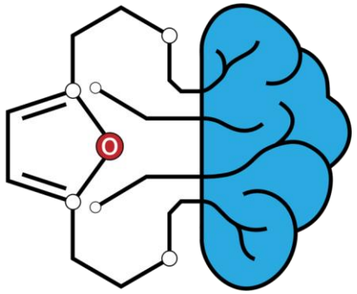
Pharmacovigilance Meets Demographics: Towards Personalized Cardiotoxicity Prediction

Mateusz Iwan, Francesca Grisoni, Marina Garcia de Lomana, Alessandra Roncaglioni



SAR 2025
MILAN 3 - 6 JUNE

Drug Toxicity

- Adverse Drug Reactions (ADRs) are an inherent part of medicines

Drug Toxicity

- Adverse Drug Reactions (ADRs) are an inherent part of medicines

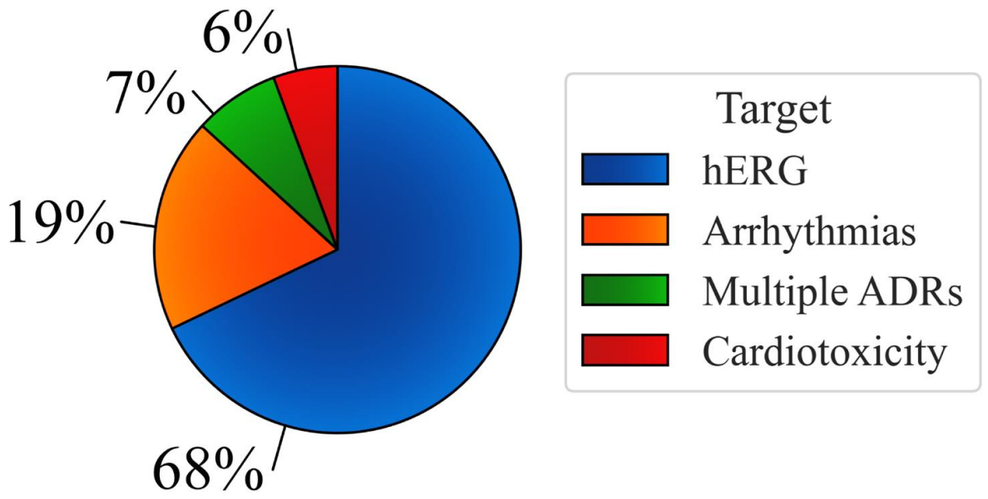
- Reactions vary depending on demographic factors

Drug Toxicity

- Adverse Drug Reactions (ADRs) are an inherent part of medicines
- Reactions vary depending on demographic factors
- Modern clinical trials include more diverse populations

Drug Toxicity

- Adverse Drug Reactions (ADRs) are an inherent part of medicines
- Reactions vary depending on demographic factors
- Modern clinical trials include more diverse populations
- *In silico* models should follow this development



