sanofi

Predicting the probability of clinical trials success from AI-based approaches using multimodal data

Nils Ternès

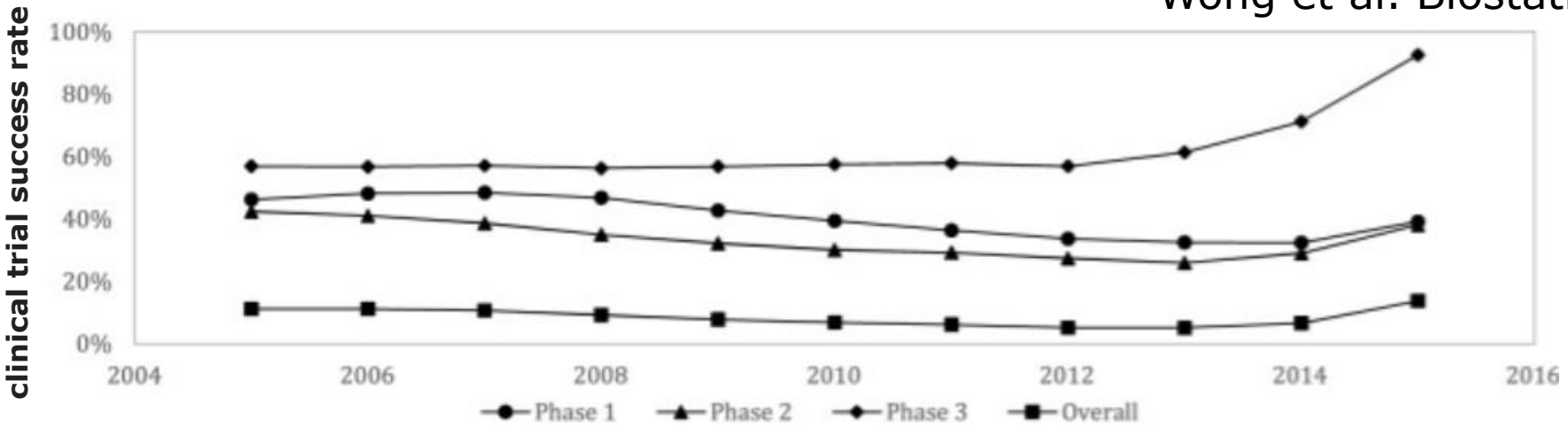
June 11th, 2025

PSI conference

Clinical trials are likely to fail!

- Overall success probability of clinical programs is relatively low (~ 10%)
  - Unnecessary cost and development time
  - Inadequate patient exposure

Wong et al. Biostatistics 2019



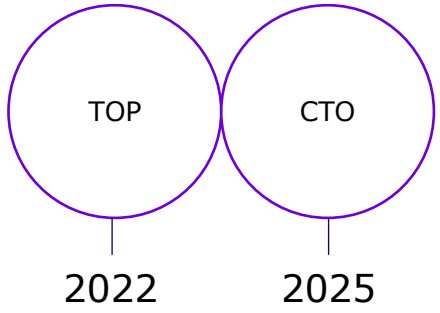
- Predicting probability of clinical trial success (PoS) could support decision making
  - Go/No Go decision, indication prioritization, compound positioning

