Overview of the TAC 2017 Adverse Reaction Extraction from Drug Labels Track

Kirk Roberts
School of Biomedical Informatics
University of Texas Health Science Center at Houston

Dina Demner-Fushman
Lister Hill National Center for Biomedical Communications
U.S. National Library of Medicine, National Institutes of Health

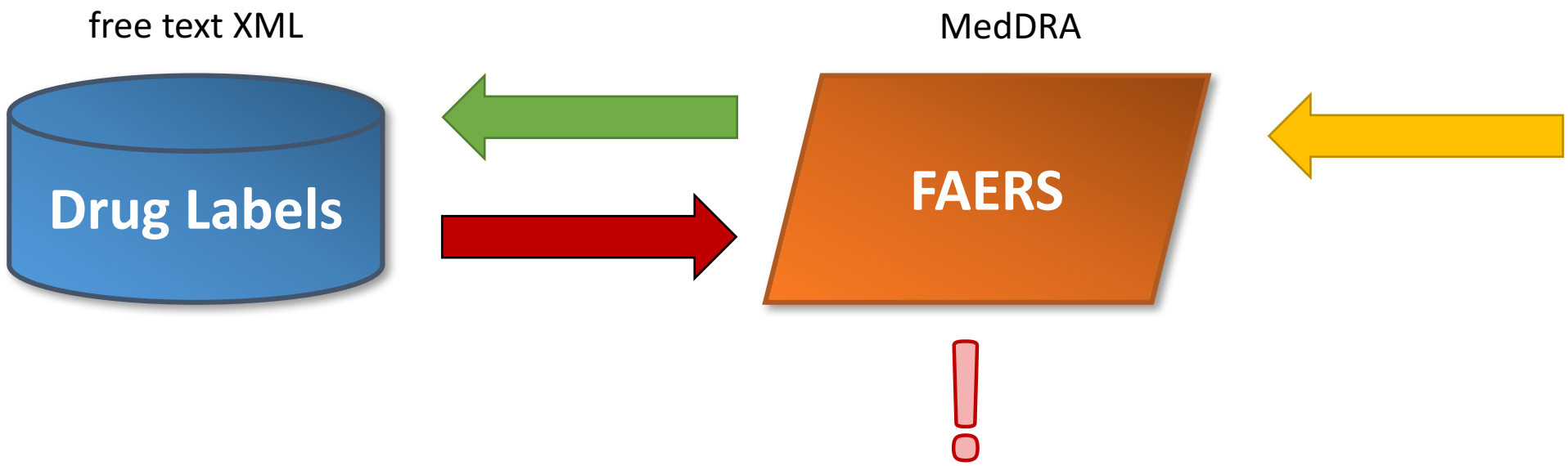
Joe Tonning
Center for Drug Evaluation and Research
U.S. Food and Drug Administration

Background: Adverse Drug Reactions

- In addition to their positive impacts, drugs often have unintended, negative side effects, sometimes very serious
- Not all adverse drug reactions (ADRs) are observed in clinical trials
- Post-marketing pharmacovigilance
- U.S. Food and Drug Administration (FDA) monitors many sources for ADRs
  - FDA Adverse Event Reporting System (FAERS)

Background: Adverse Drug Reactions

- Primary knowledge source for known ADRs is the set of drug labels (Structured Product Labels, SPLs)
- Produced by drug manufacturers based on FDA specifications



Motivation

- Extract *structured* ADR information from drug labels
  - MedDRA
- Enables automation of time-consuming step in FAERS analysis
- Complex NLP task: break into layers corresponding to typical information extraction (IE) tasks
  - *with annotated data!*
- Evaluate myriad of potential approaches within a shared task

Data

6 ADVERSE REACTIONS

The following are discussed in more detail in other sections of the labeling:

- Hypertension, Hypokalemia, and Fluid Retention due to Mineralocorticoid Excess *[see Warnings and Precautions (5.1)]*.
- Adrenocortical Insufficiency *[see Warnings and Precautions (5.2)]*.
- Hepatotoxicity *[see Warnings and Precautions (5.3)]*.

