# Overview of the TAC 2018 Drug-Drug Interaction Extraction from Drug Labels Track

Dina Demner-Fushman[1], Kin Wah Fung[1], Phong Do[2], Richard D. Boyce[3], and Travis R. Goodwin[1]

[1] U.S. National Library of Medicine nlmlhclhcques@mail.nih.gov
https://www.nlm.nih.gov/
[2] Office of Health Informatics, U.S. Food and Drug Administration
[3] University of Pittsburgh

**Abstract.** This paper describes the Drug-Drug Interaction Extraction from Drug Labels Track, part of the 2018 Text Analysis Conference (TAC). Participants were provided with an annotated set of interactions-related sections of drug labels and challenged with: (1) extracting mentions of the precipitants, triggers and effects of drug-drug interactions at sentence level; (2) identifying relations between interacting substances and their types; (3) normalizing mentions and relations to several standard terminologies; and (4) determining the unique set of drug-drug interactions across all provided sections of a drug label. Eight teams submitted at least one run, with 26 submissions in total.

## 1 Background

The U.S. Food and Drug Administration (FDA) is responsible for protecting public health by assuring safety, efficacy, and security of all FDA-regulated products, including human and veterinary drugs, prescription and over-the-counter pharmaceutical drugs, vaccines, biopharmaceuticals, blood transfusions, and biological products, among others. FDA and the National Library of Medicine (NLM) have been working together on transforming the content of Structured Product Labeling (SPL) documents for prescription drugs into discrete, coded, computer-readable data that will be made available to the public in individual SPL index documents. Transforming the narrative text to structured information encoded in national standard terminologies is a prerequisite to the effective deployment of drug safety information. Being able to electronically access labeling information and to search and sort that information is an important step toward creation of a fully automated health information exchange system. TAC 2017 addressed one of the important drug safety issues: automated extraction of adverse drug reactions reported in SPLs [1]. An equally important and complex task is automated extraction of drug-drug interaction (DDI) information. Drug-drug interactions can lead to a variety of adverse events, and it has been suggested that preventable adverse events are the eighth leading cause of death in the United States [2].

Structuring drug safety information is a task in which natural language processing (NLP) systems can provide a great benefit to the FDA and medical community in general. The purpose of this TAC track is to test various NLP approaches for their information extraction (IE) performance on drug-drug interactions in SPLs. While the ultimate goal is for NLP systems to extract and code to controlled terminologies the distinct interaction from the drug labels (the standard structured representation for drug interactions), this track also evaluates and provides data for several intermediate tasks, such as extracting entities (substances, interaction triggers and effects) and relations, as well as normalizing the extracted terms and relations to the FDA substance registration system Unique Ingredient Identifiers (UNII), National Drug File Reference Terminology (NDF-RT), SNOMED CT, and NCI Thesaurus pharmacokinetic effects. The results of this track will inform future FDA efforts at automating important safety processes.

### 1.1 Related Work

Earlier work on DDI extraction from SPLs provided some potentially useful training data [3, 4], although none of the previous annotations exactly match the FDA requirements for structuring DDIs for the SPL index files. In addition to extraction of DDIs from SPLs, two information extraction areas are closely related to the DDI TAC 2018 track: extraction of other information from SPLs and extraction of DDI from other types of text, e.g., literature and social media. DDI Extraction Challenges 2011 and 2013 focused on extracting DDI information from the literature [5]. These challenges and datasets facilitated a growing body of research, with the latest recursive neural network model that implements a tree-LSTM architecture achieving 83.8% F1-score for DDI detection and 73.5% F1-score for interaction type classification [6]. Other types of information that need to be extracted from SPLs include adverse drug reactions [1], indications [7], use in special populations [8], and several others, e.g., pharmacogenomics biomarkers or the drug's mechanism of action, that have not been explored yet.

## 2 Data

The TAC 2018 DDI track dataset consists of 325 Structured Product Labels, in which most or some of the following sections are annotated with drug-drug interactions: *Boxed Warning*, *Clinical Pharmacology*, *Contraindications*, *Dosage and Administration*, *Drug and/or laboratory test interaction*, *Drug Interactions*, *Precautions*, *Warnings and Precautions* and *Warnings*.

