Characterizing Heterogeneity in Adverse Event Risk Across Immunotherapies in Non-Small Cell Lung Cancer Using Real World Data

Abstract Code:CO63

Gal Shoham Levin[1], Tzviel Frostig[1], Keren Peri-Hanania[1], Oshri Machluf[1], Elad Berkman[1], Raviv Pryluk[1], Walid Shalata[2]
[1]PhaseV Trials Inc., Boston, MA, United States of America, [2]The Legacy Heritage Cancer Center, Soroka Medical Center, Beer Sheva, Israel

Introduction

- **Non-small cell lung cancer** (NSCLC) **is a leading cause of cancer mortality.** The advent of immune checkpoint inhibitors (ICIs) has dramatically improved clinical outcomes. However, substantial unmet need remains and **little is established regarding factors driving the differences in response** to different therapies.
- Despite important clinical benefits, ICIs are associated with a spectrum of **immune-related adverse events** (irAEs). With the exact underlying pathophysiological mechanisms unclear, it is hard to predict which patients will suffer from irAEs and how severe they will be.
- **Real world data** (RWD) collected across different treatments, **can be used to better understand response patterns** and improve treatment guidelines.
- The analyzed RWD cohort consisted of **174 patients with advanced NSCLC**, treated with chemotherapy plus one of two ICI therapies: **Pembrolizumab** ("Pembro", n=71) and a combination of **Ipilimumab and Nivolumab** ("Ipi/Nivo", n=103). The endpoint was overall survival at 24 months.

PhaseV contributed to previous findings in this cohort, reported in Shalata et. al., 2024[1].

Objective

- To **compare the Efficacy** and **Safety** (in terms of irAEs) **of Ipi/Nivo and Pembro for treatment of NSCLC.**
- To evaluate the heterogeneity in treatment effect (HTE) for efficacy or irAE and identify its key driving factors.

Methods

- To address the potential bias due to non-randomized treatment, we used **overlap weighting** for estimating the average treatment effect (ATE).
- **Advanced Causal Machine Learning** (ML) models were used to estimate the **Conditional Average Treatment Effect** (CATE). These methods are particularly suited for RWD settings, inherently accounting for propensity.
- We applied **ABC-D** - ABC (area between curves) Double - **a novel global test for heterogeneity.**
- When significant evidence was found in the global test, we used **SHapley Additive exPlanations (SHAP)** to identify the main factors driving the HTE.

[1] Shalata W, Maimon Rabinovich N, Agbarya A, et al. Efficacy of Pembrolizumab vs. Nivolumab Plus Ipilimumab in Metastatic NSCLC in Relation to PD-L1 and TMB Status. Cancers (Basel). 2024;16(10):1825. Published 2024 May 10. doi:10.3390/cancers16101825

[2] Zhao L, Tian L, Cai T, Claggett B, Wei LJ. Effectively Selecting a Target Population for a Future Comparative Study. Journal of the American Statistical Association. 2013;108(502):527-539. doi:10.1080/01621459.2013.770705

[3] Chernozhukov V, Demirer M, Duflo E, Fernández-Val I. Generic Machine Learning Inference on Heterogeneous Treatment Effects in Randomized Experiments, with an Application to Immunization in India. National Bureau of Economic Research; 2018:w24678. doi:10.3386/w24678

Results: Efficacy

ATE & CATE Estimation

- A **propensity adjustment** was performed, using overlap weights and Logistic Regression with k-fold CV. This reduced the average imbalance from 28% to less than 3% (absolute Standardized Mean Difference).
- After the adjustment, **the efficacy ATE was non significant,** with an estimate of 0.11 months [95% CI: (-2.88, 3.10)] of Restricted Mean Survival Time difference at 24 months, in favor of Ipi/Nivo (Fig 1).
- A **Causal ML model, adapted for time-to-event (TTE) endpoints,** was utilized to estimate the efficacy CATE.

