**Elucidating drug-drug interactions underlying drug polypharmacy profiles**

**Jorge Hernansanz**1**, Gianluca Mazzoni**1**, Cristina Leal**1

1 Novo Nordisk Foundation Center for Protein Research, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen DK-2200, Denmark.

Corresponding author: jorge.hernansanz@cpr.ku.dk

**ABSTRACT**

**Background:** The reuse of electronic health records (EHR) is seen as a major driver to precision medicine. Therefore, it has been collected this clinical data from in-patient Danish hospital admissions for posterior analysis. Drug-drug interaction (DDI) information remains as a non-detailed data field, thus it needs complementation from specialized drug sources, opening a novel challenge to the integration of data coming from different field sources.

**Methods:** We elaborated a compendium of DDI from a wide variety of publicly available drug sources and normalize the drugs to ATC, if needed. A second stage was spent with drug characterization analysis at a global scale covering the DDI information recollected. Finally, an integration of both clinical data and DDI data was performed resulting in a better information coverage the EHR data was not able to achieve by itself. (This last sentence is a draft of the draft of the draft…).

**Results:** Across our databases, there are **X** unique raw drug/chemical names and **X** unique raw drug-interaction pairs. Drug/chemical normalization to its ATC code reduced our drug output due to the casual exceptions of drugs/chemicals having not been assigned this sort of identifier. As stated in earlier projects of DDI extraction we found ourselves with DDI information that varied widely in coverage, leading to a little overlap/ consensus between them. (The last part to be developed, yet).

**Conclusions:** ...

**Keywords:** Drug-drug interactions, WHO-ATC identifiers, EHR,… (To be writtenmore?)

**ABSTRACT**

The study of drug-drug interactions (DDI) has the potential to open the backdoor of numerous comorbidities and confusing drug trajectories that are likely to increase in significance for the following decades. The complete drug-drug interaction profile has not been elucidated yet.

Adverse events (AE), can be the side effect of interactions between two or more drugs [2] or incompleteness regarding drug labeling [3]. These should be given a sensible focus so as to take the first steps towards a complete and structured network of DDI and its adverse events.

**Drug-Drug interactions: challenges to establish a common structure**

Previous studies already tried to integrate different data sources containing DDI information which resulted in little overlap between [3]. The integration of different dataset is challenging due to the use of different identifiers and the lack of a golden standard.

Our first stage of the project consisted of extracting all the DDI data from publicly available databases, and gather all of them by mapping each drug to a common identifier; the ATC codes. Previous research project [3] performed a similar task by using Drugbank codes. The firstThe different databases that were just one-time updated have been extracted from the GitHub’s project [3], meanwhile those one that are being continuously updated have been manually extracted and treated. Additional databases containing DDI were also included so as to have a more comprehensive dataset.

**MATERIALS AND METHODS**

**Extraction of information**

Knowledge about drug–drug interactions commonly arises from preclinical trials, from adverse drug reports, or based on knowledge of mechanisms of action [4]. The main sources comes from a free-published dataset of DDI/PDDI interactions where raw data is extracted via using REST APIs or web scraping through its website. The recollecting focuses not only on DDIs object and precipitant, but also in their adverse effects, mechanism of action, level of interaction... Additionally, single-drug information with similar features was also added as information to evaluate.

The main goal of this step consisted on linking all drugs to its ATC and establish a common dataset that relates the different DDI we can recover.

**ATC code**

Anatomical Therapeutic Chemical (ATC) Classification System is a drug-related index that classifies within different levels the chemical group of each drug, where each drug may have several ATC codes but each drug its exclusive for each ATC code (No more than one Drug per ATC code). It consists of 14 different anatomical groups, each of them reaching 5 different levels of differentiation.

Different databases such as Drugbank and KEGG annotated classifies each drug using ATC codes, but the one that covers all the identifiers and is in charge of maintaining and updating it, is the World Health Organization (WHO).

**Drugbank**

Resources come from Drugbank XML dataset that can be downloadable from the website. Using a R package specific for this document; DBparsed, it was transformed into a variety of datasets covering description of the interactions, pathways, and chemical groups that drugs belong to.

**KEGG**

Accessed the REST API of KEGG database to extract DDI interactions from each one, providing extra informa- tion such as mechanism of action, as well as other kind of interactions which may be relevant. Moreover, it was obtained information of the different pathways and diseases the drugs are involved in.

**Two sides**

Several sort of information can be recovered from Twosides webpage where we count with a dataset of DDI detailing disease provoked, and another describing the individual side effect caused by each drug. Both of them have, for each side effect, a propensity score computed that gives evidence to it.

**NDFRT**

This dataset is composed of a mayor coverage of common USA brand drugs. Single drug information like mechanism of action or therapeutical effect can be accessed from its ftp repository. DDI information was extracted from BioPortal webpage via queries on its SPARQL endpoint. From there, it was extracted object and precipitant, and intensity of reaction.

**Crediblemeds**

Crediblemeds is a website dedicated to cover and give significance to drugs and DDI related to QT interval and TDP abnormality. DDI information can be extracted directly from its website where we obtained level of risk and mechanism of action. Its REST API allows us to retrieve several list of drugs that allows us to know more individual information like TDP risk or therapeutical effect.

**ONC non-interruptive and ONC high-priority**

These two datasets comes from a comprehensive analysis between several commercial drug suppliers, supported by ONC. It is just covered the precipitant and object of the interaction, no more information.

**HIV-insite-interactions**

This website from the University of California covers a great database of antiretroviral DDI interactions. Data is in html tables where we find common features such as mechanism of action or therapeutical effect, but also new ones such us the dose used of each drug.

**HIV, HEP and CANCER interactions**

These three databases are hosted by the University of Liverpool where each of them covers a comprehensive amount of DDI interactions related to the specific disease. Here we find characteristics such as the level of interaction, an evidence level, and a description of the pharmacokinetic mechanism of the interaction. Its extraction was performed by applying web scrapping techniques.

**3/4**

**DDI Corpus 2011, 2013**

NLP DDI corpus are becoming a common source of text mining data extraction from medical abstracts. These ones were elaborated by the lab group of Isabel Segura[7] and already treated by [3]. We obtained a DDI dataset covering a description from each one from its GitHub project.

**Phaedra corpus**

This is a recent corpus that has extracted medical information from Medline abstracts. From here we recover DDI interactions followed by its adverse effect and therapeutical effect. Understanding of text mining techniques was required to extract parsed information.

**REFERENCES**

1. David C. Classen, M. . S. P. R. P. , MD & David W. Bates, M., MD. Critical drug-drug interactions for use in electronic health records systems with computerized physician order entry: Review of leading approaches. figshare https://insights.ovid.com/crossref?an=01209203-201106000-00001 (2011).
2. Percha, B. & Altman, R. B. figshare https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3808975/ (2013).
3. Serkan Ayvaz a, J. H. b. O. H. c. Q. Z. d. J. S. e. N. P. T. f. S. V. f. M. B. g. M. S. h. M. R.-M. i. M. D. j. R. D. B. k., . figshare https://www.researchgate.net/publication/274195644\_Toward\_a\_complete\_dataset\_of\_drug-drug\_ interaction\_information\_from\_publicly\_available\_sources (2015).
4. Peter Wæde Hansen, T. S. S. E. L. F. C. T.-P. L. K. G. H. G., Line Clemmensen & Andersson, C. figshare

https://www.ahajournals.org/doi/10.1161/CIRCOUTCOMES.116.003055 (2016).