Citizen + X: Volunteer-Based Crowdsourcing in Science, Public Health, and Government:
Papers from the 2014 HCOMP Workshop

# Testing Pre-Annotation to Help Non-Experts Identify Drug-Drug Interactions Mentioned in Drug Product Labeling

**Andres M. Hernandez** and **Harry Hochheiser**
Department of Biomedical Informatics
University of Pittsburgh
5607 Baum Blvd, PA 15206
Pittsburgh, Pennsylvania

**John Horn**
School of Pharmacy
University of Washington
1959 NE Pacific St
Seattle, Washington

**Rebbeca Crowley** and **Richard D. Boyce**
Department of Biomedical Informatics
University of Pittsburgh
5607 Baum Blvd, PA 15206
Pittsburgh, Pennsylvania

Drug-drug interactions (DDIs) are defined as clinically meaningful alterations in the exposure and response to a drug (called object drug) that has occurred as a result of the co-administration of another drug (called precipitant drug). According a 2011 study, drug-drug interactions (DDIs) are responsible for roughly 220,000 hospital admissions per year (Percha and Altman 2013). However, this number is possible to be an underestimate due to the uncertainty of the impact of DDI on adverse patient outcomes (Becker et al. 2007).

In this study, a system for allowing combination of text mining and crowdsourcing of annotation approaches for detection of DDIs from drug package inserts is presented. Our hypothesis was that a well-designed task for the curation of DDI data extraction from drug product labels using NLP assistance, can help for non-experts annotators to achieve comparable quality to expert annotation for some part of the DDI knowledge curation task. Therefore, non-experts, assisted by machine learning algorithms, might be able to assist experts in curating large corpora. This is the first step in order to evaluate the feasibility of crowdsourcing DDI extraction. For testing our hypothesis, an annotation study was designed to evaluate expert versus non-expert curation performance, and the impact of NLP pre-annotation on precision and recall on both groups.

The design and development of the system and annotation study, consisted of three stages. First, our existing NLP pipeline for DDI extraction was improved, and it was used to preannotate 208 drug product labels with drug mentions and DDIs (Boyce, Gardner, and Harkema 2012). This dataset contained a representative sample of sections that have unambiguously interacting, non-interacting drug pairs, and also no mention of interactions. Secondly, a DDI machine readable representation scheme was created using the Annotation Ontolgy (OA) (Ciccarese et al. 2011). This model allowed us to load the NLP preannotated drug label sections into our plugin for human curation created using the Annotation tool DOMEO (Ciccarese, Ocana, and Clark 2012). Finally, the annotation study was performed along with usability questionnaires for collecting qualitative feedback from annotators.

To our knowledge, this is the first study in comparing experts and non-experts for pharmacokinetic DDI annotation. Results showed lower performance on non-experts compared with expert annotation without the use of NLP, and an improvement of non-expert annotation performance using the NER module of the NLP assistance. Annotators reported that our tool is useful in assisting DDI extraction. However, complete NLP assistance including drug mentioned and DDIs identification is still too complex for practical applications and could impact negatively the annotation performance. Simplification of the workflow for NLP assisted annotation is necessary for scaling our approach.



## Acknowledgments

Research reported in this paper was partially supported by the Fogarty International Center of Global Health of the National Institutes of Health under the grant No. 1D43TW008443-01, National Institute on Aging grant K01AG044433, and National Library of Medicine grant 1R01LM011838-01. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

## References

Becker, M. L.; Kallewaard, M.; Caspers, P. W.; Visser, L. E.; Leufkens, H. G.; and Stricker, B. H. 2007. Hospitalisations and emergency department visits due to drug-drug in-

teractions: a literature review. *Pharmacoepidemiol Drug Saf* 16(6):641–51. Becker, Matthijs L Kallewaard, Marjon Caspers, Peter W J Visser, Loes E Leufkens, Hubert G M Stricker, Bruno H Ch Journal Article Review England Pharmacoepidemiol Drug Saf. 2007 Jun;16(6):641-51.

Boyce, R. D.; Gardner, G.; and Harkema, H. 2012. Using natural language processing to identify pharmacokinetic drug-drug interactions described in drug package inserts. 2391151 206-213.

Ciccarese, P.; Ocana, M.; Garcia Castro, L. J.; Das, S.; and Clark, T. 2011. An open annotation ontology for science on web 3.0. *J Biomed Semantics* 2 Suppl 2:S4. 2041-1480 Ciccarese, Paolo Ocana, Marco Garcia Castro, Leyla Jael Das, Sudeshna Clark, Tim Journal Article England J Biomed Semantics. 2011 May 17;2 Suppl 2:S4. doi: 10.1186/2041-1480-2-S2-S4.

Ciccarese, P.; Ocana, M.; and Clark, T. 2012. Domeo: a web-based tool for semantic annotation of online documents. *Bio-Ontologies 2011*.

Percha, B., and Altman, R. B. 2013. Informatics confronts drug drug interactions. *Trends in pharmacological sciences* 34(3):178–184.