- Integration of multi-factorial information is key to better predict PoS

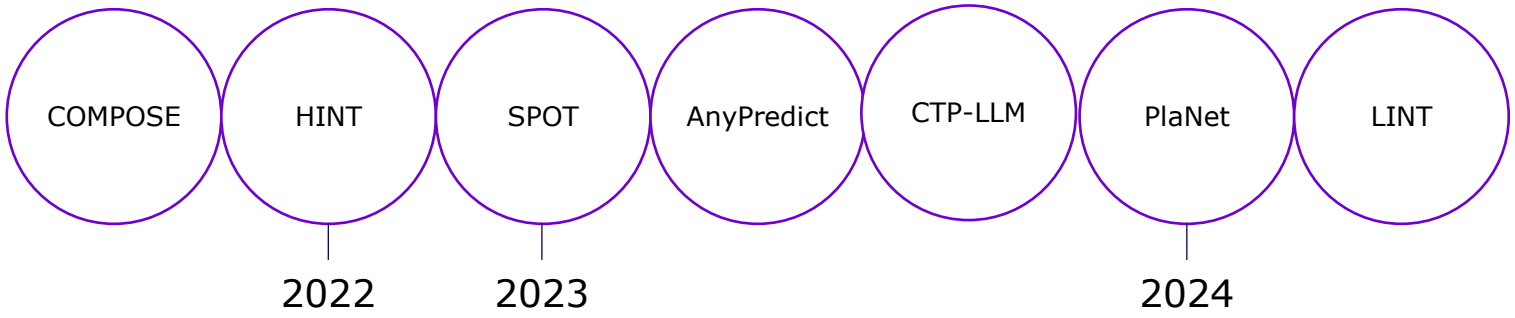
AI-based approaches to predict the PoS

- Growing efforts and researches, especially with the emergence of machine and deep learning (ML/DL) and the large amount of data available

Clinical trial benchmarks



AI-based algorithms



Our approach:

- *State-of-the-art review* of existing benchmarks and pre-trained AI algorithms *with critical thinking*
- Ensure *active connections* with other function groups working in the field and *create a group of experts*
- *Ideation phase* to define our own *fit-for-purpose* benchmark and model, and implement an *end-to-end solution*

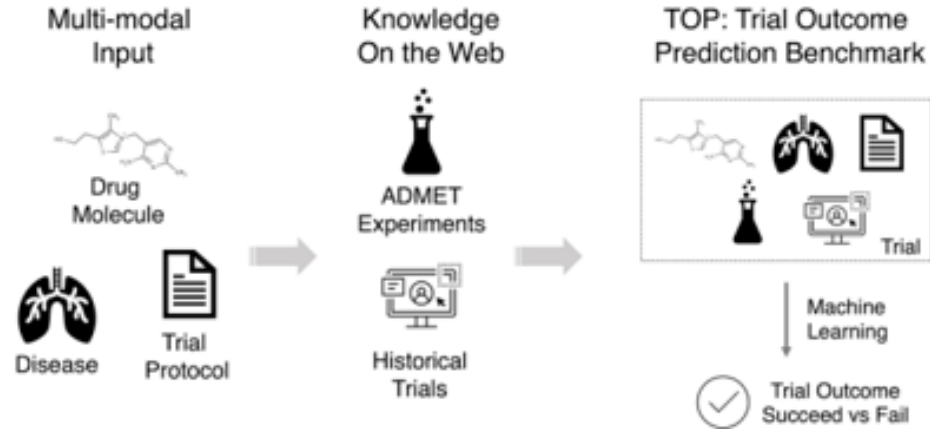
Unprecedent clinical trial benchmarks…

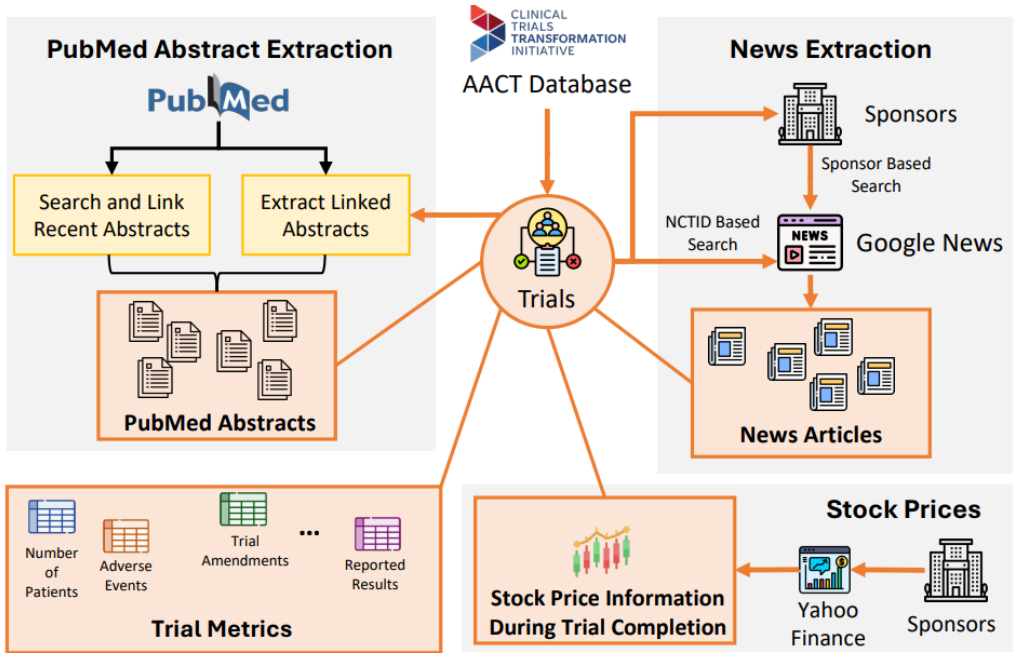
Trial Outcome Prediction (TOP[1])