https://doi.org/10.1016/j.amjcard.2009.10.007

Our Approach to DICT

Chemical structures

Standardized drug representations

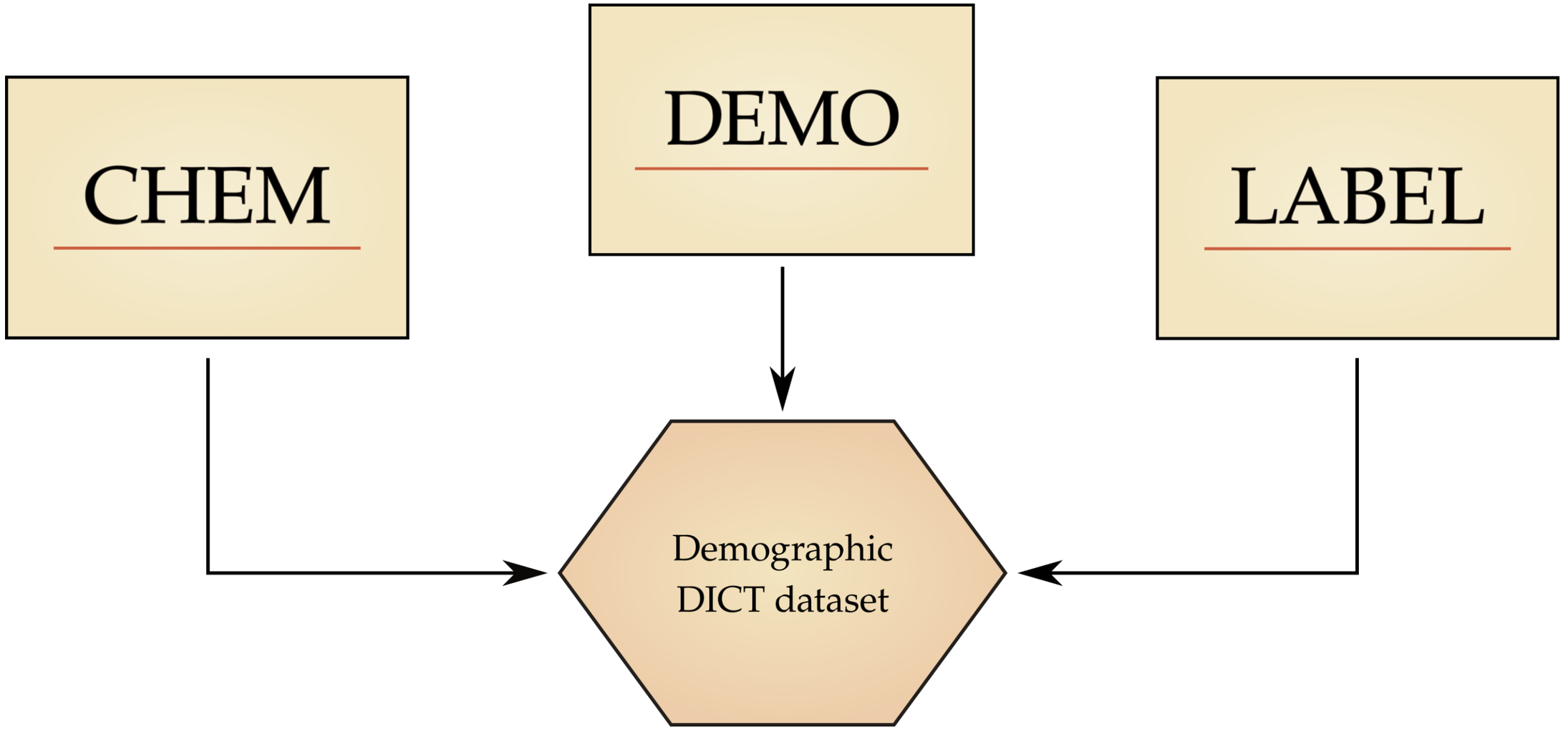
Demographic data

Patient's
Sex, Age, Weight

DICT Labels

Derived from pharmacovigilance sources

Our Approach to DICT

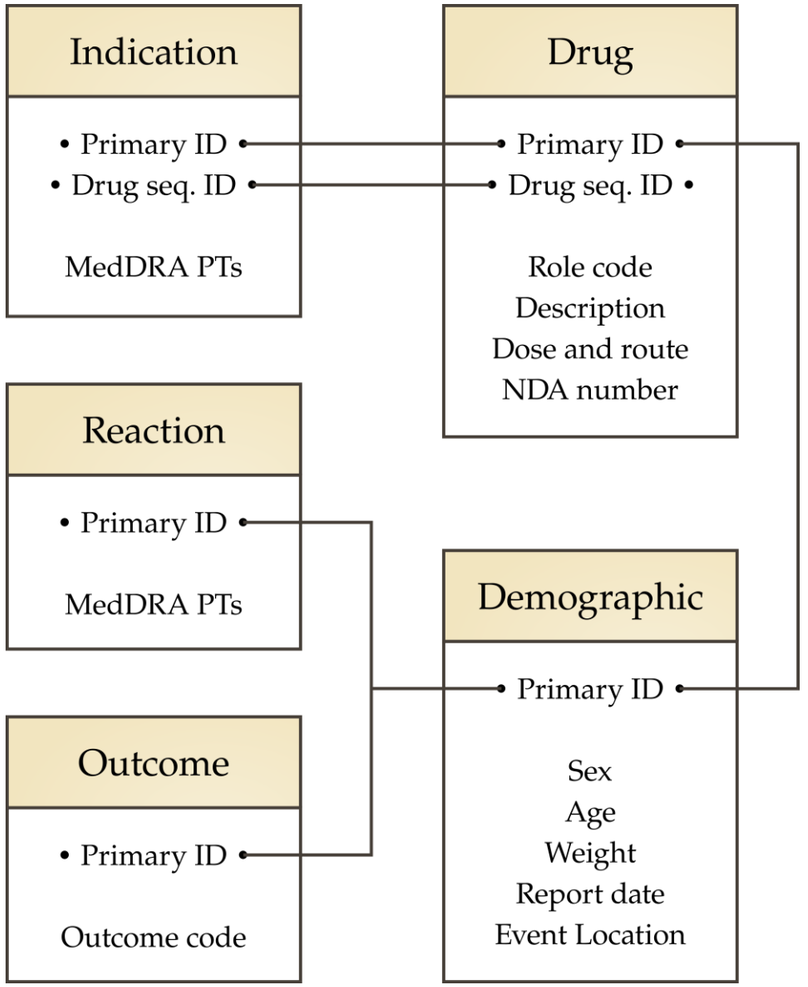