6.1 Clinical Trial Experience

Because clinical trials are conducted under widely varying conditions, adverse reaction rates observed in the clinical trials of a drug cannot be directly compared to rates in the clinical trials of another drug and may not reflect the rates observed in clinical practice.

Two randomized placebo-controlled, multicenter clinical trials enrolled patients who had metastatic castration-resistant prostate cancer who were using a gonadotropin-releasing hormone (GnRH) agonist or were previously treated with orchiectomy. In both Study 1 and Study 2 ZYTIGA was administered at a dose of 1,000 mg daily in combination with prednisone 5 mg twice daily in the active treatment arms. Placebo plus prednisone 5 mg twice daily was given to control patients.

The most common adverse drug reactions (≥10%) reported in the two randomized clinical trials that occurred more commonly (>2%) in the abiraterone acetate arm were fatigue, joint swelling or discomfort, edema, hot flush, diarrhea, vomiting, cough, hypertension, dyspnea, urinary tract infection and contusion.

The most common laboratory abnormalities (>20%) reported in the two randomized clinical trials that occurred more commonly (≥2%) in the abiraterone acetate arm were anemia, elevated alkaline phosphatase, hypertriglyceridemia, lymphopenia, hypercholesterolemia, hyperglycemia, elevated AST, hypophosphatemia, elevated ALT and hypokalemia.

Study 1: Metastatic CRPC Following Chemotherapy

Study 1 enrolled 1195 patients with metastatic CRPC who had received prior docetaxel chemotherapy. Patients were not eligible if AST and/or ALT ≥2.5× ULN in the absence of liver metastases. Patients with liver metastases were excluded if AST and/or ALT >5× ULN.

Table 1 shows adverse reactions on the ZYTIGA arm in Study 1 that occurred with a ≥2% absolute increase in frequency compared to placebo or were events of special interest. The median duration of treatment with ZYTIGA was 8 months.

Table 1: Adverse Reactions due to ZYTIGA in Study 1

	ZYTIGA with Prednisone (N=791)		Placebo with Prednisone (N=394)	
System/Organ Class Adverse reaction	**All Grades[*] %**	**Grade 3–4 %**	**All Grades %**	**Grade 3–4 %**
Musculoskeletal and connective tissue disorders				
Joint swelling/discomfort[†]	29.5	4.2	23.4	4.1
Muscle discomfort[‡]	26.2	3.0	23.1	2.3
General disorders				
Edema[§]	26.7	1.9	18.3	0.8
Vascular disorders				
Hot flush	19.0	0.3	16.8	0.3
Hypertension	8.5	1.3	6.9	0.3
Gastrointestinal disorders				
Diarrhea	17.6	0.6	13.5	1.3
Dyspepsia	6.1	0	3.3	0
Infections and infestations				
Urinary tract infection	11.5	2.1	7.1	0.5
Upper respiratory tract infection	5.4	0	2.5	0
Respiratory, thoracic and mediastinal disorders				
Cough	10.6	0	7.6	0

Data

- 2,309 drug labels
  - 101 training
  - 99 testing
  - 2,109 unannotated
- **DailyMed** XML → basic XML
  - Only maintain sections
- Three sections of interest: **Adverse Reactions**, **Warnings and Precautions**, and **Boxed Warnings**

Data: Mention-level

- **ADVERSEREACTION**: Defined by the FDA as an undesirable, untoward medical event that can reasonably be associated with the use of a drug in humans. This does not include all adverse events observed during the use of a drug, only those for which there is some basis to believe there is a causal relationship between the drug and the adverse event. Adverse reactions may include signs and symptoms, changes in laboratory parameters, and changes in other measures of critical body function, such as vital signs and ECG.