- ❑ Curation of ~18k clinical trials
- ❑ Multi-modal information (drug molecule, disease, clinical trials)
- ❑ Clinical trial outcome label manually curated

Clinical Trial Outcome (CTO[2])

- ❑ Automatic curation of clinical trial outcome for ~125k clinical trials
- ❑ LLM interpretations of trial publications, phase progression tracking, sentiment analysis from news sources, stock price movements of trial sponsors, and additional trial-related metrics
- ❑ Relatively high concordance with manual curation





1. Fu et al. 2022
2. Gao et al. 2025

… still can be further improved

TOP benchmark is considering only small molecules

- Biological drugs (e.g., monoclonal antibodies) are also included in practice
- Chemical structure based on SMILES is not suitable for biologics ⇨ other databases to be considered

Label of clinical trial outcome remains not fully reliable

- Manual curation remains subjective, e.g. with early phase clinical trials (safety)
- For CTO, weak supervision may introduce label noise
- 'P-value<0.05' reliance is limited

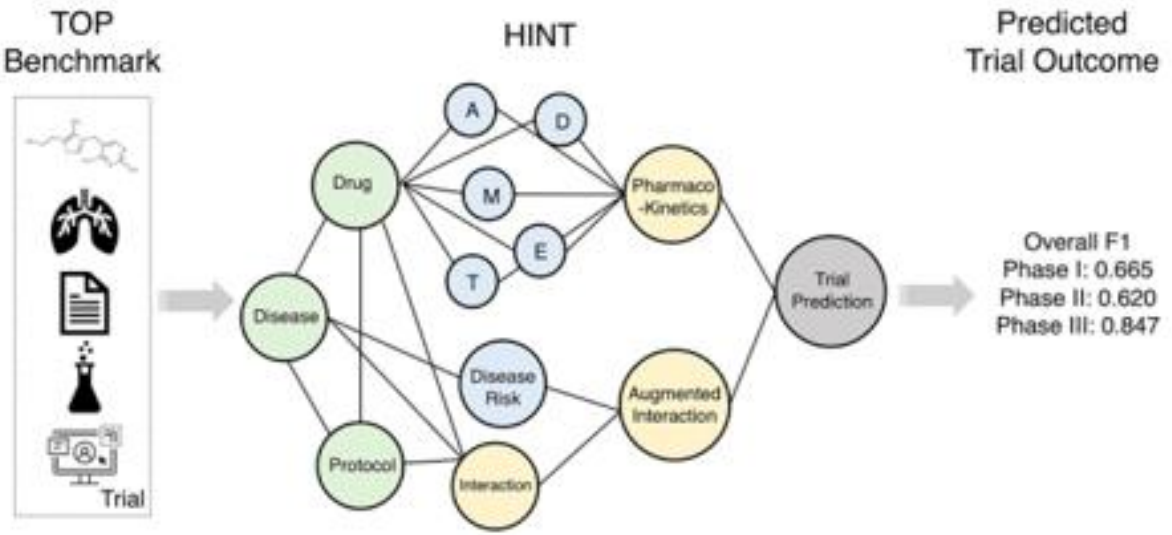
TOP and CTO are considering limited type of information

- Biological data, safety data and further efficacy data could be integrated
- Other benchmarks should be explored, e.g., PlaNet knowledge base

A plenty of existing AI-based algorithms: HINT[1]

End-to-end graph neural network integrating *multi-modal trial data* into a *hierarchical interaction graph*

- Embedding Input Data
- Enriching with Knowledge
- Construct Interaction Graph
- Graph Reasoning via Attention
- Refine Representations (GNN layers)
- Predict Outcomes