CHEM DEMO LABEL

Our Approach to DICT

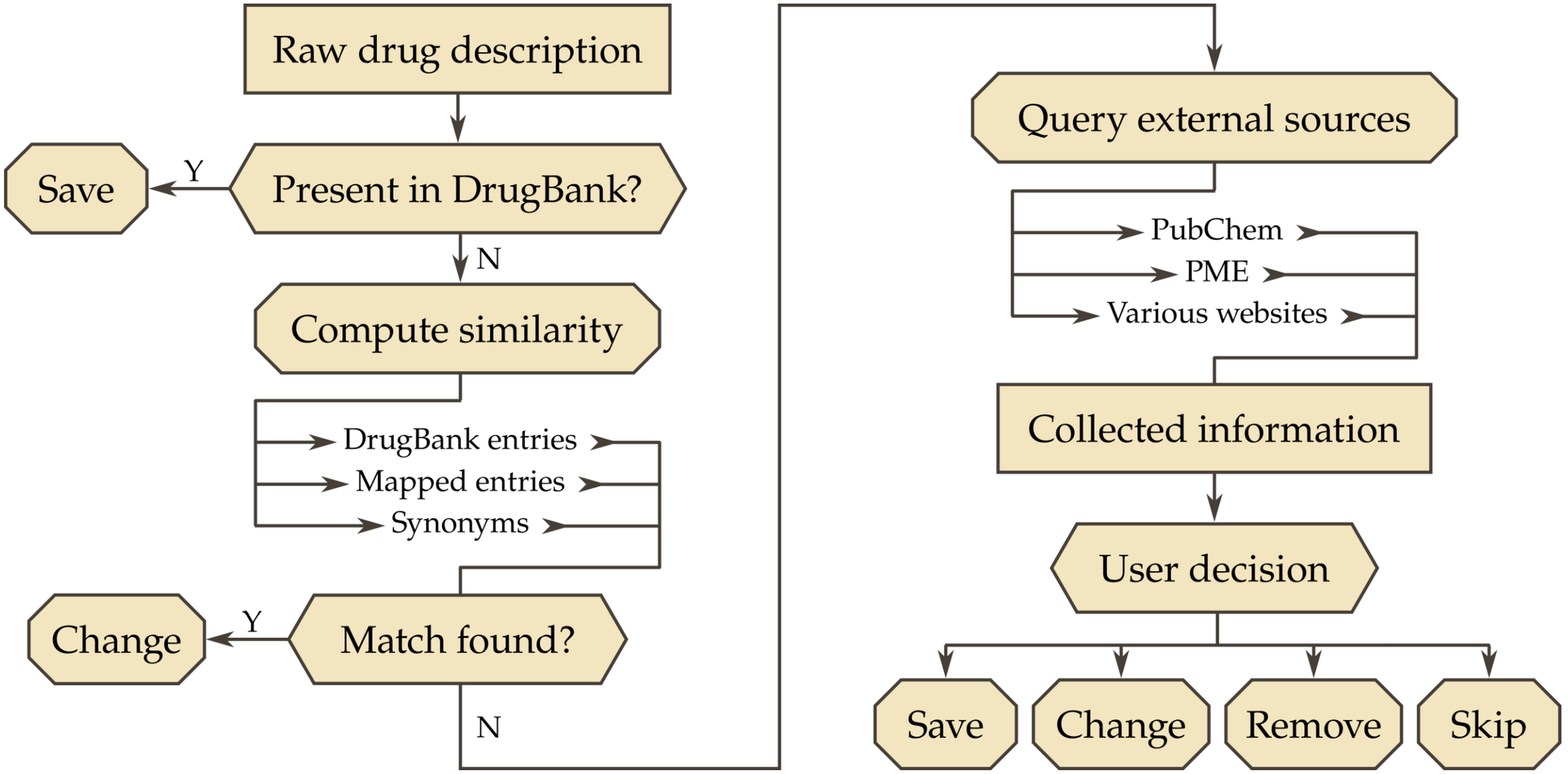
FAERS Database Overview

- Data collection: Q4 2012 – Q3 2024
- Number of unique reports: 17,687,672
- Number of unique drug descriptions: 591,402
- Number of unique adverse effects: 35,966
- Data completeness:
  - Sex: 87.2%
  - Age: 57.2%
  - Weight: 18.9%



