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

**Jorge Hernansanz**1**, Cristina Leal**1**, Gianluca Mazzoni**1**, Søren Brunak**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, drug combination clinical data of from in-patient Danish hospital admissions has been collected for posterior analysis. Drug-drug interaction (DDI) remains as a non-detailed data field, needing complementation from specialized drug sources and opening a novel challenge to the integration of data.

**Methods:** We generated a compendium of DDI integrating of **xx** publicly available drug sources: DrugBank, TwosSides, [name there]. from a clinical setting using the DDI information collected in the compendia and performed a network analysis.

**Results:** Across our databases, there are **X** unique drug/chemical names and **X** unique DDIs. ….

**Conclusions:** …

**Keywords:** Drug-drug interactions, WHO-ATC identifiers, EHR…

**BACKGROUND**

* *What are DDI and its role in pharmacovigilance*
* *Methods for obtaining DDIs, developing the importance of text-mining or NLP techniques in clinical text abstracts.*
* *The problem with drug identifiers; the challenge of normalizing DDIs to a single identifier.*
* *DDI information; explaining the different features that it can be found at the public databases.*
* *Give highlights of general findings (Of my data) related to the expected overlap of DDI or common features so as it can serve as an introduction for the section of Results.*

A DDI occurs when one drug modifies the pharmacological activity of other drug. It is a common characteristic of adverse drug reactions (ADR) [1], which is the focus of pharmacovigilance [2]. DDIs can be divided into two major groups, pharmacokinetic interactions, which involves changes in the absorption, distribution, metabolism and excretion of the drugs [3], and pharmacodynamics interactions; in which directly one drug changes the pharmacological effect of another one being additive or antagonist [4].

DDI data can be found in both public website sources (Drugbank, Crediblemeds…) and private commercial sources (Lexicomp, Stockley’s drug interactions…). Some DDI extraction challenges were hold in the past with the goal of covering different approaches when extracting DDI data from biomedical texts [5] [6]. DDI data covers the drug pair but also information related to its clinical significance, side effects, or evidence level.

Previous studies have tried to integrate 14 databases with DDI information mapped to Drugbank identifiers [7]. In a society where co-administration of drugs is becoming more and more frequent, a synthesis of this knowledge would help clinicians guide patients to effective and less noxious medication therapies [8].

In this paper, our objective is to extract all DDI data from publicly available databases, and gather all of them by mapping each drug to a common identifier; the ATC code. Having a common dataset, we will characterize the drug pairs and will factorize its DDI information so that it can be used in future exploratory analysis. Finally, we filtered our compendium to a clinical data generated by the laboratory and tried to identify significant patterns across it. Its importance in the pharmacovigilance field makes trivial to unify all this data in a unique compendium, normalized to a golden standard identifier.

**MATERIALS AND METHODS**

* Brief description of the compendia we are creating and its purpose
* *Mentioning as a reference for DDI extraction the Github project mentioned in [3]*
* *Explaining the DDI resources*
* *Parsing DDIs to ATC identifiers; explaining the different methodologies used without giving numbers. Mentioning the rest APIs, web scraping…*
* *Factorizing the different features; explaining the different methodologies without giving numbers.*

**ANATOMICAL THERAPEUTICAL CLASSIFICATION**

Anatomical Therapeutic Chemical (ATC) Classification System is a drug-related index that classifies within different levels the chemical group of each drug. Each drug may have several ATC codes but there is only one drug for each ATC code). It consists of 14 different anatomical groups, each of them reaching 4 different levels of differentiation: therapeutical, …, … and ….

Different databases such as Drugbank and KEGG work with it, but the one that covers all the identifiers and is in charge of maintaining and updating it, is the World Health Organization (WHO).

**Drug Interaction data**

We created a drug- drug interaction dataset using content from 15 openly available database resources. Many of them are continuously updating, but others are projects that are no longer maintained. A special mention to Ayvaz project [7], which already parsed some of these one-time databases that did not have specific identifiers but regular drug names.

The method each database was extracted is mentioned below. Several kind of sources were met such as regular XML format files, REST APIs, or information on website that was downloaded directly from there.

Each database contains specific information / features related to DDI pairs. From the 15 databases, we recollected five kinds of DDI features. *Clinical significance* feature associates to the DDI the level of change in the physiological effect of the drugs affected. *Evidence level* gives an estimation of how well documented is that DDI. *Management options* refers to how to proceed with the administration of the drug combination. *Mechanism of action* reflects the target/s involved due to the DDI. *Side effect* states the adverse reaction produced by the DDI.

Clinical significance, evidence level, and management features were transformed to digits in order to characterize in a numerical scale the magnitude of that information for each DDI, which could be used in posterior exploratory analyses.