The training set includes a TAC-specific training set containing 22 drug labels in XML format that exactly follows the evaluation schema and annotation requirements. All interactions are annotated with respect to the Labeled Drug, i.e., the drug for which the SPL was published. The annotations in the training set were generated semi-automatically and might be missing some interactions.

FDA experts and NLM staff and volunteers manually corrected the automatically extracted entities and relations using the interface in Fig. 1. Additional 180 labels were also available for training. These labels were fully manually annotated by NLM in a comparable format in a prior effort related to the NLM-FDA collaboration.

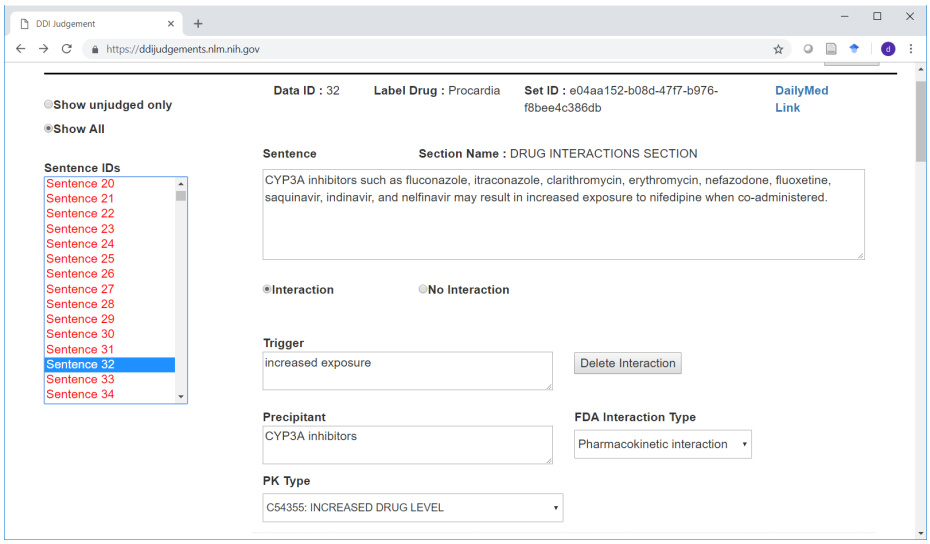


**Fig. 1.** DDI annotation interface. The online interface for registered users to annotate label sentences assigned to them. The full SPL can be reached using the DailyMed link in the upper right corner.

The dataset includes two test sets that are fully manually annotated by FDA, NLM and University of Pittsburgh using the guidelines finalized before annotation[4]. The first test set of 57 labels contains all of the above sections except *Clinical Pharmacology*. The second test set of 66 labels provides annotations from only the *Drug Interactions* and *Clinical Pharmacology* sections.

### 2.1 Annotations

*Entity Annotations* The following entities are annotated in the gold standard:

**Precipitant** – A substance interacting with the Labeled Drug could be another drug, a drug class or a non-drug substance (e.g., *alcohol*, *grapefruit juice*.)
**Trigger** – A word or phrase indicating an interaction event.
**SpecificInteraction** – Results of interactions, e.g., *severe hyperkalemia*.

[4] `https://bionlp.nlm.nih.gov/tac2018druginteractions/DDIvalidationGuidelines.docx`

*Relation Annotations* The following relations connect the above entities in an Interaction. Each relation is limited to a specific subset of entity types.

**Pharmacokinetic interactions (PK)** between the Labeled Drug and the precipitant are indicated by Triggers, e.g., *reducing diuretic absorption*, and other phrases indicating increases / decreases in function measurements.

**Pharmacodynamic (Specific) interactions** between the Labeled Drug and the precipitant are indicated by Triggers, e.g., *potentiate* or *increased risks* and result in SpecificInteraction.

**Unspecified interactions** are indicated by Triggers, e.g., *avoid use*.

*Normalization* The entities and interactions are mapped as follows:

- The interacting substances are mapped to UNII.
- Drug classes are mapped to NDF-RT NUI.
- The effect of the interaction is mapped to SNOMED CT, if it is a medical condition.
- Pharmacokinetic effects are mapped to National Cancer Institute Thesaurus codes.

