Table-to-text Generation for Biomedical Causal Inference

Andrew Bell

FDA Adverse Event Reporting System (FAERS)

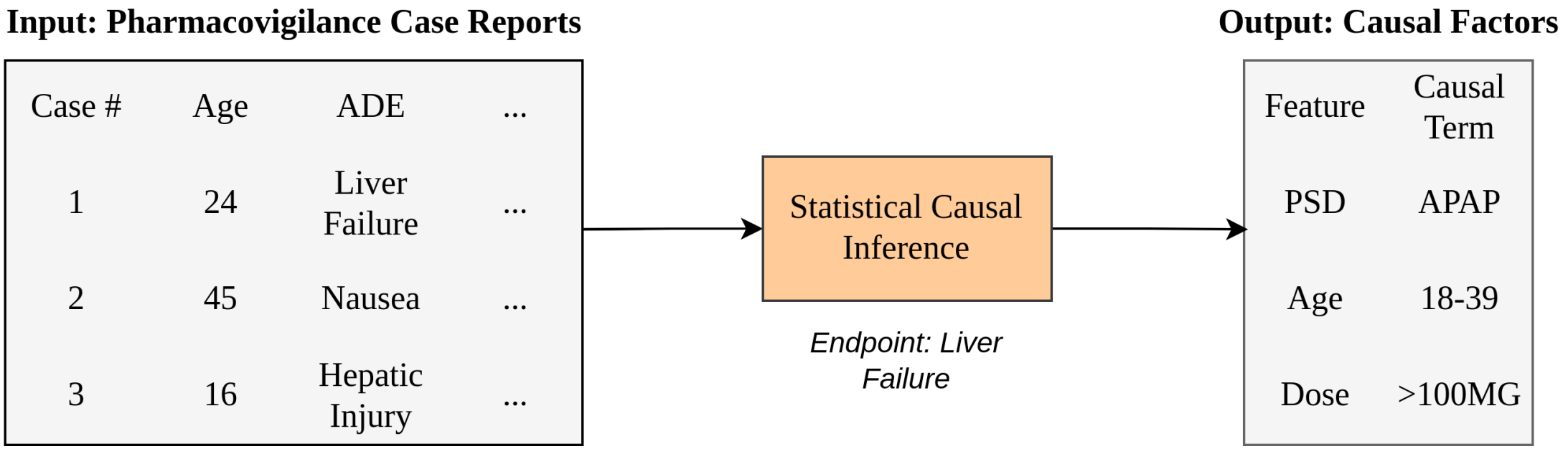
- Over 2 million reports every year
- Example Report:

CaseID	Gender	Age	PSD	ADE	...
20222515	Female	71	Nicotine, Acetaminophen	Anemia, Headache	...