https://fis.fda.gov/extensions/FPD-QDE-FAERS/FPD-QDE-FAERS.html

Drug Description Mapping

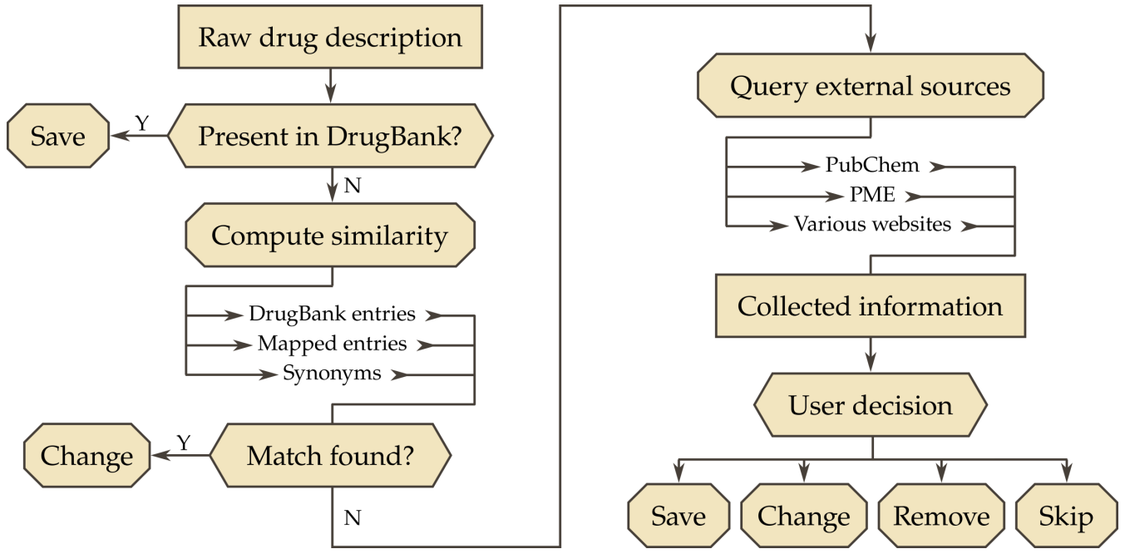


Mapping Summary

- Additional full-record linkage using string similarity

- Preprocessing steps:
  - Removal of non-drug entries
  - Retrieval and standardization of SMILES
  - Calculation of descriptors