*Interaction listing* The ultimate goal is to know which interactions are in the labels, such that the interactions may be linked to structured knowledge sources. An interaction mentioned several times should not necessarily carry more weight than an interaction mentioned once. To test the systems on finding distinct interactions, the gold standard contains a list of unique normalized interactions aggregated at the document level.

## 3 Tasks

The track contained four specific tasks, each one potentially building upon the previous tasks:

**Task 1** Extract Mentions of Interacting Drugs/Substances, interaction triggers and specific interactions at sentence level. This is similar to many NLP named entity recognition (NER) evaluations.

**Task 2** Identify interactions at sentence level, including: the interacting drugs, the specific interaction types: pharmacokinetic, pharmacodynamic or unspecified, and the outcomes of pharmacokinetic and pharmacodynamic interactions. This is similar to many NLP relation identification evaluations.

**Task 3** Normalization/Linking task. Normalize the interacting substances to FDA Substance Registration System UNII, and the drug classes to NDF-RT NUI. Normalize the consequence of the interaction to SNOMED CT if it is a medical condition. Normalize pharmacokinetic effects to National Cancer Institute Thesaurus codes.

**Task 4** Generate a global list of distinct interactions in normalized form for each label.

Tasks 1, 2 and 3 correspond to traditional NLP information extraction (IE) and entity linking tasks, while Task 4 involves document-level aggregation. While the tasks were designed to build on each other, participation was optional on a per-task basis. See Fig. 2 and Fig. 3 for examples of the sentence- and document-level annotations expected from the participating systems.

```
<Sentence id="8119" LabelDrug="Zoloft" section="34070-3">
  <SentenceText>
    ZOLOFT is contraindicated in patients: Taking, or within 14 days of stopping, MAOIs, (including the
    MAOIs linezolid and intravenous methylene blue) because of an increased risk of serotonin syndrome.
  </SentenceText>
  <Mention id="M23" type="Trigger" span="162 14" str="increased risk"/>
  <Mention id="M18" type="Precipitant" span="78 5" str="MAOIs" code="n0000000184"/>
  <Mention id="M25" type="SpecificInteraction" span="162 36" str="increased risk of serotonin syndrome"
  code="NO MAP"/>
  <Mention id="M21" type="Precipitant" span="120 26" str="intravenous methylene blue" code="N0000007449"/>
  <Mention id="M24" type="Precipitant" span="106 9" str="linezolid" code="ISQ9I6J12J"/>
  <Interaction id="I9" type="Pharmacodynamic interaction" trigger="M23" precipitant="M18" effect="M25"/>
  <Interaction id="I10" type="Pharmacodynamic interaction" trigger="M23" precipitant="M21" effect="M25"/>
  <Interaction id="I11" type="Pharmacodynamic interaction" trigger="M23" precipitant="M24" effect="M25"/>
</Sentence>
```

**Fig. 2.** Sentence-level annotations of pharmacodynamics interactions between Zoloft and Monoamine oxidase inhibitors (MAOIs). Three precipitants cause the same effect indicated by the same trigger, which results in three annotated interactions.

```
<LabelInteraction type="Unspecified interaction" precipitant="monoamine oxidase inhibitors"
precipitantCode="N0000000184"/>
<LabelInteraction type="Pharmacodynamic interaction" precipitant="pimozide" precipitantCode="1HIZ4DL86F"
effect=" 44103008: Ventricular arrhythmia (disorder)"/>
<LabelInteraction type="Pharmacodynamic interaction" precipitant="pimozide" precipitantCode="1HIZ4DL86F"
effect="111975006: Prolonged QT interval (finding)"/>
<LabelInteraction type="Pharmacokinetic interaction" precipitant="pimozide" precipitantCode="1HIZ4DL86F"
effect="C54357"/>
```

**Fig. 3.** Document-level annotations of all types of interactions between Zoloft, MAOIs and pimozide.

## 4 Evaluation

Participants submitted system results on the two test sets that differ in the number of annotated sections: in Test set 1, all sections except *Clinical Pharmacology* were annotated, if available; in Test set 2, the *Clinical Pharmacology* and *Drug Interactions* sections were annotated. We evaluated the sets independently to see if adding a section that was not annotated in the training set will influence the results.

