# Abstractive Summarization of drug dosage regimens for supporting drug comparison

Adrien UGON[a,1], Hélène BERTHELOT[a], Alain VENOT[a], Madeleine FAVRE[b,c], Catherine DUCLOS[a] and Jean-Baptiste LAMY [a]

[a] *INSERM, U1142, LIMICS F-75006, Paris, France;*

*Sorbonne Universités, UPMC Univ Paris 06, UMR_S 1142, LIMICS, F-75006, Paris, France;*

*Université Paris 13, Sorbonne Paris Cité, LIMICS, UMR_S 1142, F-93430, Villetaneuse, France*

[b] *Université Paris Descartes, Paris, France.*

[c] *Société de formation thérapeutique du généraliste*

**Abstract.** Complicated dosage regimens often reduce adherence to drug treatments. The ease-of-administration must thus be taken into account when prescribing. Given one drug, there exists often several dosage regimens. Hence, comparison to similar drugs is difficult. Simplifying and summarizing them appears to be a required task for supporting General Practitioners to find the drug with the simplest regimen for the patient. We propose a summarization in two steps: first prunes out all low-importance information, and second proceed to fusion of remaining information. Rules for pruning and fusion strategies were designed by an expert in drug models. Evaluation was conducted on a dataset of 169 drugs. The agreement rate was 27.2%. We demonstrate that applying rules leads to a result that is correct by a computational point of view, but the result is often meaningless for the GP. We conclude with recommendations for further work.

**Keywords.** Dosage Forms, Data fusion, Drug knowledge, pharmaceutical databases

## Introduction

When new drugs arrive on the market, General Practitioners (GP) need to get an objective opinion about them. Schommer et al. showed that physicians are interested in having support

1 Corresponding Author : adrien.ugon@univ-paris13.fr

for relative comparison of drugs with similar properties (1). A significant improvement in the ease-of-administration is one of the points that should be considered. One difficulty is that a cognitive effort is required to compare all existing dosage regimens given one drug and one indication; prior simplification can overcome this.

A typical drug dosage regimen depends on many parameters: (a) therapeutic indication (*e.g. angina*) (b) patient's characteristics (*e.g. adult*) (c) patient comorbidities (*e.g. renal or liver failure*), (d) administration route (*e.g. oral, intramuscular*), (e) stage in treatment process (*e.g. initial, maintenance*), (f) combination with other drugs, (g) nature of dosage regimen (*e.g. usual, maximum*). Data models used in drug databases depend on the final usage. Most of them are structured for computerized prescriptions with functional alerts (2,3). When prescribing, the dosage regimen is expressed by a dose, a frequency and a duration (*e.g. in adults, in angina, 1 tablet of 1g, 2 administrations per day during 6 days*)

Information summarization consists in condensing an information by maintaining the global meaning significant. Generally applied to text sources (4), summarization extracts most relevant passages. Abstraction step may rephrase them afterwards. It has been applied to drug information for keeping up-to-date drug databases from Medline citations (5).

The objective of this work is to apply summarization to simplify drug dosage regimens.

## 1. Methods

A 2-step process is followed: firstly, we identify the prioritized dosage regimens; others are pruned out. Secondly we fuse remaining dosage regimens with fusion strategies.

Firstly, we need to define a pattern for the resulting dosage regimen. To comply with prescriptions, we decided to express it as a dose, a frequency and a duration. Each of these 3 attributes is expressed by either a single value when possible or by a [minimum - maximum] range otherwise. The duration may have been left empty.

We worked on THERIAQUE®, a structured independent drug database widely used in France (6). In this database, dosage regimens are grouped by sheets. For given therapeutic indications, patient's characteristics and administration routes, a sheet lists a set of dosage regimens. Each dosage regimen is described by nature, dose, frequency and duration. Each of the 3 latter fields are described by minimum value and unit and maximum value and unit. At all levels comments can be added (*e.g. Stage 1, for patient with risk* or *if good tolerance)*.

All pruning rules were designed by an expert in data model for drugs representation. She was asked to indicate, by her experience and, then, by looking at a set of 15 randomly selected drugs, how to decide whether a dosage regimen has a low or high level of importance. This decision was made by considering the frequency of prescriptions following the considered dosage regimen or by prioritizing the convenience of administration for the patient.