- Final dataset summary:
  - 311,451 processed drug descriptions
  - 8,260 unique drug combinations
  - 4,333 unique drug names
  - 3,618 SMILES strings

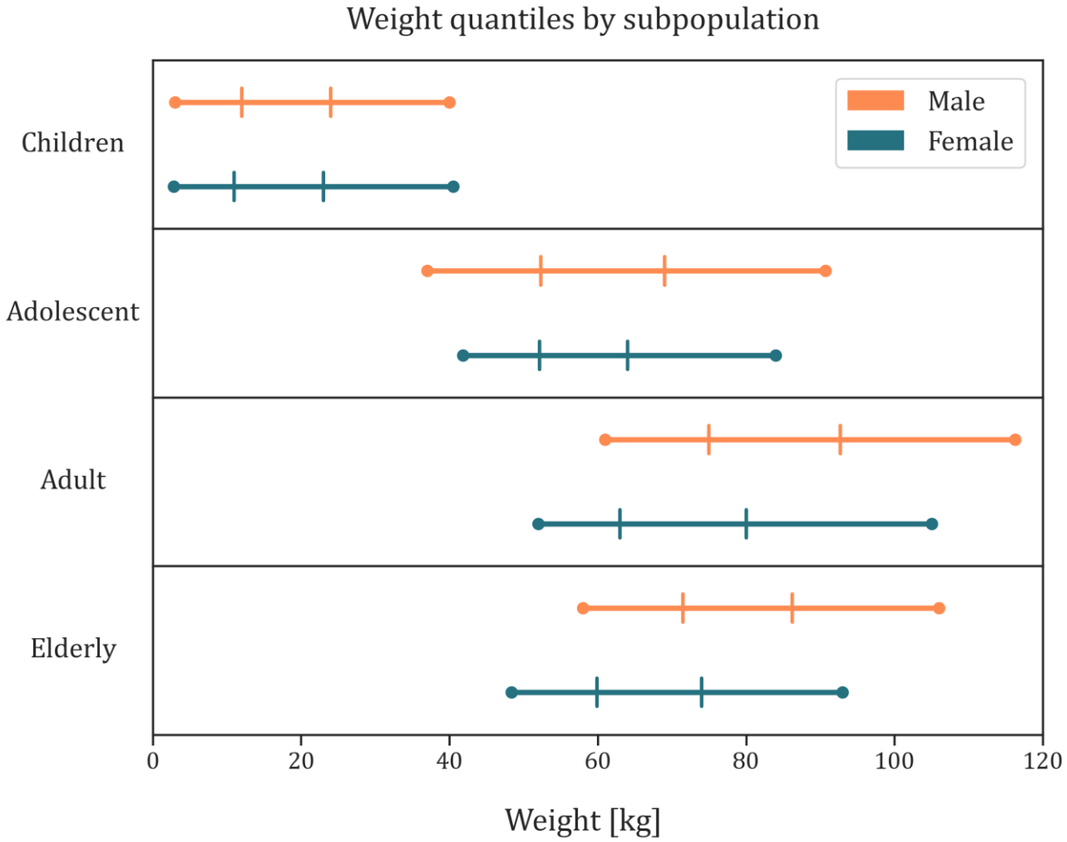


Demographic Features

- Age was binned into 4 categories:
  - Children (birth – 12 years)
  - Adolescent (12 – 21 years)
  - Adults (21 – 65 years)
  - Elderly (65 – 100 years)

- Weight was binned into 3 categories:
  - Low ($Q_{0.05}$ - $Q_{0.33}$)
  - Average ($Q_{0.33}$ - $Q_{0.67}$)
  - High ($Q_{0.67}$ - $Q_{0.95}$)

- Sex was used without any changes



Disproportionality Analysis

- What is DPA?
  - Statistical comparison of observed vs expected drug-reaction reports
  - Used for the early detection of potential adverse drug reactions

- Commonly used metrics:
  - Proportional Reporting Ratio (PRR)
  - Reporting Odds Ratio (ROR)
  - Information Component (IC)

