pHealth 2021
B. Blobel and M. Giacomini (Eds.)

© 2021 The authors and IOS Press.
This article is published online with Open Access by IOS Press and distributed under the terms of the Creative Commons Attribution Non-Commercial License 4.0 (CC BY-NC 4.0).
doi:10.3233/SHTI210572

Towards Personalized Medication

Michal HUPTYCH[a1], Jiri POTUCEK[a,b], Lenka LHOTSKÁ[a,c]
[a] *Czech Institute of Informatics, Robotics and Cybernetics, Czech Technical University in Prague, Prague, Czech Republic*
[b] *MWCare Ltd., Prague, Czech Republic*
[c] *Faculty of Biomedical Engineering, Czech Technical University in Prague, Prague, Czech Republic*

Abstract. The paper describes some aspects of precision medicine and shows the importance of pharmacokinetics and pharmacodynamics for the therapeutic drug monitoring and model-informed precision dosing. A key element in the design of the pharmacokinetics and pharmacodynamics (PKPD) models is relevant literature search that represents an essential step in the procurement and validation of a new drug. Available search engine resources do not offer specific functionalities that are required for efficient and relevant search in reliable literature sources. We present a prototype of such an intelligent search engine and show its results on real project data.

Keywords. Precision medicine, model-informed precision dosing, intelligent literature search, pharmacokinetics, pharmacodynamics

1. Introduction

Fast technological development enabled design and development of new applications across many fields during last two decades. Medicine and health care belong to these fields. Almost all medical devices can be connected to computers and communicate acquired data. Wireless technologies, sensors, wearables and Internet of Things contribute to development of continuous data acquisition and communication. Electronic health record is an inseparable part of this development. Development of new drugs is almost unimaginable without extensive computational resources. Few years ago, we spoke about the concept pf P4 health care - Participatory, Predictive, Preventive, Personalized. Recently the fifth P has been added: Precision. Precision medicine is understood as an approach to patient treatment that considers individual variability in genes, environment and lifestyle for each patient. This approach is already applied in some areas of medicine for targeted treatment, as for example in oncology. With the development of new software tools there opens a great opportunity to apply this approach to wider spectrum of treatments.

The core of the software tools for precision medicine are modules for therapeutic drug monitoring and model-informed precision dosing. Background for their development constitute pharmacokinetics and pharmacodynamics models for individual drugs. Pharmacokinetics and pharmacodynamics represent two phases of the processes

[1] Corresponding Author: Michal Huptych, PhD, Czech Institute of Informatics, Robotics and Cybernetics, Czech Technical University in Prague, Prague, Czechia; Email: Michal.Huptych@cvut.cz

between drug input and emergence of the response on the output. The pharmacokinetic phase represents all events between the input (drug dose administration) and the achievement of drug concentrations throughout the body. The pharmacodynamic phase represents all events between the arrival of the drug at its site of action and the onset, magnitude, and duration of the biological response [1]. The models have to take into account many parameters, including information about the patient, e.g. weight, age, gender, ethnicity, metabolism rate, interaction with other drugs, negative effect on potential co-morbidities, etc. Additional parameters are dose, frequency of administration and route of administration. These parameters represent variables that can be adjusted to optimize the therapeutic effects of a drug for a particular disease condition. It is possible to control the onset, intensity and duration of the positive drug effect, while minimizing any harmful effect. Obviously, for designing such models detailed understanding of human physiology, metabolism, dose-response relationships, dependence of response on drug route, drug elimination and clearance, molecular biology and genetics is needed. That means that such work is strongly interdisciplinary.

In the paper we present briefly development in the area, show the concept of a system for model-informed precision dosing and describe an important tool supporting development of models for drugs for given populations. The aim of this tools is to speed up the development of models and make the work more efficient.

2. Related Work

Last decade showed that the demand for precise drug dosage and therapeutic drug monitoring is continuously growing. There have been performed many clinical studies, in particular in oncology, showing the positive effect of targeted therapy [2, 3, 4]. Other studies include optimizing care of critically ill patients with severe infections [5], optimizing the dose for children [6], physiological regulation of drug metabolism and transport [7, 8]. Several reviews discuss current state and future of precision medicine [9, 10, 11]. Some authors analyze impact on the industry and development of new drugs [12, 13, 14].