Strengths

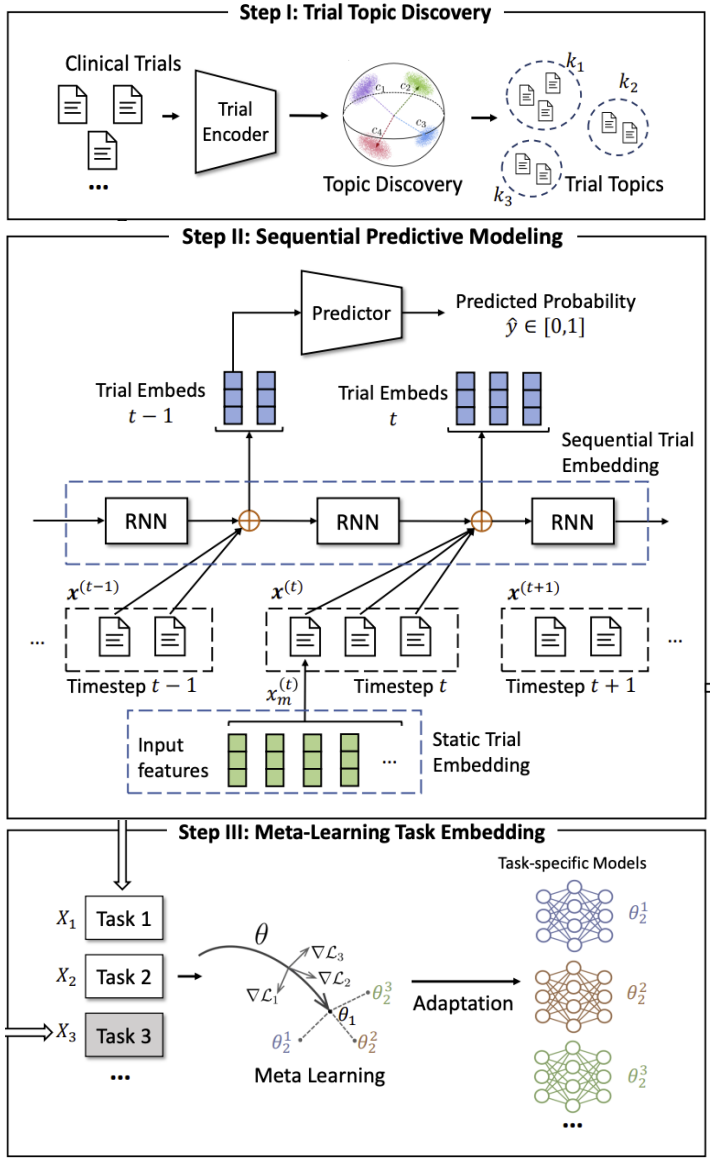
- Multi Model Data Integration
- Interaction Graph (integrate Drug, Disease and protocol data in a single node)

Limitations

- Based on TOP benchmark (not using biomedical knowledge; outcome label may not be reliable)
- Global for all diseases and not consider similarities between trials ⇨ can dilute domain specific patterns
- Struggle to predict outcome for a completely new data which is not well represented in training

A plenty of existing AI-based algorithms: SPOT[1]

Three-stage deep learning framework combining *topic-based clustering*, *temporal sequence modeling* and *meta-learning* to improve prediction accuracy and adaptability



Strengths

- Clusters Trials to capture domain specific patterns
- Meta Learning helps to perform better for disease areas
- Interpretable insights by contextualizing prediction within clusters