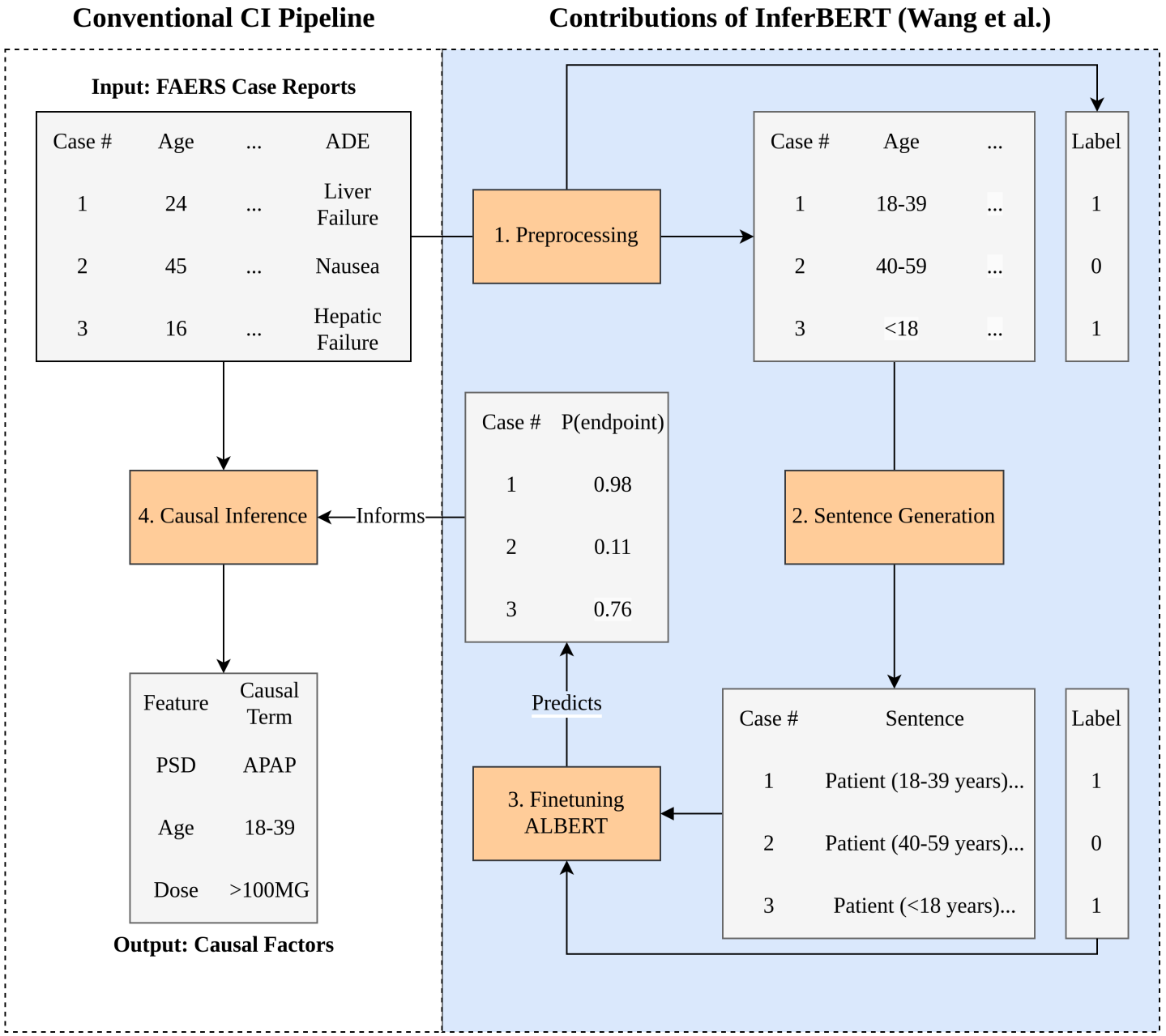
PSD = Primary Suspect Drugs

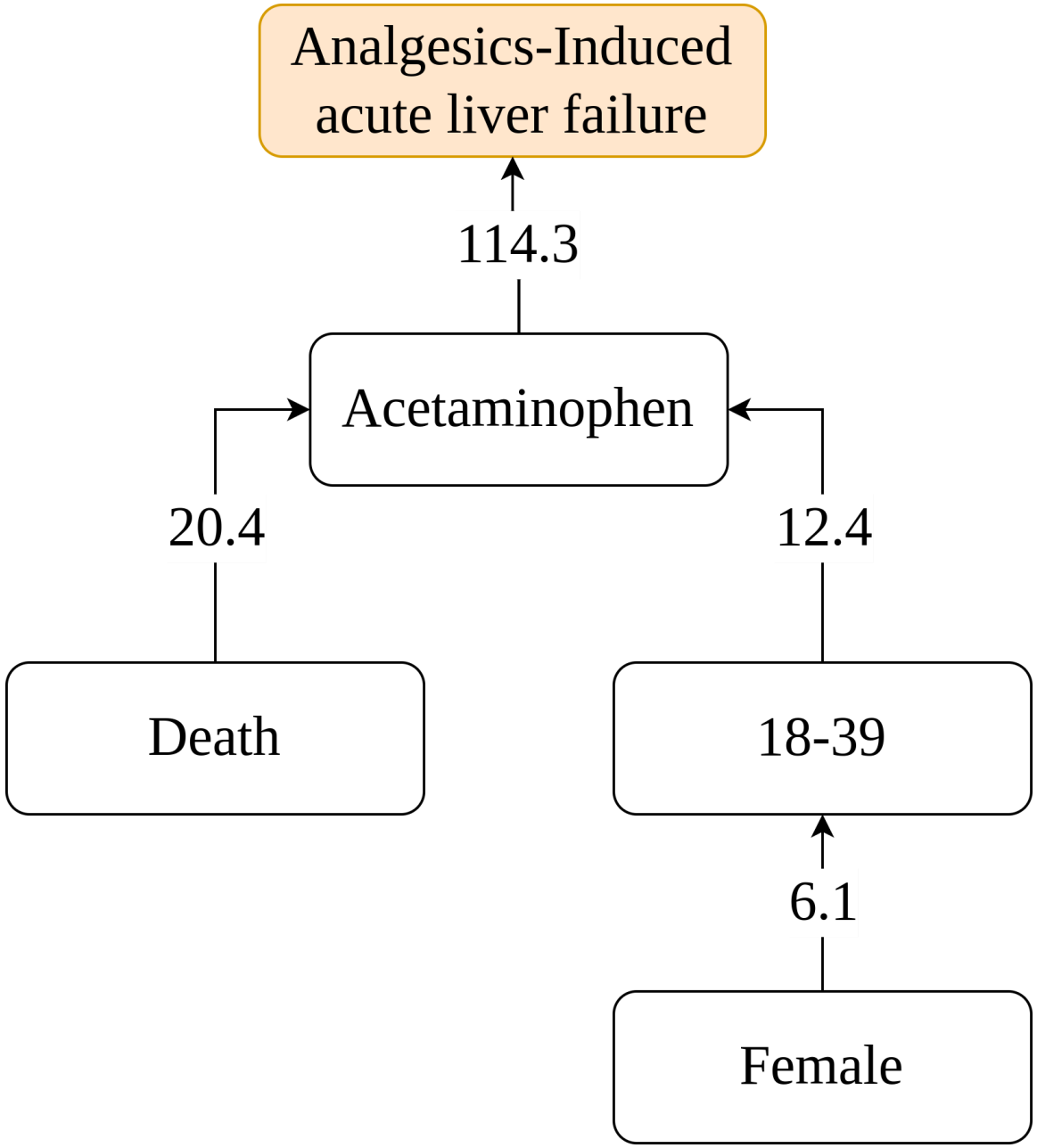
ADE = Adverse Drug Effects

Conventional CI pipeline



InferBERT informs causal inference with a language model (ALBERT)