* can be disjoint span

Data: Mention-level

- **NEGATION**: Trigger word for event negation
- **SEVERITY**: Measurement of the severity of a specific ADVERSEREACTION. This can be qualitative terms (e.g., "*major*", "*critical*", "*serious*", "*life-threatening*") or quantitative grades (e.g., "*grade 1*", "*Grade 3-4*", "*3 times upper limit of normal (ULN)*", "*240 mg/dL*")
- **ANIMAL**: Non-human animal species utilized during drug testing

* can be disjoint span
** only when in relation with ADVERSEREACTION

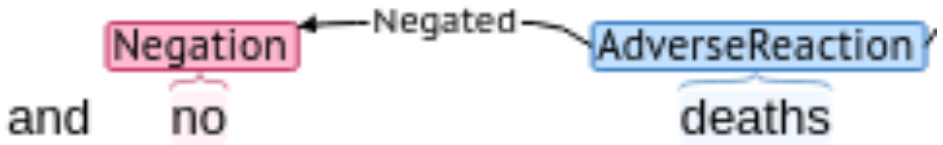
Data: Mention-level

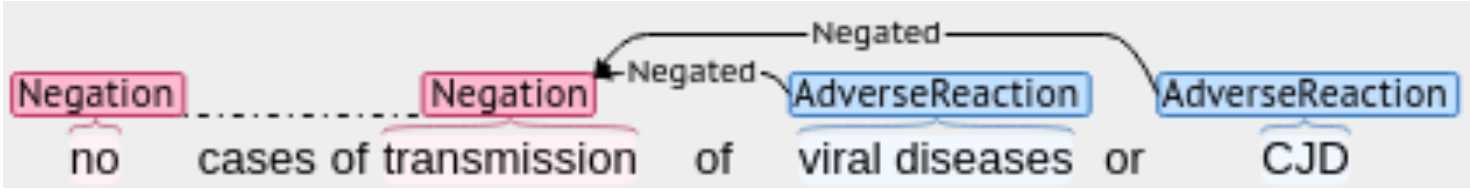
- **FACTOR**: Any additional aspect of an ADVERSEREACTION that is not covered by another mention. Notably, this includes hedging terms (e.g., "*may*", "*risk*", "*potential*"), references to the placebo arm of a clinical trial
- **DRUGCLASS**: The class of drug that the labeled drug is part of. This is designed to capture drug class effects (e.g., "*beta blockers may result in…*") that are not necessarily specific to the particular drug.

* can be disjoint span
** only when in relation with ADVERSEREACTION

Data: Relation-level

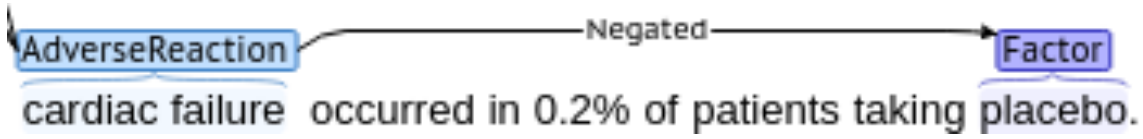
- **Negated**: A NEGATION or FACTOR that indicates the ADVERSEREACTION is absent.

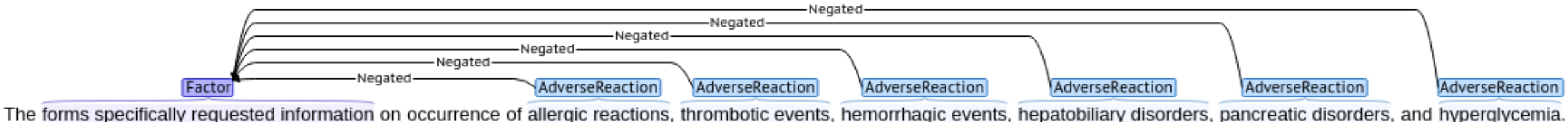




Data: Relation-level

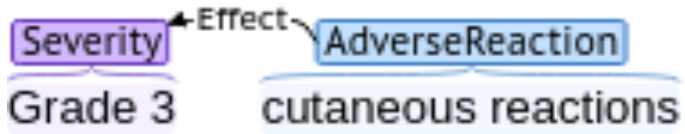
- **Negated**: A NEGATION or FACTOR that indicates the ADVERSEREACTION is absent.