Limitations

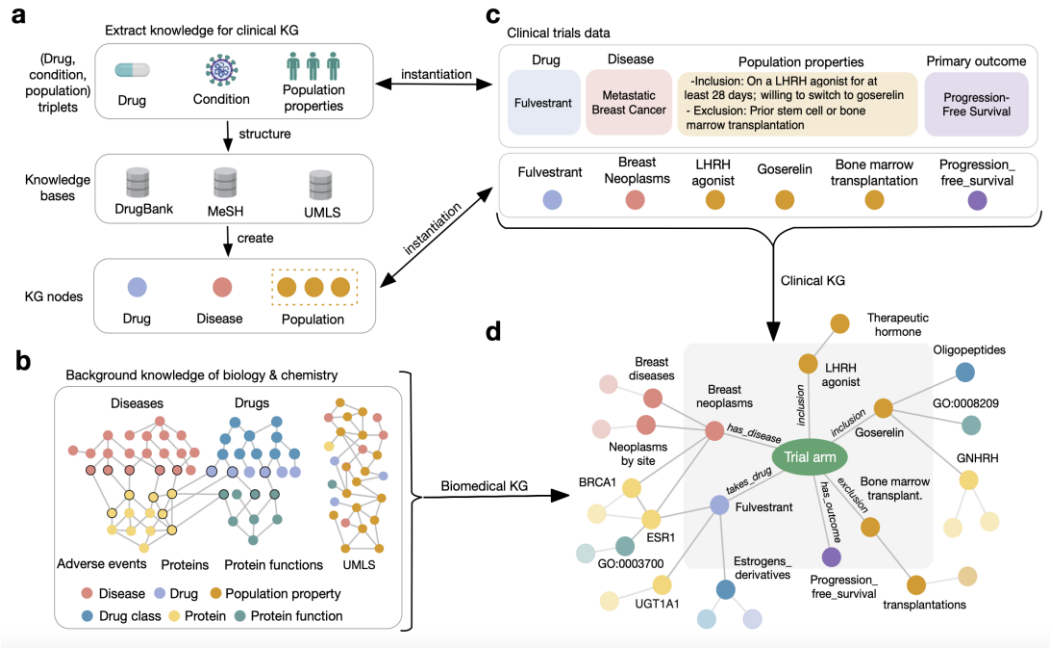
- Based on TOP benchmark (not using biological knowledge; outcome label may not be reliable)
- Additional model complexity ⇨ may require more historical trials to be trained effectively
- May struggle for minimal historical analogy

1. Wang et al. 2023

A plenty of existing AI-based algorithms: PlaNet[1]

Geometric deep-learning framework based on a *knowledge graph* (KG)

- Knowledge Graph Construction
- Self Supervised Pretraining
- Embeddings Extraction
- Fine Tuning for Efficacy and Safety



Strengths

- Integrates together clinical, biological and chemical data with proper edges between all nodes
- Self-Supervised Pretraining for link prediction: which helps in learning rich biomedical representations
- Can easily be finetuned for new drugs/combinations by its prior knowledge of biomedical data
- Easily traceable and more flexible than neural networks

Limitations

- Mostly a static graph reasoning model

1. Brbić et al. 2024

Ideation phase for the development of a new model

Keep the flexibility of a knowledge graph

- Same principle as PlaNet with clinical and biomedical data
- Additional nodes to be added: e.g., for year (sequential info) and full protocol embeddings

Group trials together into disease areas (topics)

- Extract subgraphs of trials from the whole KG
- Whole KG with global learning parameters and topic-specific edge weights

Interpretable model

- Use of back propagation to assess which node contributes the most to the prediction

Conclusion

- Predicting clinical trials success ⇨ hot topic (potential to support clinical decision making)
- Construction of a complete and reliable benchmark is even more important than model development itself
- Interpretable model is crucial to gain the trust of project teams and to be adopted
- Challenging and time-consuming activity: large collaboration and multi-disciplinary group of experts are essential to the project's success

References

- Brbić, M., Yasunaga, M., Agarwal, P. and Leskovec, J., 2024. Predicting drug outcome of population via clinical knowledge graph. *medRxiv*.
- Fu, T., Huang, K., Xiao, C., Glass, L.M. and Sun, J., 2022. Hint: Hierarchical interaction network for clinical-trial-outcome predictions. *Patterns*, *3*(4).
- Karlsson, C., Greasley, P.J., Gustafsson, D. and Wåhlander, K., 2019. Development of human target validation classification that predicts future clinical efficacy. *The Journal of Pharmacology and Experimental Therapeutics*, *368*(2), pp.255-261.
- Wang, Z., Xiao, C. and Sun, J., 2023, September. Spot: Sequential predictive modeling of clinical trial outcome with meta-learning. In *Proceedings of the 14th ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics* (pp. 1-11).
- Wong, C.H., Siah, K.W. and Lo, A.W., 2019. Estimation of clinical trial success rates and related parameters. *Biostatistics*, *20*(2), pp.273-286.