The evaluation measures were:

**Task 1** Precision/Recall/F1-measure on annotated entities (triggers, substances and effects) using IE-style measurement (i.e., offset-dependent). Both mentions with type and without type were evaluated. The primary evaluation

metric was micro-averaged F1 across the exact matched entity-level annotations with type.

**Task 2** Precision/Recall/F1-measure on relations. Both the full relation (all elements of interaction, i.e., the precipitant, the trigger, the effect and the interaction type) and the presence of relations were evaluated, both with and without type. The primary evaluation metric was micro-averaged F1 across full relations with type.

**Task 3** Precision/Recall/F1-measure on linking entities to the specific terminologies. The primary evaluation metric was F1 macro-averaged across labels.

**Task 4** SPL-level Precision/Recall/F1-measure on unique normalized interactions. The primary evaluation metric was F1 macro-averaged across labels.

## 5 Participants

BUPT-PRIS *Pattern Recognition and Intelligence System Lab, Beijing University of Posts and Telecommunications.* The team participated in Task 1 using bidirectional LSTM-CRF system.

gwm *Institute of Technology Tallaght Dublin, Ireland.* The team participated in Task 1 using a third-party bidirectional LSTM-CRF system with word and characters embeddings [9]. Additional dependency-based information was integrated into word embeddings.

HIKE_DCD_ZJU *DCD Lab, Zhejiang University, China.* The team participated in Tasks 1 and 2 using an encoder/decoder to recognize precipitants first. Then the same architecture was used to extract effects of specific interactions and triggers for Specific, PK and Unspecified interactions. This architecture jointly learned named entity types such that triggers were learned as Specific, PK or Unspecified trigger. For each sentence, the inputs to CNN were word and character embeddings, and capitalization features. For precipitant mention prediction, position features of label drug were added to inputs. The interactions were derived from the extracted named entities based on rules. For PK interactions, a rule based system assigned the interaction code. In addition to provided training data, the team downloaded DailyMed labels, preprocessed the raw texts and annotated them manually, following the official annotation guidelines. The team sampled several sentences from each of the downloaded labels, annotating a total of 1148 sentences. It turned out, that the manually annotated sentences were extracted from some of the test set labels, but no more than 3 or 4 sentences from each.

IBMResearch *IBM research.* The only team that participated in all four tasks. This is also the only team that submitted valid XML documents in the exact required format. The team converted the 180 training labels to TAC format, and, due to the specifics of the 180-set annotations, decided to identify specific interactions and their triggers jointly, and split the two in a post-processing

step. For Tasks 1 and 2, the team used a BiLSTM-CRF model trained on the 180 labels to recognize all entities. Then another BiLSTM-CRF model used words, part-of-speech tags, dependency features and type features to recognize interaction spans. Next, a Piecewise Attention-LSTM model determined the relations between recognized Biomedical Entities and Interactions, using words, part-of-speech tags, dependency features, type features and positional indicators. The team pretrained embeddings using all FDA SPLs and used the embeddings to initialize the models. The 22-label set was used as the development dataset. In post-processing, entities that do not participate in a relation were removed. All references to the Labeled drug and its class were also removed. Finally, a hybrid linguistic approach that combined shallow parsing and syntactic simplification with pattern matching was used to extract triggers from the recognized interactions and to further restore discontinuous spans.

For Tasks 3 and 4, the team used learning to rank to select the best term with the highest-ranking score from the corresponding knowledge source. First, Lucene BM25 model was used to retrieve the top 10 candidate terms for a given mention. Then, for each mention-candidate term pair, four scores were computed: BM25 ranking score, Jaccard similarity score, Longest common subsequence and word2vec similarity. Linear RankSVM then assigned a final ranking score to each candidate term. The top term for each mention was chosen as the normalization for the mention. For PK interactions, a heuristics based on the mention span and its associated relationships was used to match pharmacokinetic effects to the National Cancer Institute Thesaurus codes.