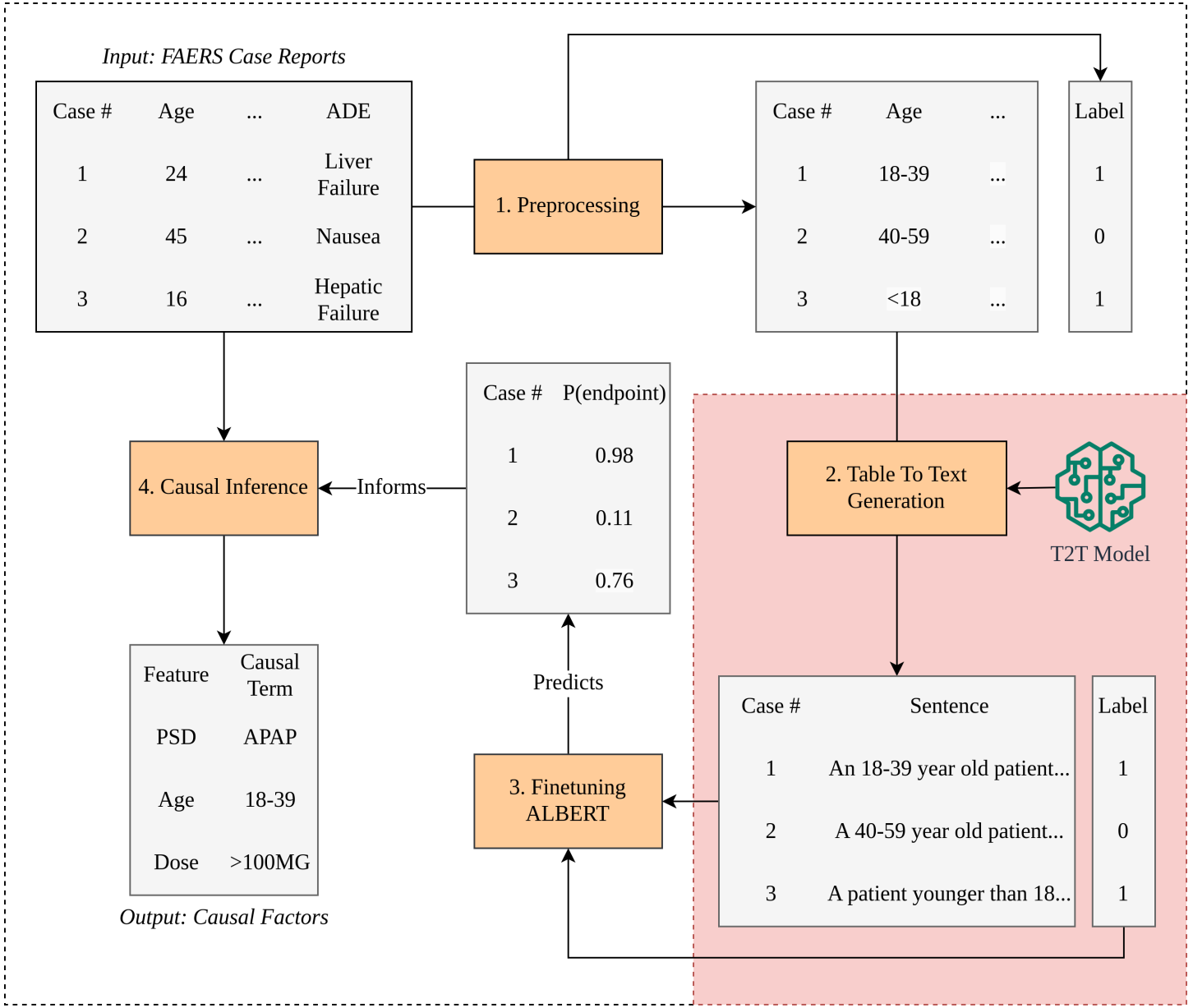




InferBERT Results

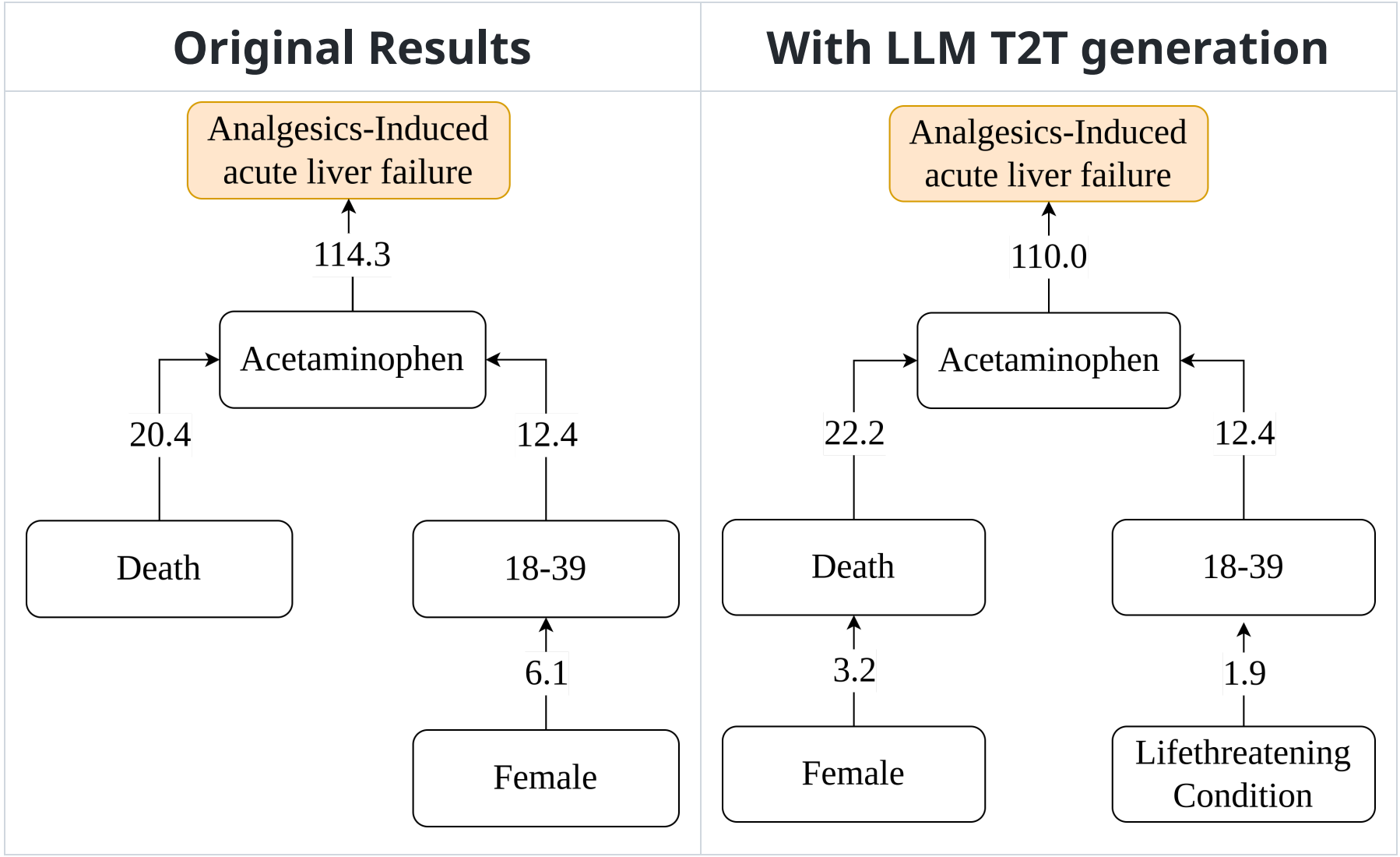
Feature	Term	Z-score
PSD	APAP	114.95
Outcome	Death	77.94
Age	18-39	27.29
Indication	Suicide Attempt	16.85
Gender	Female	10.26
Dose	>100mg	5.01