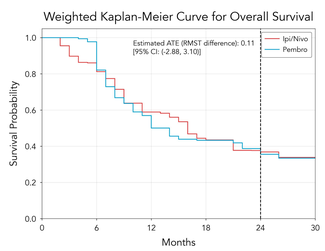

Fig. 1: Overall Survival comparison

Testing for Heterogeneity

- To test whether the HTE estimated by the efficacy CATE model is statistically significant, we adapted the ABC-D test to suit TTE endpoints (see box below).
- **The test yielded a p-value of 0.12**. While this might provide some evidence for heterogeneity, **we cannot reject the null hypothesis of a constant effect.**

Global Heterogeneity Testing

- We performed **a simulation study to compare the performance of different testing procedures** for the presence of HTE captured by a CATE model:
  - **ABC-D** (Area Between Curves - Double): a permutation test, using a diagnostic based on the area between the cumulative ATE curve and the constant ATE line as the test statistic[2].
  - **BLP** (Best Linear Predictor): uses the estimated CATE as a covariate in a linear model, where the resulting coefficient serves as the test statistic[3].
  - **GATES** (Sorted Group Average Treatment Effects): forms heterogeneity groups based on the estimated CATE and uses the variance in ATE between these groups as a test statistic[3].
- **The simulation results show that the ABC-D test offers superior power while tightly controlling the type-I error rate,** especially when compared to asymptotic tests that do not involve permutations (Table 1, Fig 2).

Test	Type	Type-I Error Rate	Power
ABC-D	Perm.	0.0496	0.45
BLP	Perm.	0.0420	0.27
BLP	Asymp.	0.0409	0.14
GATES	Perm.	0.0798	0.12
GATES	Asymp.	0.1132	0.13

Table 1: Type-I error rate & power for 0.05 significance level. Data generated from a linear model with 10 i.i.d. uniform covariates, non trivial propensity and 200 observations.

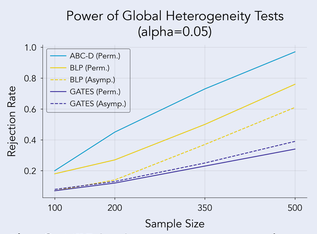

Fig. 2: HTE tests power comparison

Results: Immune-Related Adverse Events

ATE Estimation

- The observed irAE occurrence rates in the cohort were 72% for the Ipi/Nivo group and 51% for the Pembro group.
- **After propensity adjustment, we can conclude that Pembro causes significantly less irAE than Ipi/Nivo:** the irAE ATE estimate was 0.21 absolute risk reduction [95% CI: (0.06, 0.35)] for Pembro.

Testing for Heterogeneity

- A Causal ML model was used to estimate the irAE CATE. When applied to this model, **the ABC-D test yielded a p-value of 0.046** (Fig 3).
- The results indicate **significant heterogeneity in the occurrence of irAE,** explained by the observed baseline covariates.

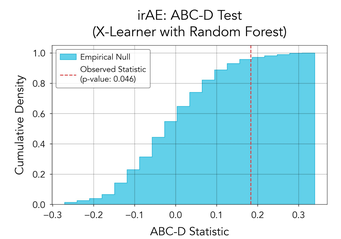

Fig. 3: ABC-D test for irAE HTE

CATE Model Interpretation

- We leveraged **SHapley Additive exPlanations (SHAP)** to further analyze the CATE (Fig 4).
- The SHAP values suggest that smoking and contralateral lung metastases contribute to Pembro's advantage, while factors like lymph node involvement hinder it.

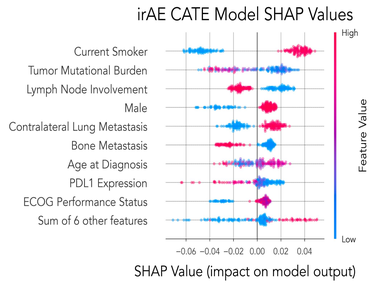

Fig. 4: SHAP values for the irAE CATE model

Summary and Conclusions

- **Our retrospective analysis approach helps to overcome the unique challenges and risks of real-world clinical data, while remaining suitable for randomized controlled trial data.**
- **Through the use of advanced causal ML techniques and innovative HTE testing, we found significant evidence for patient heterogeneity in the occurrence of immune-related adverse events between two key NSCLC treatments.**

For more information, visit us at Booth #306 or contact us at info@phasevtrials.com