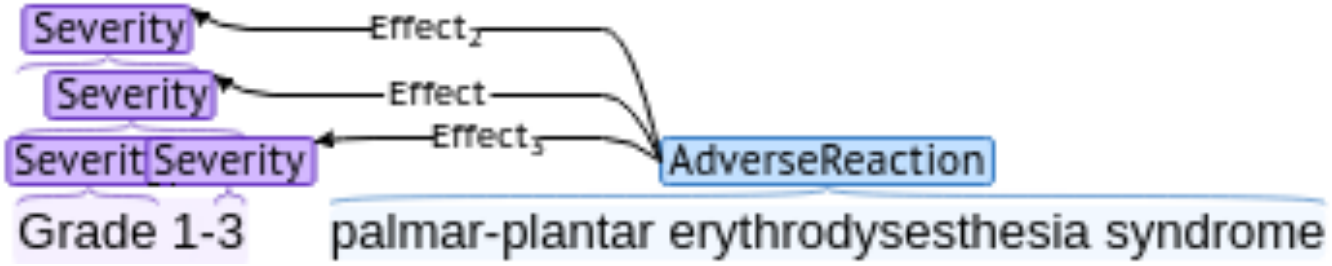




Data: Relation-level

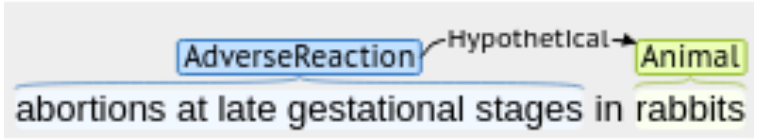
- **Effect**: Indicates SEVERITY of the ADVERSEREACTION.

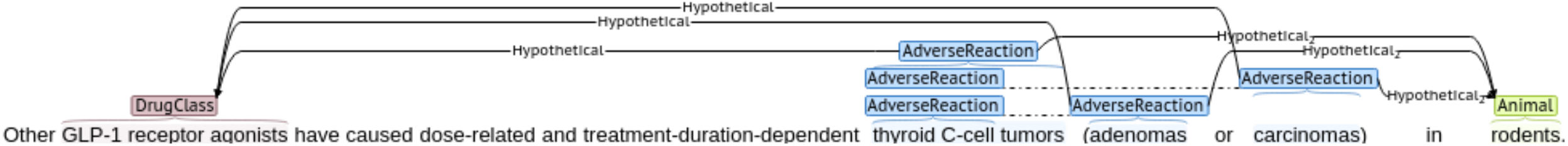




Data: Relation-level

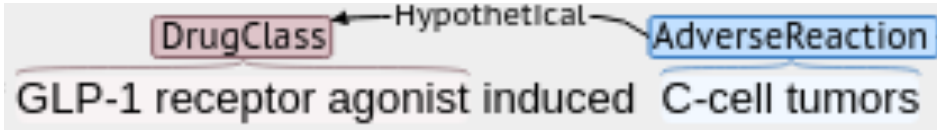
- **Hypothetical**: ANIMAL, DRUGCLASS, or FACTOR that indicate an ADVERSEREACTION is possible, but has not actually been seen in humans.

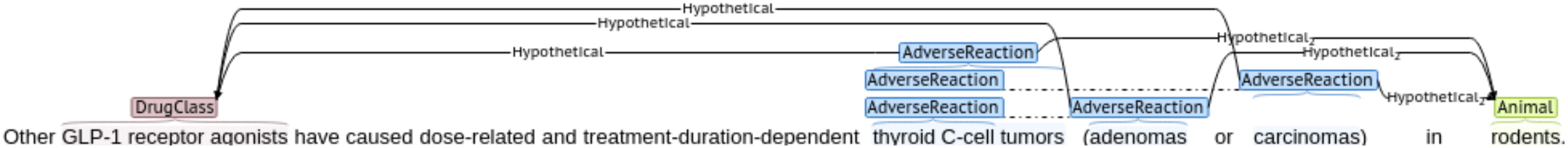




Data: Relation-level

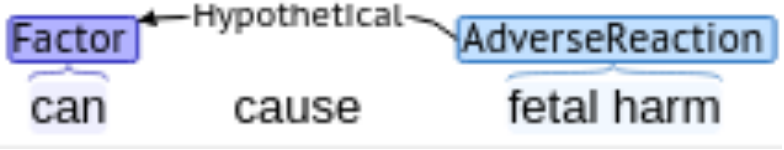
- **Hypothetical**: ANIMAL, DRUGCLASS, or FACTOR that indicate an ADVERSEREACTION is possible, but has not actually been seen in humans.

