = 198440]) and bring together all drugs sharing these characteristics in a single concept. In France, the non-profit association Medicabase[2], which aims to build and make available semantic resources concerning medication, has developed a database of virtual drugs. These virtual drugs group similar drugs composed of the same active substance(s), the same strength in active substance and the same pharmaceutical form (e.g. Paracétamol 500 mg comprimé [id_MVN=MV00002306]).

[1] Corresponding Author, Yvon K. AWUKLU; E-mail: yvon.awuklu@outlook.com
[2] https://www.medicabase.fr/

Our objective was to assess the homogeneity of drug indications in each virtual drug class. For this purpose, we have created this class since the groupings of drugs as virtual drugs made by Medicabase have not yet been published. Thus, in this article, we will describe the process of creating and evaluating a class that aggregates drug information: the "Virtual Drug".

2. Materials and Methods

Grosjean et al. have defined the virtual drug as [4]: *"a drug that combines drug's brand names having: (i) the same active ingredient(s) or salts of the active ingredient(s) that are clinically equivalent in terms of iatrogenic risks; (ii) the same strengths of active ingredients in active base; and (iii) a galenic form considered clinically equivalent from the point of view of iatrogenic risks."*

The virtual drug "Paracetamol 500 mg, tablet" thus includes especially the products "DAFALGAN 500 mg, scored effervescent tablet", "DOLIPRANE 500 mg, tablet" and "EFFERALGAN 500 mg, orodispersible tablet" because they are equivalent according to this combination.

To build the "Virtual Drug" class, we used the French Public Drug Database (BDPM[3]) that provides online data on the composition of active ingredients, the strength of these active ingredients and the pharmaceutical form of the drugs marketed in France.

2.1. Implementing the Virtual Drug class

Before implementing this class, a preliminary step was necessary. It consisted in: (i) normalizing the pharmaceutical forms of the drugs, and (ii) normalizing and converting the strengths of the active ingredients. The normalization of pharmaceutical forms was achieved by the automatic grouping of forms sharing the same root (e.g. "scored film-coated tablet" and "scored effervescent tablet" have been grouped under the form "tablet"). With regard to the units of concentration of the active substances, an automatic standardization has been carried out in accordance with the Unified Code for Units of Measure terminology (UCUM[4]) (e.g. "µg" and "microgram" have been normalized to "ug"), as well as an automatic conversion of these strengths, if applicable (e.g. "1000 mg/2 mL" was converted to "500 mg/mL").

After this preliminary step, the "Virtual Drug" class was built, consisting of a group of drugs sharing the same composition in active ingredients, the same strength of these active ingredients and the same pharmaceutical form.

2.2. Evaluation

In an attempt to assess the overall consistency of this class, we hypothesized that any drug sharing the same composition in active ingredients, the same strength in these active ingredients and a similar pharmaceutical form should have the **same indications**.

[3] https://base-donnees-publique.medicaments.gouv.fr/

[4] A code system intended to include all units of measurement currently used in science, engineering and international trade (https://ucum.org/trac)

To test this hypothesis, we used the Anatomical Therapeutic Chemical classification system (ATC[5]), which gives the anatomical site of action, as well as the therapeutic class of a drug. Thus, ATC provides information concerning the indication of a drug from a generic perspective. The ATC codes were then automatically extracted from ROMEDI [5], Thesorimed[6] and the drug's summaries of product characteristics (SPCs[7]) (Figure 1). As the anatomical site and the therapeutic class of drugs can be identified in ATC codes of levels 2 to 5, ATC codes of the first level (e.g. N-Nervous system) were not included. The purpose of the evaluation was thus to determine whether or not the ATC codes associated with drugs belonging to a single "Virtual Drug" class were homogeneous (i.e. whether the drugs included in a virtual drug had all the same ATC code(s)).