Mechanism of action and Side effect features were also treated to be used in exploratory analyses. Mechanism of action output was restricted / formatted to a general vocabulary of actions easy to deal with. Side effect output was filtered to the ones that had MedDra identifiers.

**DRUGBANK**

The DRUGBANK database [9] is a unique bioinformatics and cheminformatics resource that combines detailed drug data with comprehensive drug target information. Information retrieval was performed with the R package “DBparsed” [10], obtaining a brief description for each DDI. The version used in this study (5.0) was downloaded from the Drugbank website on January 2020.

**KEGG DRUG**

The KEGG DRUG database [11] is a comprehensive drug information resource for approved drugs in Japan, USA, and Europe. Information is unified based on the chemical structure and/or the chemical component of active ingredients. Information retrieval was performed with its REST API [12], such as drugs’ mechanism of action and management DDI (“Contraindicated”, “precaution”). The version used in this study (93.0) was downloaded from the KEGG API in February 2020.

**TWOSIDES**

The TWOSIDES project [13] consists of a comprehensive database of DDI side effects elaborated by Tatonetti’s laboratory. Twosides reports a total of 40 million DDI-related side effects, with their corresponding propensity scores for the evidence level. We downloaded the dataset from their lab repository in February 2020.

**NDF-RT**

NDF-RT [14] is the National Drug File – Reference Terminology from the U.S. Department of Veteran Affairs, Veterans Health Administration (VHA). It is an extension of the VHA National Drug File (NDF) that combines its hierarchical drug classification with a multi-category reference model. “VA Drug interactions” was the category extracted, which contains clinical significance concept for each of the DDIs. The version used in this study dates from July 2014 as this was the last release VA was maintaining NDF-RT interactions, resulting in its removal from their posterior updates. Data was downloaded from the National Institutes of Health (NIH) repository in March 2020.

**CREDIBLEMEDS**

CREDIBLEMEDS [15] is a clinically oriented information source that is used to guide clinical decision-making and safe use records drugs with risk of QT prolongation and/or torsades de pointes (TdP). Crediblemeds comprises a small DDI dataset and contains information such as Precipitant and Object drugs, as well as mechanism of action and management options features for each of the DDIs. The newest version of it dates from May 2010 with no identifiers, and it was extracted from [] as they already had parsed the drugs to their Drugbank identifiers.

**INTERAKTION DATABASEN**

The Danish DDI database [16] is an electronic search tool that describes evidence-based interactions documented by clinical/case studies. It provides a common frame of reference for the on the handling of drug interactions in the Danish healthcare system. Data was retrieved from a public document provided by the Danish Medicines Agency. Information provided includes clinical significance, evidence level, and management administration features for each DDI. The version used in this study (XML\_dato\_3 release 2.7) was downloaded in March 2020.

**ONC HIGH-PRIORITY / NON-INTERRUMPTIVE**

ONC HIGH-PRIORITY (ONC-HP) [17] consists of a set of high-severity DDIs for use in electronic health records (EHR). This dataset is characterized by contraindicated and highly clinical significant DDIs. Nevertheless, most of the DDIs included in this resource lack of primary literature supporting their evidence. We retrieve information regarding clinical significance, evidence level, and management options features. It is a one-time updated project published in April 2012.

ONC NON-INTERRUMPTIVE (ONC-NI) [18] consists of a set of low priority DDIs feasible for non-interruptive alerts that aims to reduce alert fatigue for the provider’s workflow of EHRs. We retrieved information regarding the clinical significance and evidence level features. Like ONC HIGH-PRIORITY, this is also a one-time updated project published in September 2012.

These two projects come from research organized by the Office of The National Coordinator for Health Information Technology (ONC) where in both cases the set of DDIs was a consensus between the different commercial drug providers that participated. The projects were already treated by [], parsing their drug names to Drugbank identifiers. DDI datasets were extracted from that source in February 2020.

**DDI CORPUS 2011 / 2013**

DDI CORPUS 2011 [5] relates to a project included in the DDI Extraction 2011 workshop focused on natural language processing (NLP) techniques for drug-drug extraction from text selected from the Drugbank database. This project dates from September 2011

DDI CORPUS 2013 [6] relates to a project included in the SemEval 2013 DDI Extraction challenge for the evaluation of NLP techniques applied to recognition of pharmacological substances and drug-drug extraction from Drugbank and Medline databases. This project dates from July 2013.

Isabel Segura’s Lab carried out both projects and its raw data can be accessed from her own GitHub. Those projects were already treated by [7], parsing their drug names to Drugbank identifiers. DDI datasets were extracted from that source in February 2020.

**NLM CV CORPUS**

NLM CV CORPUS [19] is a DDI corpus used for training NLP techniques that identifies possible DDIs where cardiovascular drugs are involved. This project was already treated by [7], parsing their drug names to Drugbank identifiers. DDI datasets were extracted from that source in February 2020.