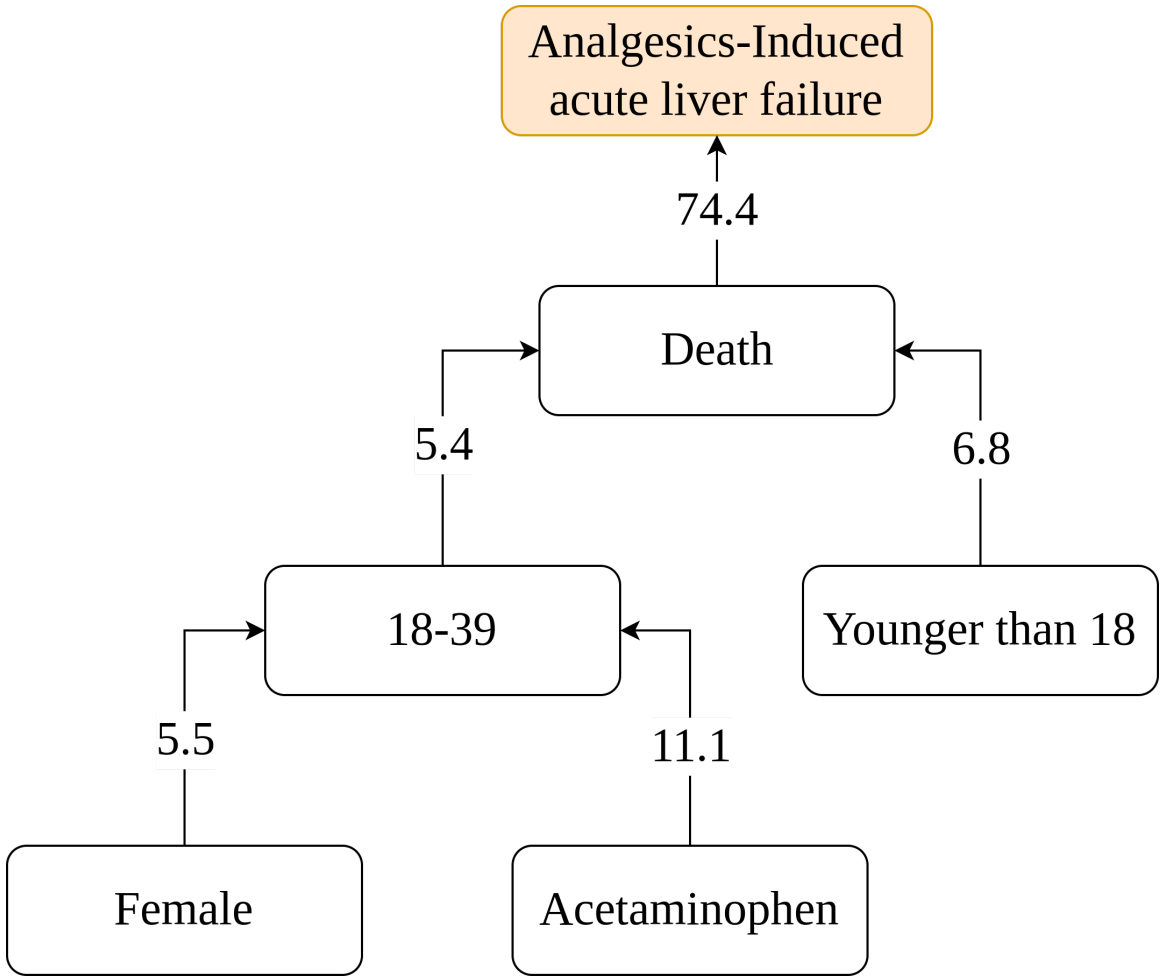
We modify InferBERT using Table-to-text generation



Advantages of our method

- T2T generation can be applied to any input features - This allows us to easily apply the framework to new studies or features
- Using a pre-trained language model, T2T is able to infer implicit relations between report terms
- These relations are made explicit in the generated sentences, improving endpoint prediction of the ALBERT model


Original Results
With LLM T2T generation
Analgesics-Induced acute liver failure
114.3
Acetaminophen
20.4
12.4
Death
18-39
6.1
Female
Analgesics-Induced acute liver failure
110.0
Acetaminophen
22.2
12.4
Death
18-39
3.2
1.9
Female
Lifethreatening Condition



With More Features

Feature	Value	Z-score
SSD	Diphen-hydramine	5.13
Weight	50-70kg	5.75
Route	Intravenous	15.55

SSD= Secondary Suspect Drug

Future Work

- Applying this framework to further case studies
- What are the limits for the number of features/terms we can study in one go?
- What would be the impact of using prompting instead of finetuning for endpoint prediction?
    - Can we use a single model from report $\rightarrow$ prediction?