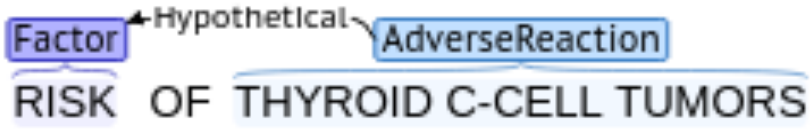


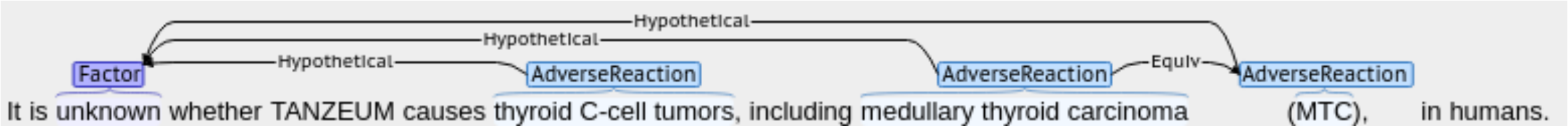


Data: Relation-level

- **Hypothetical**: ANIMAL, DRUGCLASS, or FACTOR that indicate an ADVERSEREACTION is possible, but has not actually been seen in humans.







Data: Document-level

- All unique ADVERSEREACTION strings in the drug label that are *positive*: not NEGATED (with NEGATION or FACTOR) and not HYPOTHETICAL with ANIMAL or DRUGCLASS.
  - Note HYPOTHETICAL with FACTOR is fine
- All unique **MedDRA PT** (Preferred Term) and **LLT** (Lower Level Term) mappings for the above positive reactions.

Data

Annotation	Training	Testing	Total
# SPLs	101	99	200
# Sections	239	237	476
# AdverseReaction	13,795	12,693	26,488
# Animal	44	86	130
# DrugClass	249	164	413
# Factor	602	562	1,164
# Negation	98	173	271
# Severity	934	947	1,881
# Effect	1,454	1,181	2,635
# Hypothetical	1,611	1,486	3,097
# Negated	163	288	451
# Reactions	7,038	6,343	13,381
# MedDRA PTs	7,092	6,409	13,501

Tasks

- **Task 1** [Mention]: ADVERSEREACTION, SEVERITY, FACTOR, DRUGCLASS, NEGATION, ANIMAL
  - micro-average F1 on exact spans
- **Task 2** [Relation]: NEGATED, HYPOTHETICAL, EFFECT
  - micro-average F1 on full relations
- **Task 3** [Document]: positive ADVERSEREACTION strings
  - macro-average F1
- **Task 4** [Document]: MedDRA Preferred Terms
  - macro-average F1

Participants

System	Affiliation	T1	T2	T3	T4
BUPT_PRIS	Beijing University of Posts and Telecommunications	✓	✓		
CHOP	Children's Hospital of Philadelphia	✓		✓	✓
CONDL	University of North Dakota	✓		✓	✓
GN_team	University of Manchester	✓			
IBM_Research	IBM Research	✓	✓		
MC_UC3M	MeaningCloud; Universidad Carlos III de Madrid	✓	✓	✓	✓
Oracle	Oracle Health Sciences			✓	
PRNA_SUNY	Philips Research North America; SUNY Albany	✓	✓	✓	✓
TRDDC_IIITH	TCS Research; IIT Bombay; IIT Hyderabad	✓			
UTH_CCB	University of Texas Health Science Center at Houston	✓	✓	✓	✓