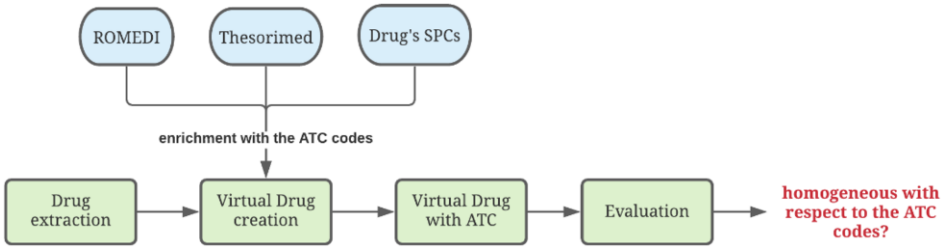

Figure 1. The "Virtual Drug" evaluation process

3. Results

The "Virtual Drug" class included 6,784 classes that were created from 15,318 drugs (Table 1). After pre-processing, 51 distinct drug forms and 1,367 distinct active substance strengths were obtained. The minimum number of drugs contained in a virtual drug is 1, while the maximum number is 41 (Table 2). "PARACETAMOL 500 mg : tablet" is the virtual drug that contained the most drugs (Table 3).

Table 1. Overview of the information about active substances, dosages and pharmaceutical strengths during the creation process of the "Virtual Drug" class

Drugs	Distinct Ingredients in drugs	Pharmaceutical forms		Strengths		Virtual drugs
		Before pre-processing	After pre-processing	Before pre-processing	After pre-processing	
15,318	3,531	414	51	2,181	1,367	6,784

Table 2. Information about the distribution of drugs in the "Virtual Drug" class

	Total	Min.	Q1	Q2	Mean	Q3	Max.
Virtual drugs	6,784	1	1	1	2.26	2	41
Virtual drugs with more than one drug	1,995	2	2	3	5.28	7	41

[5] https://www.whocc.no/atc/structure_and_principles/

[6] https://v3.prod-un.thesorimed.org/

[7] See for example, the SPC of "DOLIPRANE 500 mg, tablet" https://base-donnees-publique.medicaments.gouv.fr/affichageDoc.php?specid=63368332&typedoc=R

Table 3. Top 5 virtual drugs with the highest number of drugs

Virtual drug	Number of drugs
PARACETAMOL 500 mg: tablet	41
PANTOPRAZOLE 20 mg: tablet	39
DICLOFENAC SODIQUE 1 %: gel	36
OLANZAPINE 10 mg: tablet	34
IBUPROFENE 400 mg: tablet	33

The 1,995 virtual drugs comprising more than one drug included 10,529 (68.74%) drugs. Of these, 249 (12.48%) did not have an ATC code, including 240 (96.39%) classes containing only homeopathic drugs. Of the virtual drugs that had an ATC code, 1,555 (89.06%) contained drugs that all had at least one ATC code and 94.84% of their ATC codes were level 5 (being the most specific level).

The evaluation of these 1,555 "Virtual drug" classes showed that 111 (7.14%) of them are heterogeneous with respect to the ATC codes. The causes of heterogeneity were the following:

Source (46.0%). This case covers problems inherent in the source of ATC codes (i.e. incorrect ATC). For example, the virtual drug "ADAPALENE 0.1%: cream" has an ATC code D01AD03 that does not exist. The correct code is D10AD03.

Context (44.2%). This situation corresponds to the heterogeneity due to different ATC codes but having the same meaning in the "Virtual Drug" class in which they are found. For example, the virtual drug "PARACETAMOL | CODEINE 400 mg/20 mg: tablet" contains all three ATC codes (whose meaning in that specific class is the same):

- N02AA59: Codeine in combination with the exception of psycholeptics,
- N02AJ06: Codeine and paracetamol,
- N02BE51: Paracetamol in combination with the exception of psycholeptics.

Route (5.3%). This heterogeneity is related to the route of administration. For example, "ACETYLCYSTEINE 0.2 g / mL: solution" has two ATC codes R05CB01: ACETYL-CYSTEINE (respiratory route) and V03AB23: ACETYLCYSTEINE (venous route).