	Reaction	No Reaction
Drug	a	b
No Drug	c	d

$$PRR = \frac{a\,/\,(a+b)}{c\,/\,(c+d)}$$

$$ROR = \frac{a\,/\,b}{c\,/\,d}$$

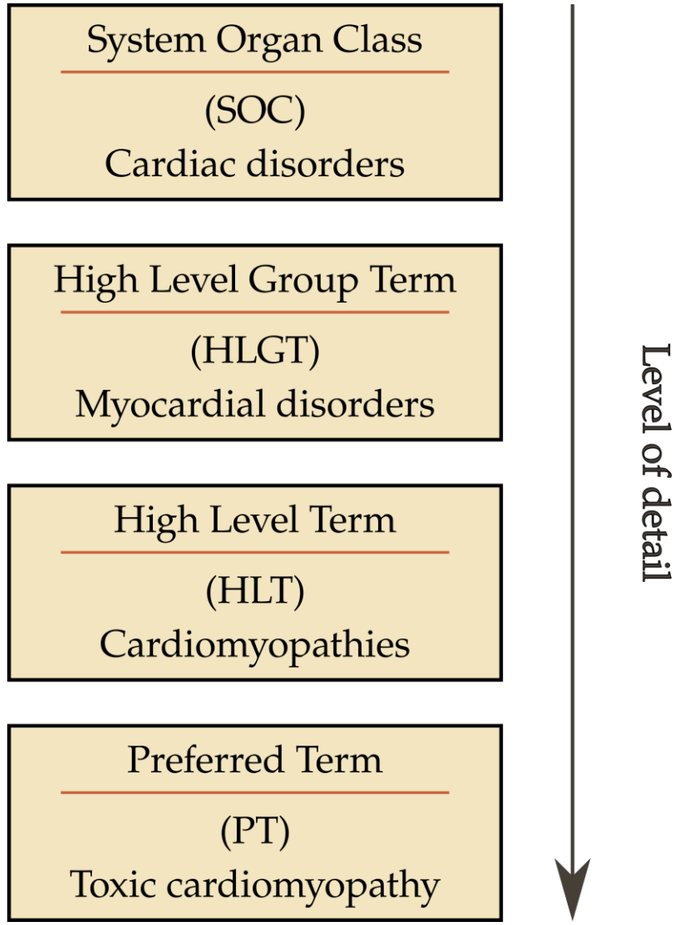
$$IC = \log_2\left(\frac{a+\kappa}{N_{exp}+\kappa}\right)$$

https://github.com/fusarolimichele/pvda/

Cardiotoxicity definition

- MedDRA:
  - Standardized medical terminology by the International Council for Harmonisation
  - Enables grouping of related adverse effects across multiple levels

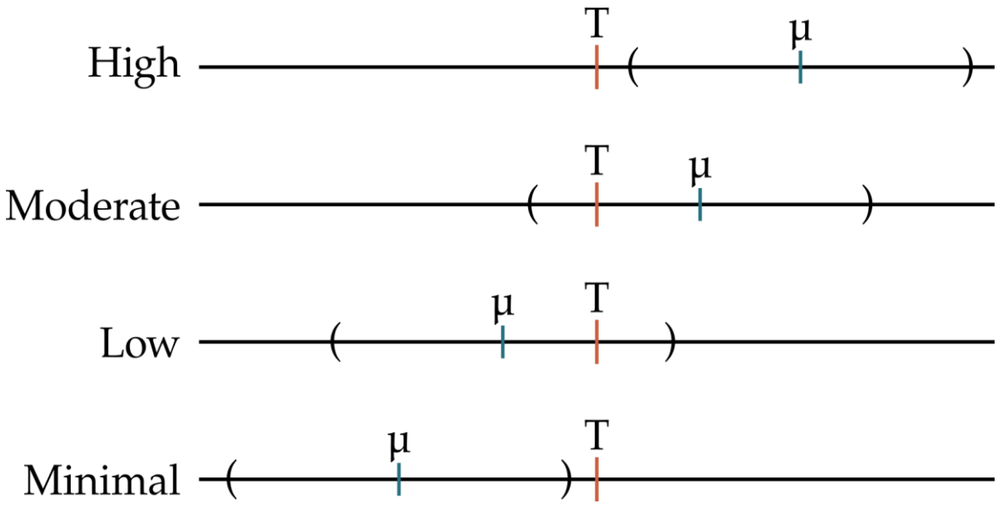
- Label curation steps:
  - Progressivly traversed and reviewed terms at each level
  - Selected terms related to drug-induced cardiotoxicity
  - Removed unrelated PTs (e.g., congenital or infectious diseases, mechanical injuries)