Development of software tools for therapeutic drug monitoring and model-informed precision dosing is not an easy task. It requires extensive and deep knowledge from different disciplines, which means gathering the experts willing to collaborate on the development, and also extensive literature search for drug models and their validation. There are not so many tools available as two evaluation studies, performed in the last decade show [15, 16]. In [15], published in 2013, 12 software tools were tested and evaluated using 12 criteria: pharmacokinetic aspects (population and drug, models, modularity, plot, and various) general characteristics (user interface, computing aspects, interfacing, cost, report, and storage) and expertise of authors. The authors of [16] developed with the help of 22 experts more detailed evaluation criteria, finally grouped in eight categories: user friendliness and utilization, user support, computational aspects, population models, quality and validation, output generation, privacy and security, and cost. They evaluated 10 software tools, three out of them were already evaluated in 2013, however, since then new versions were developed with more advanced functionalities. They were MwPharm (upgraded to MwPharm++), MM-USC*PACK© (now BestDose) and T.D.M.S. (now PrecisePK). MwPharm/MwPharm++ achieved in both evaluations one of the highest rankings (first, resp. second). MwPharm++ [17] is continuously complemented with new drug models. In this work it appeared that a tool for supporting

the model development is needed. Thus, in collaboration of the MediWare company with the Czech Technical University in Prague research and development of an intelligent search robot was initiated.

3. Intelligent Literature Search

The MwPharm++ project (software for drug metabolism modelling, in which validated models are implemented) has been expanding into foreign markets over the last 2 years and is currently used in 18 countries worldwide. In connection with the completion of the cloud version of MwPharm Online, the project will upgrade to the category of "software as a service". The project consists of 2 parts:

- a custom algorithm for patient dose adjustment based on a pharmacokinetics and pharmacodynamics (PKPD) model using genetic knowledge related to drug metabolism and
- an ever-expanding database of drug substances that contains PKPD parameters, knowledge of possible polymorphisms, and the literature and factual sources from which these values were determined.

The main task is to specify PKPD models for a selected set of drugs and for the given population (ethnicity is one of the parameters that influences the drug pharmacokinetics and pharmacodynamics). Literature search is an essential step in the procurement and validation of a new drug. The intention of the designed tool is to speed up and improve the process of conducting a literature search by automatically extracting relevant articles using both the tools utilized in natural language processing and by taking advantage of the automatic processing capabilities of the retrieved results. It is worth mentioning that the publishing activity in the area of interest is very intense, with about 10 articles per week. This results in a high number of articles (depending on the number of drugs) related to the problem. With such a large number of papers, it does not make sense to present the results only in the form of a list of publications, but it is necessary to evaluate and structure the results for maximum clarity and the possibility of systematic work. The work to date has established a basic terminological basis for the representation of (abstracts of) articles, the method and conditions for their selection, and the method and possible forms of creating a structured, systematic review for the selected literature.

So far, the available search engine resources (Google, PubMed) have been used to build the database, with subsequent manual selection of suitable papers. As the requests for new drugs and new population groups are increasing, we decided to simplify this process by implementing a robot whose function aims not only at searching and selecting articles, but also at systematically classifying the retrieved literature sources in order to generate a library of papers on the population parameters of the proposed new drug (with the possibility of user annotation). Currently, the system is mainly used for TDM (therapeutic drug monitoring), which involves about 200 drugs out of the total number of drugs in use. As the new cloud-based version of MwPharm Online also allows for first dose reduction in case of renal or metabolic insufficiency of the patient, we are extending its functionality to a total of 1500-2000 drug substances for each population group. Initially, the information needed to determine the PKPD parameters of new drug substances was obtained manually at a cost of approximately 400 EUR / drug. With the new robot, this cost should be reduced to 100 EUR/drug. For the 2 population groups used so far - the Transcaucasian population and the Asian population, this means a saving of almost 2/3 of the costs incurred so far. For example, for the Transcaucasian population

and 2,000 drugs, this is a saving of more than 350 thousand EUR and a similar benefit can be expected for the Asian and other populations. Considering 500 most used medicines for 1 population, the savings compared to the original costs is about 40 - 80 thousand EUR. If we consider that this is an ongoing process where we have to add and update a lot of information on specific medicines every year, this is an ongoing annual saving.

