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

**Jorge Hernansanz**1**, Cristina Leal**1, **Gianluca Mazzoni**1**, Søren Brunak1**

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:** First, we generated a complete compendium of DDI data through the integration of XX a wide variety of publicly available drug sources and normalize drugs to ATC, if needed. Second, we characterized DDI from a clinical setting using the DDI information collected in the compendia and performed a network analysis resulting in a better information coverage the HER data was not able to achieve by itself.

**Results:** Across our the databases studied, we found **X** unique drugs and a total of **X** unique DDIs. Drug/chemical normalization to its ATC code reduced our drug output due to the casual exceptions of drug/chemicals lacking of 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. (More to add)

**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) so as it can serve as an introduction for the section of Results.*

**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.*

**DRUGBANK**

The DRUGBANK database [1] 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”, 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 [2] 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, such as the drugs’ mechanism of action and management of DDI (‘contraindicated’, ‘precaution’). The version used in this study (93.0) was downloaded from the KEGG API in February 2020.

**TWOSIDES**

The TWOSIDES project [3] 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 is the National Drug file – Reference Terminology (NDF-RT) from the U.S. Department of Veteran Affairs, Veterans Health Administration (VHA) [4]. 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, regulated by the Veteran Administration (VA), which contains clinical significance concept for each 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 [5] is a clinically oriented information resource that is used to guide clinical decision-making and safe use of medicines specially for drugs with risk of QT prolongation and/or torsades de pointes (TdP). Crediblemedis comprises a small DDI dataset and contains information such as precipitant and object drugs, 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.

**INTERAKTIONSDATABASEN**

The Danish DDI database [6] is an electronic search tool that describes evidence-based interactions documented by clinical/case studies. It sprovides a common frame of reference for 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) [7] 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 retrieved information regarding clinical significance, evidence level, and management options features. It is a one-time updated project published in April 2012.

Next, ONC NON-INTERRUMPTIVE (ONC-NI) [8] consists of a set of low priority DDIs feasible for non-interruptive alerts, aiming 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. These 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 [9] relates to a project included in the DDI Extraction 2011 workshop focused on natural language processing (NLP) techniques for drug-drug extraction from texts selected from the Drugbank database. This project dates from September 2011

DDI CORPUS 2013 [10] 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 [], 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 [11] [12] [13] 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.

* *To be added more databases, and explain the reason of eliminating some for this thesis.*

**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.*

**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?*

**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. 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>
2. 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
3. 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
4. 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
5. <https://www.crediblemeds.org/healthcare-providers/drug-drug-interaction>
6. <https://laegemiddelstyrelsen.dk/en/sideeffects/danish-drug-interaction-databases/>
7. 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
8. 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
9. 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*.
10. 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>
11. <https://www.hiv-druginteractions.org/checker>
12. <https://www.hep-druginteractions.org/>
13. <https://cancer-druginteractions.org/>