**PK CORPUS**

PK CORPUS [20] is a specific-pharmacokinetic DDI corpus from a pharmacokinetic ontology used for DDI text mining analysis in drug product labels. This project was already treated by [7], parsing their drug names to Drugbank identifiers. DDI datasets were extracted from that source in February 2020.

**HIV / HEP / CANCER DRUG INTERACTIONS**

HIV / HEP /CANCER DRUG INTERACTIONS [21] [22] [23] correspond to three DDI resources from the Liverpool university. They report the clinical significance between a selected small amount of disease-specific drugs and a bigger set of standardized drugs, independently of if there is interaction or not. We extracted information related to the clinical significance and the evidence level for each DDI. The three of them were downloaded in April 2020.

**NETWORK ANALYSIS**

Cytoscape [24] is an open source software platform used for visualizing and integrating networks. It provides features called “apps” used for network analysis. For our project, we used “MCODE” [25] app to identify clusters in our DDI network. It is a cluster algorithm that detects densely connected regions in large interaction networks. Although it is mainly used for protein-protein interactions, it can be used for DDI in order to identify significant drug clusters according to a specific DDI feature.

**RESULTS**

**Parsing DDIs and factorizing DDI features**

* Showing the different quantities for each DDI database, giving also the owner of each one and version used. Adding there also the different features extracted for each database.
* Give a glance (Maybe another picture of the DDI overlap achieved between the different databases (UpsetR, table with percentages?).

**Characterizing DDIs**

* *Given a bit of introduction of ATC levels at the section “materials and methods”. I will go through the different distribution of the ATC 1 level along the different databases, emphasizing in those ones that are designed for a specific class-disease. Also observing how interactions are distributed depending the ATC 1 level (If a ATC class is more likely to interact with itself or the others) (Using also figures created for the second presentation with Søren).*
* *Emphasize on observed ATC levels that are overrepresented or underrepresented. Assesing importance by observing its risk ratio.*
* *Are you planning on using the circular plots figure we did? I think that one is a cool one.*

**Network analysis**

* *Determining important hubs (Tendency to selected ATC level class?), looking to their factorized features and check tendency.*
* *Can be included here the enrichment analysis with the clinical data (To be discussed further next meeting).*
* *Any more ideas of what I could do, or is it enough?*
  + *FOLLOW TEMPLATES OF THE PAPERS YOU HAVE AS LITERATURE/REFERENCE LIST…. USE THEM AS THE TEMPLATE FOR WRITING BY “REWRITING WHAT THEY DID” WITHOUT PLAGIARISM OR GETTING THEIR CONCEPT IDEA*

**DISCUSSION**

* *Comment the future potential of this project, how it could have been improved ( More time for the project, could lead to more databases parsed, COVID situation didn’t allowed me to use some of the lab tools…), where it could be implemented (Mention further analysis that could be done with the clinical data).*