4. Description of the Intelligent Search Engine

As a reference project, we take article retrieval for model validation for a list of drugs provided by the Korean project partner (part of the project No. TF05000020 MwPharmASIA - database extension of drug substances and their MwPharm models for East Asian population and development of NGS diagnostic panel and algorithm for predicting statin pharmacokinetics/dynamics of the Technology Agency of the Czech Republic). This set contains 38 drugs for which we wanted to see if PKPD models can be specified/validated. Furthermore, collaboration with experts in the field of PKPD model building was then used to determine their part of the terms for the text search. Collaboration with these experts was done in an iterative manner to obtain the best quality term sets for text retrieval. Until now, no such standardized sets existed (at least for our subject area).

These documents are used in the form of a dictionary. The dictionary forms the basis of each project created within the application developed. The use of dictionaries is one of the standard methods for text processing by assigning entities to the required (searchable) parts of the text. The user can basically view the dictionary in its three levels. Each dictionary is thus divided into three levels, representing categories, entities and the search words themselves. Primarily, further processing works with entities that represent a defined set of words. However, from the searches carried out, it became clear that for a clearer orientation in the selection result it is advisable to define a higher level, i.e. the category of entities, in certain parts.

After defining the dictionary, the next steps are available. The first step is to collect the articles. We chose the PubMed database as the primary source of articles for this prototype because it contains articles that have been subjected to peer review. Thus, the resulting extraction contains articles that can be considered valid in terms of expertise. The PubMed database has an advanced search system that allows searching for combinations of terms using the logical AND and OR operators.

The query is created automatically either based on entities in mandatory categories or (if no category is marked as mandatory) based on entities in all categories. The complication is that the search

- is either very rigid and not many relevant papers are included in the resulting selection,
- or the search rule is loose and the search results in thousands, tens of thousands or even hundreds of thousands of articles.

With a broader coverage, which is the reference in a case like this, the numbers of articles from a PubMed search are very high, and even automatically clustering them according to the occurrence of each class of interest would lead to an unprocessable result. Therefore, the step of selecting articles from PubMed by our independently developed system is quite relevant.

To illustrate the complexity, we show an example: As of December 23, 2019, there are 14980 articles on PubMed corresponding to our query, of which 14912 abstracts are in the internal working database - where the texts are already there after tokenization and lemmatization. The same number (14912) was screened and tagged, and 414 articles were finally selected as best meeting the dictionary-defined requirements. These 414 articles are being worked on further. It is obvious that the reduction of number of articles that have to be reviewed by human experts is significant.

Part of the file tagging is the calculation of the score of each article. For this purpose, a weight is added to each defined entity. However, the user does not directly define the numerical value of the weights, but gives a hierarchy of entities, whereby any number of entities can be at the same level.

The user is not required to define the weights and can leave all entities with the same weight (weight 1). However, s/he can use the weights even if none of the dictionary categories are defined as mandatory and the level of importance of the article is then defined only by the resulting score of the article. In our benchmark example, a combination of both approaches has been chosen, i.e. we have defined 3 categories as mandatory and the weights of each entity are defined at the same time.

If a dictionary is defined and the supporting documents are ready, the user can specify browsing and tagging of articles by dictionary. Here, the result of the search through the PubMed database interface serves as input to our extraction algorithm. The final selection of the articles is based on the dictionaries defined, the weights of each class (entity) defined in the dictionaries, and the algorithm for arranging the results. The output of this part of the processing is a primary ordered list of entities from each category for each of the selected articles. This tagging is available to the user in three different forms of representation.

The first of the possible representations is a tabular form. This output is easy to see for many users, but it is true that the results of the article tagging are stored in Microsoft's Excel format. It is therefore possible to open them in this program and make any further modifications to the output. A far more meaningful representation of the results is the tree structure, which is formed by creating classes for combinations of entities (there can be, but also one single entity) from each category. An example of detail of such an output is shown in Figure 1, in which it can be seen that the order of the categories is population, drugs, pharmacological information, characterization and the last one is the article itself.