Results

Task 1

System (Run)	Precision	Recall	F1
UTH_CCB (3)	82.54	82.42	82.48
UTH_CCB (2)	80.22	84.40	82.26
UTH_CCB (1)	83.78	79.74	81.71
IBM_Research	80.90	75.30	78.00
CONDL (1)	76.45	77.49	76.97
GN_team (1)	80.19	72.23	76.00
GN_team (2)	76.84	74.36	75.58
PRNA_SUNY (1)	77.71	63.90	70.13
PRNA_SUNY (3)	77.71	63.90	70.13
CONDL (3)	65.19	69.77	67.41
CONDL (2)	65.47	61.40	63.37
PRNA_SUNY (2)	64.25	61.58	62.89
MC_UC3M (1)	54.79	66.33	60.01
MC_UC3M (2)	54.79	66.33	60.01
trddc_iiith	79.14	43.12	55.83
CHOP	57.95	29.64	39.22
BUPT_PRIS	40.47	11.81	18.29

Results

Task 2

System (Run)	Precision	Recall	F1
UTH_CCB (3)	50.24	47.82	49.00
UTH_CCB (1)	51.67	44.45	47.79
UTH_CCB (2)	46.24	48.32	47.26
IBM_Research	48.13	32.54	38.83
PRNA_SUNY (1)	50.48	22.36	30.99
PRNA_SUNY (3)	50.48	22.36	30.99
PRNA_SUNY (2)	31.28	9.34	14.39
MC_UC3M (2)	10.41	10.95	10.67
BUPT_PRIS	0.97	0.38	0.55

Results

Task 3

	Micro			Macro		
System (Run)	P	R	F1	P	R	F1
UTH_CCB (3)	80.97	84.87	82.87	80.69	85.05	82.19
UTH_CCB (1)	82.83	81.76	82.29	82.61	81.88	81.65
UTH_CCB (2)	79.68	85.57	82.52	78.77	85.62	81.39
Oracle (3)	81.18	79.69	80.43	81.47	79.28	79.67
Oracle (2)	82.71	78.05	80.31	82.64	77.73	79.42
Oracle (1)	81.28	79.32	80.28	81.10	78.81	79.20
CONDL (1)	87.77	67.33	76.21	87.34	67.64	75.15
PRNA_SUNY (1)	73.05	69.90	71.44	73.23	68.91	70.29
PRNA_SUNY (3)	73.05	69.90	71.44	73.23	68.91	70.29
MC_UC3M (1)	70.03	71.42	70.71	69.23	72.93	70.13
MC_UC3M (2)	70.03	71.42	70.71	69.23	72.93	70.13
CONDL (2)	70.86	69.76	70.31	70.16	70.29	69.35
CONDL (3)	70.86	69.76	70.31	70.16	70.29	69.35
PRNA_SUNY (2)	59.57	71.91	65.16	58.16	70.96	63.25
CHOP	64.29	39.57	48.99	62.97	39.95	47.99

