Identifying and Characterising Response in Clinical Trials: Development and Validation of a Machine Learning Approach in Colorectal Cancer

[1]Adam Marcus and [2]Paul Agapow

[1]Department of Computing, Imperial College London [2]Oncology R&D ML&AI, AstraZeneca

Objectives

- Develop an approach using repeated measures to identify and characterise responders in clinical trial data.
- Validate the approach using simulated data.
- Apply the approach to colorectal cancer chemotherapy trial data.

Introduction

Precision medicine promises to transform health care by offering individualised treatments that will dramatically improve clinical outcomes. A necessary prerequisite is the ability to identify subgroups of patients who will respond differently to different therapies. Current approaches have overlooked repeated measures commonly found in clinical trials. By ignoring these measures, a potential opportunity is lost to find better subgroups and an implicit assumption is made that response to treatment is fixed and unable to vary with time.

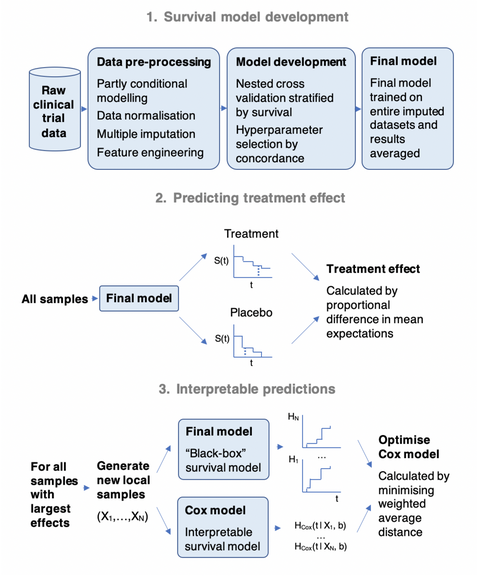

Figure 1:Main stages of the approach

Methods

The approach Consists of several stages:

1. Survival model development – that utilises the concept of partly conditional modelling [1] to develop a high performing predictive model able to utilise temporal data.
2. Predicting treatment effect – that identifies responders to treatment by using the Virtual Twins method [2].
3. Interpretable predictions – that characterises the resulting time-specific responses to treatment using survLIME, an extension of the Local Interpretable Model-agnostic Explanations (LIME) technique to survival data [3].

Validation using simulated data The performance of the approach was evaluated using a simulated 2-arm randomised control trial with 1,000 patients and repeated 100 times. The goal was to identify a region representing responders to treatment of the covariate space. This was assessed using binary classification metrics such as area under the curve (AUC).

Application to clincial trial Data were obtained from Project Data Sphere for four colorectal cancer trials examining the effectiveness of panitumumab to treat metastatic colorectal cancer. The approach was applied to each study independently and the findings were assessed for plausibility.

Results

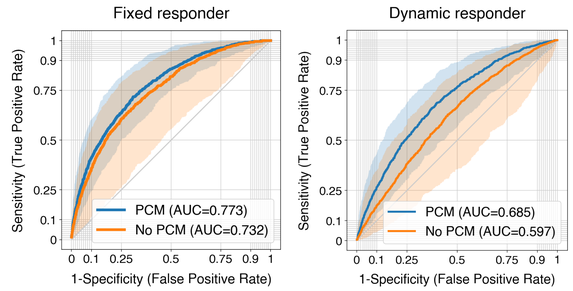

Figure 2:ROC curves for identification performance in responder simulation with 95% confidence bands

The AUC for identifying fixed responders in the simulation study was 0.773. For dynamic responders, partly conditional modelling (PCM) increased the AUC from 0.597 to 0.685.

Important Result

The approach is able to use temporal information within clinical trials to better identify and characterise responders to treatment.

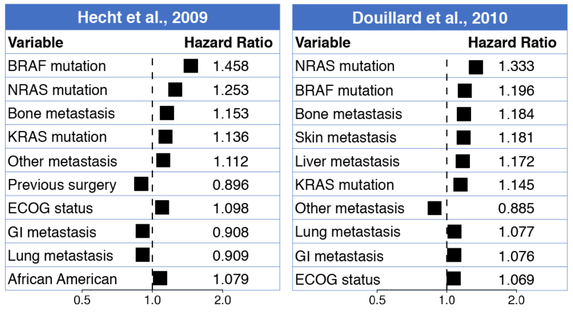

Figure 3:Results of the approach applied to clinical studies

Applying the approach to colorectal cancer trials found genetic mutations (BRAF, NRAS, and KRAF), sites of metastasis (central nervous system, bone, and skin), and ethnic group as important factors influencing response to treatment.

Conclusion

The approach can accommodate a dynamic response to treatment while potentially providing better performance than existing methods in instances of a fixed response to treatment. Applying the approach to colorectal cancer clinical trials led to identified factors that were consistent with the literature. In the future, it is hoped that refining the approach to consider response as continuous not binary and using a tree-based surrogate model to aid interpretability may see it become an essential adjunct to drug development.

References

[1] Yingye Zheng and Patrick J Heagerty. Partly conditional survival models for longitudinal data. *Biometrics*, 61(2):379–391, 2005.

[2] Jared C Foster, Jeremy MG Taylor, and Stephen J Ruberg. Subgroup identification from randomized clinical trial data. *Statistics in medicine*, 30(24):2867–2880, 2011.

[3] Lev V Utkin, Maxim S Kovalev, and Ernest M Kasimov. Survlime-inf: A simplified modification of survlime for explanation of machine learning survival models. *arXiv preprint arXiv:2005.02387*, 2020.

Acknowledgements

Adam Marcus is supported by the UKRI CDT in AI for Healthcare http://ai4health.io (Grant No. P/S023283/1)

Contact Information

- Web: https://github.com/amarcu5
- Email: adam.marcus11@imperial.ac.uk

Imperial College London AI Health AstraZeneca