Iles *LIMSI, France.* The team participated in Tasks 1 and 2 using word and character embedding as input to a CNN layer followed by a CRF to identify entity mentions. Logistic regression was used to identify interactions.

joslin93720 *Peking University and Tulane University.* The team participated in Task 1 using dictionary- and rule-based preprocessing and SVM classifiers.

KlickLabs *Klick Labs.* The team participated in Task 1 using a two-step process. First, a sentence classifier predicted whether a given sentence describes an interaction. If the sentence was classified as having an interaction, the noun chunks were considered interacting entities.

ttran *The University of Kentucky and NLM.* The team participated in Tasks 1 and 2, using a BiLSTM for joint entity recognition and interaction type prediction. A CNN with two separate dense output layers (one for PK and one for PD interactions) was used to predict PD effects.

## 6 Results

The results for all runs are shown in Tables 1 – 4. Task 4 was clearly the most challenging (attempted only by one team with the best F1 of 11.8% compared to F1 >= 40% for tasks 1 and 2) This is likely due to the fact that many interactions are repeated in several sections. An optimistic view would be to

assume that the most important and severe distinct interactions were captured because these are usually repeated in all annotated sections. The results on Task 1, although the highest for this evaluation, indicate that this new task is challenging, even compared to the same DDI extraction from the literature, and needs more attention.

**Table 1.** Task 1 (Named Entity Recognition) results sorted by primary F1 score.

| | Test 1 | | | | | | Test 2 | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Primary | | | Relaxed | | | Primary | | | Relaxed | | |
| Run | R(%) | P(%) | F1(%) | R(%) | P(%) | F1(%) | R(%) | P(%) | F1(%) | R(%) | P(%) | F1(%) |
| HIKE_DCD_ZJU1 | **42.4** | 51.2 | **46.4** | **42.6** | 51.5 | **46.6** | **43.8** | 49.9 | **46.7** | **44.2** | 50.6 | **47.2** |
| HIKE_DCD_ZJU3 | 38.3 | **55.2** | 45.3 | 38.5 | **55.5** | 45.5 | 41.0 | **51.5** | 45.7 | 41.2 | **51.9** | 45.9 |
| HIKE_DCD_ZJU2 | 39.8 | 48.6 | 43.8 | 40.0 | 49.0 | 44.0 | 42.0 | 45.0 | 43.5 | 42.4 | 45.7 | 44.0 |
| ttran2 | 37.4 | 29.5 | 33.0 | 37.8 | 29.8 | 33.3 | 40.0 | 36.6 | 38.2 | 40.7 | 37.4 | 39.0 |
| ttran1 | 31.5 | 29.6 | 30.5 | 31.8 | 29.9 | 30.8 | 33.4 | 35.3 | 34.3 | 34.1 | 35.9 | 35.0 |
| IBMResearch2 | 23.2 | 41.9 | 29.9 | 23.4 | 42.2 | 30.0 | 29.3 | 44.6 | 35.4 | 29.5 | 45.0 | 35.6 |
| ttran3 | 27.5 | 28.6 | 28.0 | 27.7 | 29.0 | 28.3 | 30.3 | 34.9 | 32.4 | 31.1 | 35.8 | 33.3 |
| IBMResearch1 | 24.8 | 32.1 | 28.0 | 25.1 | 34.6 | 29.0 | 27.9 | 29.7 | 28.8 | 28.1 | 31.1 | 29.5 |
| gwm1 | 20.4 | 38.0 | 26.5 | 20.5 | 38.2 | 26.7 | 24.0 | 37.8 | 29.4 | 24.1 | 37.9 | 29.5 |
| joslin937201 | 17.0 | 15.9 | 16.4 | 17.9 | 16.7 | 17.3 | 21.9 | 17.1 | 19.2 | 23.0 | 18.1 | 20.3 |
| KlickLabs1 | 17.0 | 6.2 | 9.0 | 23.8 | 8.7 | 12.8 | 15.9 | 5.4 | 8.0 | 21.9 | 7.4 | 11.1 |
| BUPT_pris1 | 7.3 | 4.0 | 5.2 | 9.1 | 5.2 | 6.6 | 0.5 | 0.4 | 0.4 | 0.9 | 0.7 | 0.8 |
| Iles1 | 0.3 | 0.4 | 0.3 | 0.3 | 0.4 | 0.3 | 0.1 | 0.3 | 0.2 | 0.1 | 0.3 | 0.2 |

