

Drug-Drug Interaction Data Source Survey and Linking

Serkan Ayyvaz, MS¹, Qian Zhu, PhD², Harry Hochheiser, PhD³, Mathias Brochhausen, PhD⁴, John Horn, PharmD⁵, Michel Dumontier, PhD⁶, Matthias Samwald, PhD⁷
Richard D. Boyce, PhD³

¹Kent State University, Kent, OH; ²Mayo Clinic, Rochester, MN; ³University of Pittsburgh, Pittsburgh, PA; ⁴University of Arkansas for Medical Sciences, Little Rock, AK; ⁵University of Washington, Seattle, WA; ⁶Stanford University, Palo Alto, CA; ⁷Medical University of Vienna, Vienna, Austria

Abstract

As an initial step towards the goal of a common data model for potential drug-drug interactions, we surveyed the data elements provided by the publicly available sources. Our analysis found that there is very little overlap between or across publicly available resources and that the information provided is very heterogeneous.

Introduction

Health care providers often have inadequate knowledge of what drug interactions can occur, patient specific factors that can increase the risk of harm from an interaction, and how to properly manage an interaction when patient exposure cannot be avoided. As a result, many thousands of lives are negatively affected by preventable drug-drug interactions each year. Addressing these problems is urgent as the majority of United States healthcare organizations strive to include potential drug-drug interaction (PDDI) screening in their strategies to achieve effective use of electronic health records.

We propose a new PDDI knowledge representation paradigm that we hypothesize would reduce preventable medication errors by more effectively synthesizing existing available PDDI knowledge, and more rapidly producing evidence to fill in knowledge gaps. A key component of the new paradigm is the ability to connect PDDI information from multiple sources towards the goal of obtaining more complete understanding of PDDIs. Our objective was to investigate *publicly available* (i.e., non-proprietary) PDDI information sources that may be linkable and evaluate their information coverage. We also sought to survey the data elements provided by each source as a first step toward a common data model for representing PDDIs. Our motivation for focusing on non-proprietary sources was that the number of such sources has grown in recent years and the PDDIs they provide might enhance other widely used public information systems such as Wikidata

Methods

We conducted a search of Google, PUBMED, and Embase to identify sources of PDDI information. The reference lists of relevant articles retrieved by this search were scanned for additional sources. This search was supplemented by a scan of resources provided by the Bioportal, OntoBee, and datahub for drug interaction data sets.

We downloaded public PDDI datasets identified from the aforementioned search that were available in file format or via an API. We then developed a simple PDDI data model (as a Python dictionary) that combined the data elements provided from each source. Custom Python scripts were used to translate the PDDIs listed in each source to the model. The proportion of PDDIs common between and across the downloaded datasets was examined. To enable cross-dataset comparisons, drug identifiers in each dataset were mapped to DrugBank identifiers wherever possible. For datasets where this was not the case, custom mappings were generated by finding “hub” resources on the Semantic Web that enabled a mapping from the PDDI dataset to DrugBank.

Results and Discussion

Our analysis found that there is very little overlap between or across publicly available PDDI resources and that the information each source provides is very heterogeneous. In spite of this, our results suggest that making the sources interoperable will indeed enable a better synthesis of PDDI knowledge and making it easier to identify gaps that can be directly investigated using pharmacoepidemiology. Moreover, combining the information available across the multiple sources into the simple PDDI data model provided much richer description of these interactions. Our results indicate the importance of further research on generating high quality, complete, and consistently updated mappings between the drug terms in these information sources.

Acknowledgements

This work was supported by the NIH/NIGMS (U19 GM61388; the Pharmacogenomic Research Network), the NLM (R01LM011838) and the Agency for Healthcare Research and Quality (K12HS019461).