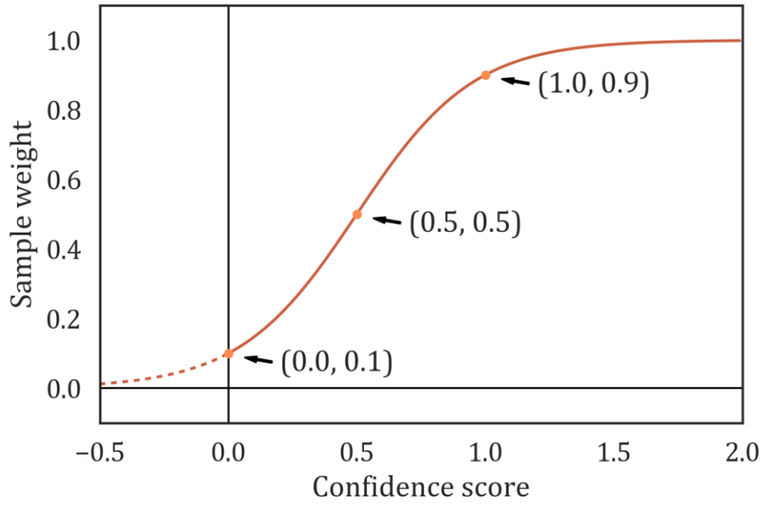
https://doi.org/10.2165/00002018-199920020-00002

Label assignment

- Drugs were classified into 4 risk classes: *High*, *Moderate*, *Low*, *Minimal*
- Thresholds were derived from literature