**Table 2.** Task 2 results (Interaction extraction) sorted by primary F1 score.

| | Test 1 | | | | | | Test 2 | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Primary | | | Relaxed | | | Primary | | | Relaxed | | |
| Run | R(%) | P(%) | F1(%) | R(%) | P(%) | F1(%) | R(%) | P(%) | F1(%) | R(%) | P(%) | F1(%) |
| HIKE_DCD_ZJU3 | **32.8** | **54.4** | **40.9** | 45.5 | **71.7** | 55.6 | 28.2 | **45.4** | 34.8 | 45.4 | **65.2** | 53.5 |
| HIKE_DCD_ZJU1 | 34.5 | 48.9 | 40.5 | **49.6** | 67.0 | **57.0** | **30.7** | 42.7 | **35.7** | **50.7** | 63.3 | **56.3** |
| HIKE_DCD_ZJU2 | 30.9 | 45.1 | 36.7 | 45.5 | 63.2 | 52.9 | 28.7 | 38.6 | 32.9 | 48.2 | 58.1 | 52.7 |
| ttran2 | 21.1 | 22.1 | 21.6 | 38.4 | 40.6 | 39.4 | 22.5 | 24.7 | 23.6 | 44.3 | 49.5 | 46.7 |
| IBMResearch1 | 16.6 | 24.1 | 19.7 | 31.7 | 44.5 | 37.0 | 16.3 | 20.8 | 18.3 | 36.9 | 42.4 | 39.5 |
| IBMResearch2 | 16.1 | 25.2 | 19.7 | 29.8 | 46.6 | 36.3 | 16.8 | 23.0 | 19.4 | 36.9 | 50.1 | 42.5 |
| ttran1 | 18.1 | 21.3 | 19.6 | 33.8 | 38.2 | 35.9 | 17.7 | 22.5 | 19.8 | 37.6 | 44.9 | 40.9 |
| ttran3 | 16.6 | 21.9 | 18.9 | 30.6 | 38.9 | 34.3 | 15.7 | 21.4 | 18.1 | 36.1 | 46.7 | 40.8 |
| Iles1 | 0.1 | 0.2 | 0.1 | 0.2 | 0.5 | 0.3 | 0.0 | 0.0 | 0.0 | 0.2 | 0.4 | 0.2 |

## 7 Conclusion

The goal of the TAC Drug-Drug Interaction Extraction from Drug Labels Track was to evaluate and draw attention to the important problem of identifying the drug interactions described in SPLs. Eight teams submitted a total of twenty six runs across the four tasks. The results clearly indicate that the ultimate goal

**Table 3.** Task 3 results (Normalization) sorted by primary Test 1 macro F1 score.

| | Test 1 | | | | | | Test 2 | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Micro | | | Macro | | | Micro | | | Macro | | |
| Run | R(%) | P(%) | F1(%) | R(%) | P(%) | F1(%) | R(%) | P(%) | F1(%) | R(%) | P(%) | F1(%) |
| IBMResearch1 | **21.7** | 30.8 | **25.5** | **24.0** | **31.9** | **26.4** | 26.0 | 28.8 | 27.3 | **24.7** | 25.0 | 24.1 |
| IBMResearch2 | 20.0 | **32.4** | 24.7 | 20.1 | 31.9 | 23.4 | **26.2** | **30.8** | **28.4** | 23.9 | **26.5** | **24.5** |