**CONCLUSION**

* *General conclusion of the significant results obtained from the analysis.*

**REFERENCES**

1. Montastruc F, Sommet A, Bondon-Guitton E, et al. The importance of drug-drug interactions as a cause of adverse drug reactions: a pharmacovigilance study of serotoninergic reuptake inhibitors in France. European Journal of Clinical Pharmacology. 2012 May;68(5):767-775. DOI: 10.1007/s00228-011-1156-7.
2. Lopes P, Nunes T, Campos D, Furlong LI, Bauer-Mehren A, Sanz F, et al. (2013) Gathering and Exploring Scientific Knowledge in Pharmacovigilance. PLoS ONE 8(12): e83016. <https://doi.org/10.1371/journal.pone.0083016>
3. Palleria C, Di Paolo A, Giofrè C, et al. Pharmacokinetic drug-drug interaction and their implication in clinical management. *J Res Med Sci*. 2013;18(7):601‐610.
4. Niu J, Straubinger RM, Mager DE. Pharmacodynamic Drug-Drug Interactions. *Clin Pharmacol Ther*. 2019;105(6):1395‐1406. doi:10.1002/cpt.1434
5. Segura-bedmar, I. (2011). Proceedings of the 1st Challenge task on Drug-Drug Interaction Extraction. *Proceedings of the 1st Challenge Task on Drug-Drug Interaction Extraction*, *January*.
6. Herrero-Zazo, M., Segura-Bedmar, I., Martínez, P., & Declerck, T. (2013). The DDI corpus: An annotated corpus with pharmacological substances and drug-drug interactions. *Journal of Biomedical Informatics*, *46*(5), 914–920. <https://doi.org/10.1016/j.jbi.2013.07.011>
7. Ayvaz, S., Horn, J., Hassanzadeh, O., Zhu, Q., Stan, J., Tatonetti, N. P., Vilar, S., Brochhausen, M., Samwald, M., Rastegar-Mojarad, M., Dumontier, M., & Boyce, R. D. (2015). Toward a complete dataset of drug-drug interaction information from publicly available sources. *Journal of Biomedical Informatics*, *55*(May), 206–217. https://doi.org/10.1016/j.jbi.2015.04.006
8. Hines, L.E., Malone, D.C. and Murphy, J.E. (2012), Recommendations for Generating, Evaluating, and Implementing Drug‐Drug Interaction Evidence. Pharmacotherapy, 32: 304-313. doi:[10.1002/j.1875-9114.2012.01024.x](https://doi.org/10.1002/j.1875-9114.2012.01024.x)
9. Wishart, D. S., Feunang, Y. D., Guo, A. C., Lo, E. J., Marcu, A., Grant, J. R., Sajed, T., Johnson, D., Li, C., Sayeeda, Z., Assempour, N., Iynkkaran, I., Liu, Y., MacIejewski, A., Gale, N., Wilson, A., Chin, L., Cummings, R., Le, Di., … Wilson, M. (2018). DrugBank 5.0: A major update to the DrugBank database for 2018. *Nucleic Acids Research*, *46*(D1), D1074–D1082. <https://doi.org/10.1093/nar/gkx1037>
10. Mohammed Ali, Ali Ezzat (). dbparser: DrugBank Database XML Parser. R package version 1.1.3.9000.
11. Kanehisa, M., Goto, S., Furumichi, M., Tanabe, M., & Hirakawa, M. (2009). KEGG for representation and analysis of molecular networks involving diseases and drugs. *Nucleic Acids Research*, *38*(SUPPL.1), 355–360. https://doi.org/10.1093/nar/gkp896
12. <https://www.kegg.jp/kegg/rest/keggapi.html>
13. Tatonetti NP, Ye PP, Daneshjou R, Altman RB. Data-driven prediction of drug effects and interactions. *Sci Transl Med*. 2012;4(125):125ra31. doi:10.1126/scitranslmed.3003377
14. Olvey, E.L., Clauschee, S. and Malone, D.C. (2010), Comparison of Critical Drug–Drug Interaction Listings: The Department of Veterans Affairs Medical System and Standard Reference Compendia. Clinical Pharmacology & Therapeutics, 87: 48-51. doi:10.1038/clpt.2009.198
15. <https://www.crediblemeds.org/healthcare-providers/drug-drug-interaction>
16. <https://laegemiddelstyrelsen.dk/en/sideeffects/danish-drug-interaction-databases/>
17. Phansalkar, S., Desai, A. A., Bell, D., Yoshida, E., Doole, J., Czochanski, M., Middleton, B., & Bates, D. W. (2012). High-priority drugedrug interactions for use in electronic health records. *Journal of the American Medical Informatics Association*, *19*(5), 735–743. https://doi.org/10.1136/amiajnl-2011-000612
18. Phansalkar, S., van der Sijs, H., Tucker, A. D., Desai, A. A., Bell, D. S., Teich, J. M., Middleton, B., & Bates, D. W. (2013). Drug-drug interactions that should be noninterruptive in order to reduce alert fatigue in electronic health records. *Journal of the American Medical Informatics Association*, *20*(3), 489–493. https://doi.org/10.1136/amiajnl-2012-001089
19. Johann Stan, A Machine-Learning Approach for Drug–Drug Interaction Extraction from FDA Structured Product Labels, Presented at the 2014 National Library of Medicine Training Conference, Pittsburgh PA, USA, 17-Jun-2014.
20. Wu, H., Karnik, S., Subhadarshini, A. *et al.* An integrated pharmacokinetics ontology and corpus for text mining. *BMC Bioinformatics* **14,**35 (2013). https://doi.org/10.1186/1471-2105-14-35
21. <https://www.hiv-druginteractions.org/checker>
22. <https://www.hep-druginteractions.org/>
23. <https://cancer-druginteractions.org/>
24. Paul Shannon, 1, Andrew Markiel, 1, Owen Ozier, 2 Nitin S. Baliga, 1 Jonathan T. Wang, 2 Daniel Ramage, 2, Nada Amin, 2, Benno Schwikowski, 1, 5 and Trey Ideker2, 3, 4, 5 (1971). Cytoscape: A Software Environment for Integrated Models. *Genome Research*, *13*(22), 426. https://doi.org/10.1101/gr.1239303.metabolite
25. Bader, G.D., Hogue, C.W. An automated method for finding molecular complexes in large protein interaction networks. *BMC Bioinformatics* **4,**2 (2003). <https://doi.org/10.1186/1471-2105-4-2>