Granularity (3.5%). This cause includes problems due to the ATC code level provided for a drug. For example, "MIDODRINE HYDROCHLORIDE 2.5 mg: tablet" has the following two ATC codes: C01CA (level 4): Adrenergic and dopaminergic agents and C01CA17 (level 5): Midodrine.

Mode (0.9%). This heterogeneity is due to the mode of administration of a drug. For example, "ETHINYLESTRADIOL | DESOGESTREL 150 ug / 20 ug: tablet" has the ATC codes G03AB05: Desogestrel and ethinylestradiol (for sequential administration) and G03AA09: Desogestrel and ethinylestradiol (for fixed administration).

4. Discussion

As stated earlier, a class comparable to "Virtual Drug" is represented in SNOMED CT ("Clinical Drug") and RxNorm ("Semantic Clinical Drug") [2]. These two classes, although similar, do not incorporate pharmaceutical form grouping. In our study, this aggregation reduced the number of distinct pharmaceutical forms from 414 to 51, which resulting in better aggregation of information. Thus, it would be interesting to compare our virtual drugs to the RxNorm and SNOMED CT classes based on the other features we used to aggregate the information.

More than two thirds (70.59%) of the "Virtual Drug" classes contains only one drug. Despite this, the average number of drugs in the "Virtual Drug" and in his subgroup containing more than one drug is 2.26 and 5.28 respectively. These values reflect the aggregation capacity of this class. The five virtual drugs containing the largest number of drugs are dominated by analgesics, antisecretory drugs, nonsteroidal anti-inflammatory drugs and antipsychotics. These drugs represent the most prescribed drugs in France, which reflects the overall consistency of the "Virtual Drug" class.

To measure the internal consistency of each "Virtual Drug" subclass, it is necessary to assess whether we aggregated drugs that should not be aggregated and whether drugs that should be aggregated were not. The latter seems more difficult to measure because we do not yet have a method to do so. However, it was possible to assess the grouping of drugs that do not need to be aggregated. For this purpose, we used the ATC classification.

Virtual drugs containing only one drug logically have the maximum internal consistency. These classes were removed from our final evaluation dataset because they would have artificially overestimated the quality of the evaluation. The evaluation performed on the classes that could show low internal consistency identified only 7.14% of heterogeneous classes. Among the causes of heterogeneity, only those related to the route and the mode of administration (corresponding to only 0.45% of 1,555 virtual drugs used for the evaluation) are real sources of heterogeneity because they are directly linked to the drug. Therefore, the "Virtual Drug" class presents a good intra-class consistency regarding the indication.

5. Conclusions

In our study, we have shown that drugs belonging to the same virtual drug class have the same indications according to the ATC classification. As future work, we plan to verify this hypothesis with the specific indications of each drug, thus at a finer granularity level.

Acknowledgments

This work was partially supported by the INTENDED AI Chair (ANR-19-CHIA-0014).

References

[1] Rambaran KA, Huynh HA, Zhang Z, et al. The Gap in Electronic Drug Information Resources: A Systematic Review. Cureus. 2018;10:e2860.
[2] Nikiema JN, Bodenreider O. Comparing the representation of medicinal products in RxNorm and SNOMED CT – Consequences on interoperability. Proc 8th Int Conf Biomed Ontol ICBO 2019. 2019;6.
[3] Bodenreider O, Cornet R, Vreeman DJ. Recent Developments in Clinical Terminologies — SNOMED CT, LOINC, and RxNorm. Yearb Med Inform. 2018;27:129–139.
[4] Grosjean J, Letord C, Charlet J, et al. Un modèle sémantique d'identification du médicament en France. Proc Atelier IA Santé. 2019;9.
[5] Cossin S, Lebrun L, Lobre G, et al. Romedi: An Open Data Source About French Drugs on the Semantic Web. Stud Health Technol Inform. 2019;264:79–82.