Results

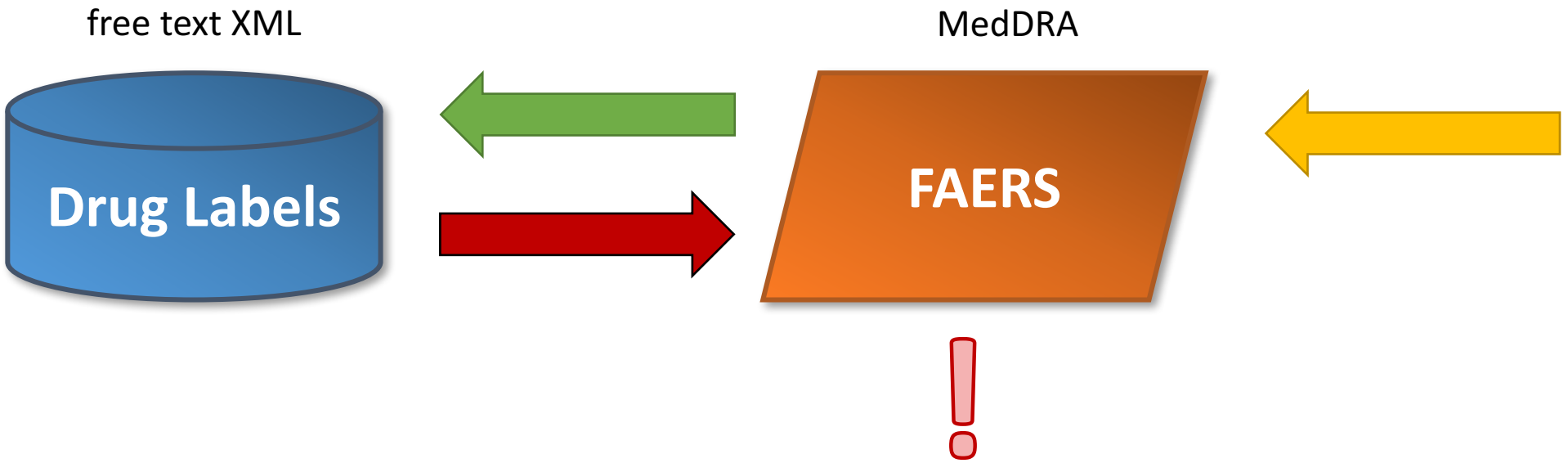
Task 4

	Micro			Macro		
System (Run)	P	R	F1	P	R	F1
UTH_CCB (3)	84.17	89.84	86.91	83.02	89.06	85.33
UTH_CCB (1)	85.00	87.75	86.35	84.04	86.67	84.79
UTH_CCB (2)	82.42	90.78	86.40	80.83	89.90	84.53
CONDL (1)	88.81	77.16	82.58	88.20	75.76	80.50
PRNA_SUNY (1)	86.14	74.89	80.12	85.32	72.76	77.97
PRNA_SUNY (2)	81.55	78.24	79.86	79.80	76.03	77.25
PRNA_SUNY (3)	83.60	74.14	78.59	82.22	71.44	75.87
CONDL (2)	74.56	80.96	77.63	73.06	79.92	75.55
CONDL (3)	74.56	80.96	77.63	73.06	79.92	75.55
MC_UC3M (1)	73.40	80.25	76.67	72.10	80.38	75.29
MC_UC3M (2)	73.40	80.25	76.67	72.10	80.38	75.29
CHOP	71.78	50.14	59.04	70.12	49.84	57.27

Further Evaluation

- In the process of conducting further evaluation based on **post-hoc sample** of outputs on unannotated data
- Chose 50 "**most controversial**" labels, i.e., those with lowest agreement
  - "Hard" labels might better distinguish systems
- Same **manual annotation** process as original 200 labels
- Roughly 2000 ADVERSEREACTIONS on this data
- Analysis to come….

Discussion



Will an ~0.85 F1 system be sufficient for this?

Future Work (FDA)

- A **scalable** system to analyze ADRs across all labels is needed
  - drug safety is not "*one size fits all*"
- Various types of ADRs may be of lesser or greater interest to a researcher or FDA reviewer
  - Pre-clinical studies (ADRs in animals)
  - Pre-market approval (identifying ADRs of concomitant drugs in clinical trials)
  - Post-market pharmacovigilance (e.g., FAERS)

Future Work (FDA)

- **Automation** of some current **manual processes**
  - Analysis of ADRs of concomitant drugs in clinical trials
  - Pharmacovigilance of post-marketing reports
- **Data mining** of ADRs across all labels
  - Determining whether a drug could be **repurposed** (i.e., for a new indication)
  - Finding patterns to predict drug interactions or other toxicity by pharmacologic class or similar chemical moieties

Future Work (NLP)

- Lots of other information in drug labels where **NLP** could be useful
  - ADRs in specific populations
  - Overdose information
  - Drug-drug interactions
  - Clinical trial data
  - Contraindications

Conclusion

- **Goal**: evaluate and draw attention to the important problem of identifying ADRs in drug labels
- Having an accurate list of known ADRs will be of tremendous value to FDA for **pharmacovigilance** and **other activities**
- **Good participation**: T1- 17 submissions; T2- 9 submissions; T3- 15 submissions; T4- 12 submissions
- Top submission on T4: ~85 F1

Acknowledgments

- Funding:
  - FDA/NLM Interagency Agreement: IAA 224-15-3022S
  - NLM Grant: 4-R00-LM012104-02
  - NLM Intramural Research Program
- Annotators: Alan Aronson, Sonya Shooshan, Laritza Rodriguez, Dina Demner-Fushman
- Development: Willie Rogers, Francois Lang
- NIST