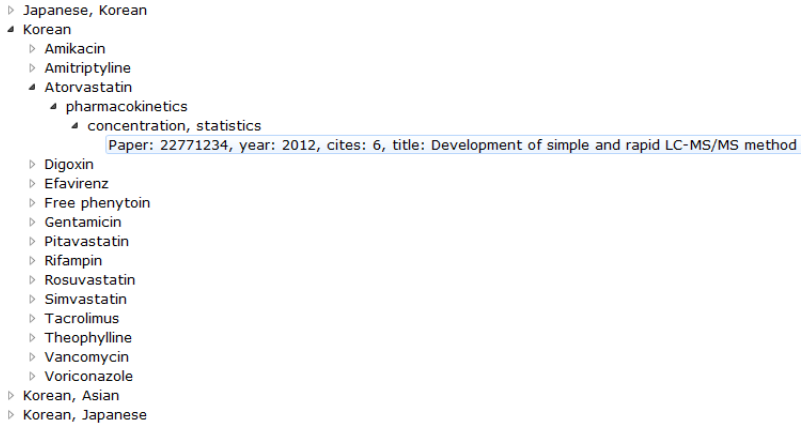

Figure 1. Detail of the resulting tree

However, this order is not fixed and the user can modify the order of the categories. The order of the entities has a relatively strong influence on the clarity of the result tree, so it is a good idea to choose it appropriately.

In any case (for any dictionary), the article itself is part of the result. In the tree representation, a link to the PubMed database is added to this tree node, so the user can open the article abstract by double-clicking.

The last form of representation of the results from dictionary tagging is a graph. This form of representation can be very useful for identifying common articles and selecting those for download and study. However, it is true that this form of representation needs to be handled very judiciously to make the resulting graph workable. Therefore, in order to give the user the best possible experience when reading the results of the selection in graph form, the user has two options for editing the input data (labelling results). The first is the option to select the entities to be omitted from the graph. Thus, the user can define that s/he wants, for example, only articles containing the entity Korean and no others. This does not only mean removing the nodes of the other populations, but a very significant reduction in the total number of nodes, because the entity Korean does not connect to many entities from other categories (there are no links connecting them) and so these entities (nodes of the graph) also drop out. The second option is then to select the links themselves. Again, the individual articles in the leaf are linked to the url and a double-click leads to the selected article in the PubMed database.

However, the results of tagging articles using dictionaries have several drawbacks. The biggest of these is the fact that there is no simple coincidence of words within the text. A dictionary is thus a very quick and convenient tool in terms of definition by experts, but it is advisable to complement the results given by it with some more comprehensive form of text representation. In this project, we made two assumptions:

- we are interested in the relationship of entities, or their proximity in the text, and the filtering of the text into classes is of secondary interest to us (the dictionary has proven itself in this respect);
- we work primarily with short texts (abstracts), for which forms of representation such as term_frequency/inverse_document_frequency are common.

We have therefore chosen the standard text representation in the form of word n-grams. In this representation, the text is sequentially split word by word into a combination of n words. The n can be chosen as a number from 2 (bi-gram), 3 (tri-gram), 4 (quatro-gram) and even higher. However, there is of course a logical limitation in the number of words for which text sampling make sense. We have chosen to create 2-gram, 3-gram and 4-gram models and the user has the option to choose which n-gram model to use to label the selected articles. An example of a window listing 4-grams defined based on dictionary-defined articles is shown in Figure 2.

The process of marking and clustering cells is then divided into the following steps:

- Each n-gram has stored with it which entities represent it and which articles contain it. Here it is very important to note that n-grams are made up of the words of articles, but these are represented by entities in our system. Thus, two different n-grams can be described by the same entities.
- A transformation is made where entities are the key and n-grams and articles are subsets of them. This step leads to the clustering of the n-grams. This reduces the dimension of this representation (example from the reference set: the original 13624 4-grams are represented by 334 entities or combinations of entities)

- In the next step, a transformation is performed where entities are atomized (until now they could represent n-grams in combinations) and assigned the value 1 if they are represented in the given n-gram.
- The final step is to convert the previous representation into articles, i.e., each n-gram is replaced by the articles it represents. Since one article can be represented by multiple n-grams, the occurrences of each entity are summed. To reinforce the role of the combination of entities within an n-gram, this is weighted - here, if an entity occurs repeatedly in an article but alone, its value within the vector of that article is increased by +1. If an entity occurs in combination with another entity, the value within the vector for both is increased by +10, in the case of three entities it is increased by +100, and in the case of 4 (we are working with a 4-gram in the description) it is increased by +1000.