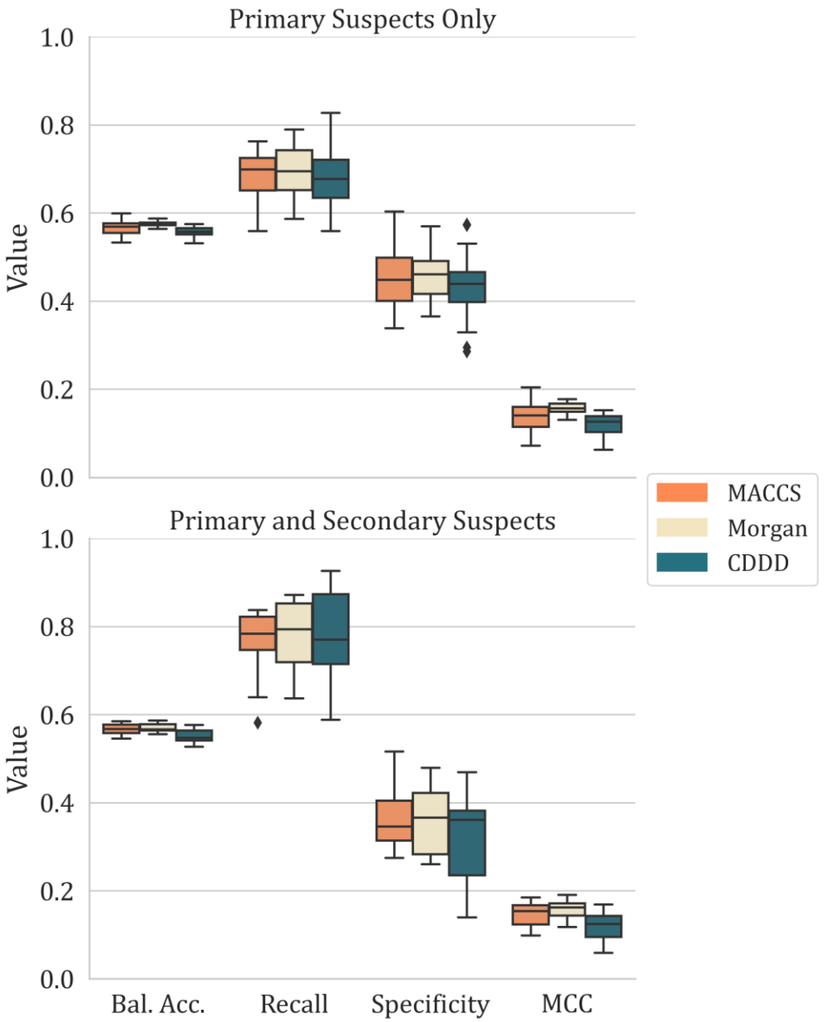


- Confidence score quantifies signal distance from the threshold, normalized by CI width
- Transformed using a modified sigmoid



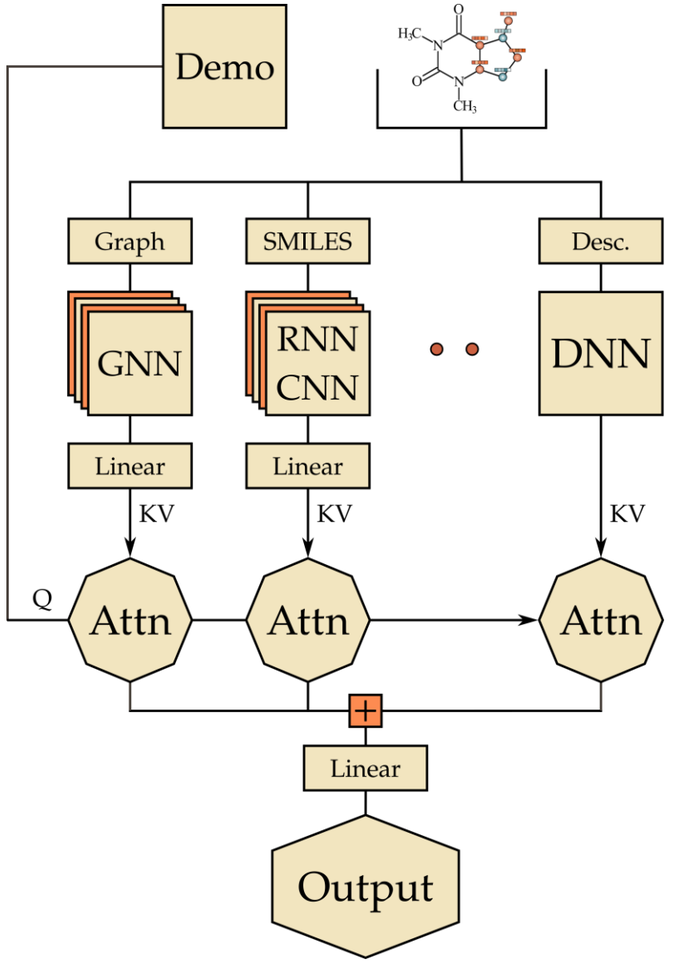
Impact of Design Choices

- ❖ Tested variants:
  - Inclusion criteria (e.g., suspect roles)
  - SMILES standardization and weighting scheme
  - Cardiotoxicity definitions (MedDRA level)
  - DPA metric (PRR, ROR, IC)
- ❖ No major shifts in overall performance
- ❖ Signal remains stable across variants



Conclusions & Future

- Next steps:
  - Large-scale HP search for classical models
  - Architecture search for DL models
- Goal: assess limits of structure-based approaches
- Chemical structure alone may not fully capture the complexity of cardiotoxicity risk prediction
- Future directions:
  - Drug-protein interactions
  - Genetic features



Acknowledgements

Alessandra Roncaglioni

Marina Garcia de Lomana

Francesca Grisoni

Simone Stefano

This study was partially funded by the Horizon Europe funding programme, under the Marie Skłodowska-Curie Actions Doctoral Networks grant agreement “Explainable AI for Molecules - AiChemist” no. 101120466.