**Table 4.** Task 4 results (Distinct normalized label-level interactions) sorted by primary Test 1 macro F1 score.

| | Test 1 | | | | | | Test 2 | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Micro | | | Macro | | | Micro | | | Macro | | |
| Run | R(%) | P(%) | F1(%) | R(%) | P(%) | F1(%) | R(%) | P(%) | F1(%) | R(%) | P(%) | F1(%) |
| IBMResearch2 | **9.67** | **17.35** | **12.42** | 9.7 | **17.4** | **11.83** | **9.59** | **12.66** | **10.91** | **7.79** | **9.74** | **8.45** |
| IBMResearch1 | 8.86 | 14.99 | 11.13 | **9.96** | 15.81 | 11.68 | 8.94 | 11.68 | 10.13 | 7.74 | 9.24 | 8.16 |

of producing index files coded to multiple terminologies fully automatically is unattainable at this time. The results achieved by half of the teams, however, show that automated systems could help FDA produce the files faster using a semi-automated approach. Extraction of drug names generally corresponds to the state-of-the-art established on other text collections, such as clinical text and the literature. Extracting triggers and effects of the interactions proved to be somewhat harder, as reported by the IBM team. For the most part, results on the second test set indicate there are some variations in how the interactions are described in the *Clinical Pharmacology* section and the absence of the training data for that section might explain the lower numbers for tasks 2,3, and 4 on the second test set. The results of this evaluation have already informed the FDA and NLM collaboration on the the next steps. We hope the availability of the training and test collections will further encourage research of this imprtoant problem.

## Acknowledgements

The organizers would like to thank the corpus annotators: Mark Sharp (who finalized the guidelines, annotated a large part of the test sets and provided quality assurance as NLM Special Volunteer), Phong Do, Farinaz Beniesfahany, Wujin Kim, Mohammed Abuassi, Melissa Teng, Markos Gebru, Jessica Kim, Julia Xu, Kin Wah Fung, Britney Ann Stottlemyer and Amy Grizzle. We would like to thank Soumya Gayen for developing the annotation interface and ensuring its availability during annotation.

This project was primarily supported through an Inter-agency Agreement between the U.S. Food and Drug Administration (FDA) and the U.S. National Library of Medicine (NLM), part of the National Institutes of Health (NIH). This work was also partially supported by the intramural research program at

the National Library of Medicine. Research reported in this publication was partially supported by the National Library of Medicine of the National Institutes of Health under award number R01LM011838. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

## References

1. Roberts, K., Demner-Fushman, D., Tonning, JM.: Overview of the TAC 2017 Adverse Reaction Extraction from Drug Labels Track. `https://tac.nist.gov/publications/2017/additional.papers/TAC2017.ADR_overview.proceedings.pdf`. Last accessed 30 Nov 2018
2. Goldstein, J., Jaradeh, I., Jhawar, P., Stair, T. ED Drug-Drug Interactions: Frequency & Type, Potential & Actual, Triage & Discharge. The Internet Journal of Emergency and Intensive Care Medicine **8**(2), (2004)
3. Boyce, RD., Horn, JR., Hassanzadeh, O., De Waard, A., Schneider, J., Luciano, JS., Rastegar-Mojarad, M., Liakata M. Dynamic enhancement of drug product labels to support drug safety, efficacy, and effectiveness. Journal of Biomedical Semantics **4**(5), (2013)
4. Ayvaz, S., Horn, J., Hassanzadeh, O., Zhu, Q., Stan, J., Tatonetti, NP., Vilar, S., Brochhausen, M., Samwald, M., Rastegar-Mojarad, M., Dumontier, M., Boyce, RD. Toward a complete dataset of drug-drug interaction information from publicly available sources. J Biomed Inform. **6**(55), (2015)
5. Segura-Bedmar, I., Martnez, P., Zazo, MH. Semeval-2013 task 9: Extraction of drug-drug interactions from biomedical texts (ddiextraction 2013). In Proceedings of the Seventh International Workshop on Semantic Evaluation (SemEval 2013)
6. Lim, S., Lee, K., Kang, J. Drug drug interaction extraction from the literature using a recursive neural network. PLoS ONE **13**(1), (2018)
7. Fung, KW., Jao, CS., Demner-Fushman, D. Extracting drug indication information from structured product labels using natural language processing. J Am Med Inform Assoc. **20**(3), (2013)
8. Rodriguez, LM., Demner-Fushman D. Automatic Classification of Structured Product Labels for Pregnancy Risk Drug Categories, a Machine Learning Approach. AMIA Annu Symp Proc. (2015)
9. Ma, X., Hovy, E. End-to-end Sequence Labeling via Bidirectional LSTM-CNNs-CRF. in Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (2016)