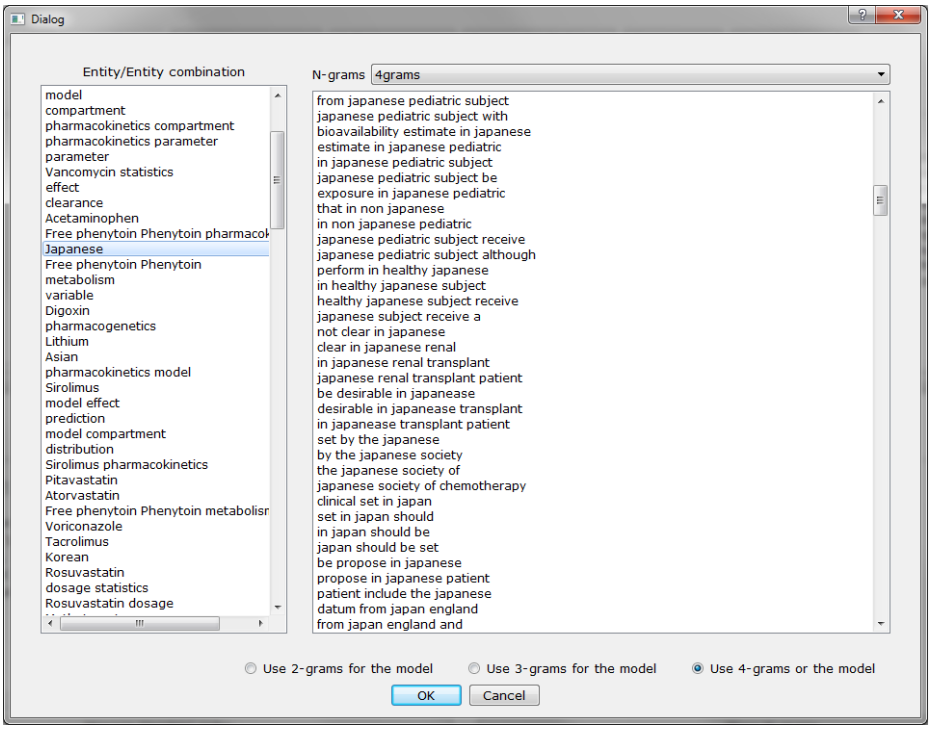

Figure 2. Overview of 4-grams extracted from the articles selected by the dictionary

By this procedure, we have obtained a vector representation of articles that takes into account combinations of entities in n-grams. We can now perform the standard hierarchical clustering process, using cosine similarity as the similarity metric. The result of this process is represented within the application in two ways. The first is a tree representation, an example of its detail is shown in Figure 3.

The second is the dendrogram representation of the clustering result. In this form of representation, the user can again define which entities will be included in the comparison and can then choose which articles they want to see in the display. The items

in the article list are ordered according to the ordering (clustering) in the dendrogram and again serve as links to the PubMed database for the selected article.

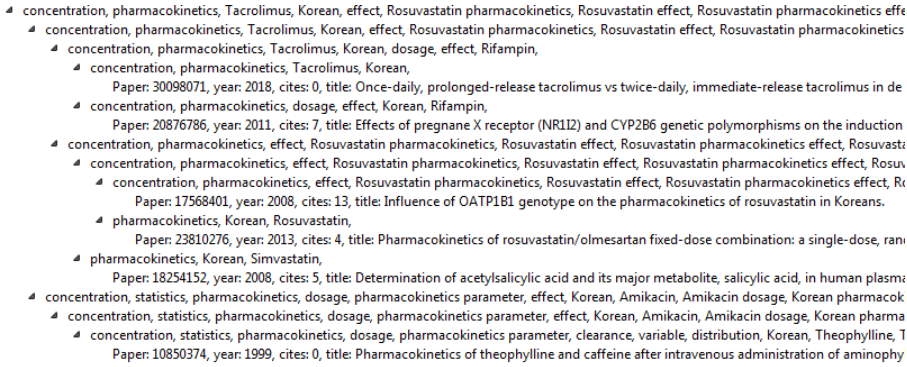

Figure 3. Detail of a tree generated from the clustering results for a model based on 4-grams from the reference example

5. Conclusion

The proposed system allows for functional search, sorting and visualization of literature resources with output storage that allows to refine one's own library of articles. The system is in the prototype stage, which enables all the declared functionalities. Currently, the next phase of evaluation of the system is underway and the evaluation of functionalities and elements that should undergo further development for the most optimal use of the tool.