Remaining dosage regimens are then fused by indication. Dose, frequency and duration need to be treated differently. Obviously, for all dosage regimens dose and frequency are numeric-valued. On the contrary duration may be empty. Given one therapeutic indication, all remaining dosage regimens are reviewed to generate a [minimum, maximum] range.

## 2. Results

Examples of rules for pruning are listed in Table 1. There are 2 types of rules: (a) elimination rules define a field-value couple that indicates to prune the whole dosage regimen, (b) preference rules order the possible values of one field by importance. Only dosage regimens with the highest importance are retained; others are pruned out.

**Table 1.** Rules for pruning

| Rule type | Concerned field | Rules |
|---|---|---|
| Elimination | **Dose or Frequency** | Dosage regimens whose dose or frequency equals to "To adapt" are pruned out |
| Preference | **Injectable administration route** | "Intramuscular" is preferred to "intravenous" |
| Preference | **patient characteristics** | 1. "Adult" is preferred to "Elderly" and "Child"<br>2. "Elderly" and "Child" are preferred to others (premature, neonate, …) |
| Preference | **Patient comorbidities** | "Patient without comorbidity" is preferred to "Patient with any comorbidity" |

Our method was evaluated on a dataset of new 169 drugs marketed since 2011, chosen after having designed rules. For each drug, predefined rules were applied and the expert defined the expected result. For 46 drugs (27.2%), there was agreement with the expert.

## 3. Discussion

All the rules proposed by the expert could have been implemented. Results are correct, by a computational or mathematical point of view, but may be meaningless for the GP. Sources of meaninglessness are imprecision, confusion and heterogeneity.

We identified different reasons that could harm the progress of the algorithm: (a) a lack of precision in indication in terms of categories potentially leading to misinterpretation. (e.g. *Expressions like "Adult until 16 years" used instead of "adolescent"*). (b) Dosage regimens with different periodicity should be treated separately and prioritized following new rules.

(c) New efficient calculation methods need to be created to convert all encountered non-homogeneous units into administrable units, taking into account extra field such as dose of each tablet or volume of bottle or ampoule. (d) Another pitfall is the difficulty to deal with several-steps treatments. (e) Some dosage regimens depend on physiological parameters such as creatinine clearance or weight. (f) The conversion into administrable units is also an issue, and requires to take into account the divisibility of a tablet. (g) Our approach did not distinguish preventive and curative treatments. It turned out to be suitable to have done it. (h) Taking into account the maximum dosage regimen at different level of time is another issue. (e.*g. 16 tablets per day maximum, 2 tablets maximum per administration*)

This work has pointed out that, to remain significant, more fields need to be exploited, which was not possible in this work, since many information is stored in comments fields. Thus, data need to be structured in an exhaustive proper and scalable way. A clear definition should be given for all concepts. New rules should be designed to exploit all fields. We also need a similarity function for dosage regimens, including the measurement of similarity of doses taking into account the unit. Default values may be used for required extra parameters.

**Acknowledgements**

This work was funded by the French drug agency (ANSM, Agence Nationale de Sécurité du Médicament et des produits de santé) through the VIIIP project (AAP-2012-013).

**References**

1. Schommer JC, Worley MM, Kjos AL, Pakhomov SVS, Schondelmeyer SW. A thematic analysis for how patients, prescribers, experts, and patient advocates view the prescription choice process. Res Soc Adm Pharm RSAP. juin 2009;5(2):154‑69.
2. Martin P, Haefeli WE, Martin-Facklam M. A Drug Database Model as a Central Element for Computer-Supported Dose Adjustment within a CPOE System. J Am Med Inform Assoc JAMIA. 2004;11(5):427‑32.
3. Séné B, Venot A, de Zegher I, Milstein C, Errore S, de Rosis F, et al. A general model of drug prescription. Methods Inf Med. sept 1995;34(4):310‑7.
4. Salton G, Singhal A, Mitra M, Buckley C. Automatic text structuring and summarization. Inf Process Manag. mars 1997;33(2):193‑207.
5. Fiszman M, Rindflesch TC, Kilicoglu H. Summarizing Drug Information in Medline Citations. AMIA Annu Symp Proc. 2006;2006:254‑8.
6. Husson M-C. Thériaque® : base de données indépendante sur le médicament, outil de bon usage pour les professionnels de santé. Ann Pharm Fr. nov 2008;66(5-6):268‑77