The studies presented in the Section 2 Related work show that the demand for model-informed precision dosing software tools will grow and that it may positively influence the whole area of precision medicine. Thus, all software tools that can support the process of development of new drug PKPD models are and will be useful and effective contribution.

Acknowledgements

Described research is supported by the project No. FV40426 “Supplementary Platform for Physiologically Based Pharmacokinetic Modeling (PBPK)” funded by the TRIO program of the Ministry of Industry and Trade of the Czech Republic.

References

[1] Rosenbaum SE. (ed.). Basic Pharmacokinetics and Pharmacodynamics. 2nd edition. Wiley, 2017
[2] Polasek TM, Ambler K, Scott HS, et al. Targeted pharmacotherapy after somatic cancer mutation screening. F1000Res. 2016;5:1551
[3] Darwich AS, Ogungbenro K, Hatley OJ, et al. Role of pharmacokinetic modeling and simulation in precision dosing of anticancer drugs. Transl Cancer Res. 2017;6(S10):S1512–S1529
[4] Rini BI, Garrett M, Poland B, et al. Axitinib in metastatic renal cell carcinoma: results of a pharmacokinetic and pharmacodynamic analysis. J Clin Pharmacol. 2013;53(5):491–504

[5] Tängdén T, Ramos Martín V, Felton TW, et al. The role of infection models and PK/PD modelling for optimising care of critically ill patients with severe infections. Intensive Care Med. 2017;43(7):1021–1032
[6] Vinks AA, Emoto C, Fukuda T. Modeling and simulation in pediatric drug therapy: application of pharmacometrics to define the right dose for children. Clin Pharmacol Ther. 2015;98(3):298–308
[7] Morgan ET, Dempsey JL, Mimche SM, et al. Physiological regulation of drug metabolism and transport: pregnancy, microbiome, inflammation, infection and fasting. Drug Metab Dispos. 2018;46:503–513
[8] Peck RW. The right dose for every patient: a key step for precision medicine. Nat Rev Drug Discov. 2016;15(3):145–146
[9] Polasek TM, Shakib S, Rostami-Hodjegan A. Precision dosing in clinical medicine: present and future. Expert Review of Clinical Pharmacology. 2018; 11(8): 743-746. https://doi.org/10.1080/17512433.2018.1501271
[10] Elemento O. The future of precision medicine: towards a more predictive personalized medicine. Emerg Top Life Sci. 2020 Sep 8;4(2):175-177. doi: 10.1042/ETLS20190197.
[11] Akhoon N. Precision Medicine: A New Paradigm in Therapeutics. Int J Prev Med. 2021 Feb 24; 12: 12. doi: 10.4103/ijpvm.IJPVM_375_19. eCollection 2021.
[12] Allerheiligen SR. Impact of modeling and simulation: myth or fact? Clin Pharmacol Ther. 2014;96(4):413–415
[13] Dugger SA, Platt A, Goldstein DB. Drug development in the era of precision medicine. Nat Rev Drug Discov. 2018 Mar; 17(3): 183–196. doi: 10.1038/nrd.2017.226
[14] Peck RW. Precision Dosing: An Industry Perspective. Clin Pharmacol Ther. 2021 Jan; 109(1): 47-50. doi: 10.1002/cpt.2064.
[15] Fuchs A, Csajka C, Thoma Y, Buclin T, Widmer N. Benchmarking therapeutic drug monitoring software: a review of available computer tools. Clin Pharmacokinet 2013 Jan;52(1):9-22. doi: 10.1007/s40262-012-0020-y
[16] Kantasiripitak W, Van Daele R, Gijsen M, Ferrante M, Spriet I, Dreesen E. Software Tools for Model-Informed Precision Dosing: How Well Do They Satisfy the Needs? Front. Pharmacol., 07 May 2020, |https://doi.org/10.3389/fphar.2020.00620
[17] Mediware. MwPharm++ Therapeutic Drug Monitoring Tool, retrieved from: https://mediware.cz/en/mwpharm-documentation