Predictive Modeling as an OHDSI Network Study: Treatment Non-Response in NSCLC patients treated with ALK Inhibitors

Aaron Galaznik[1], Greg Klebanov[2], Eldar Allakhverdiiev[2], Yuri Khoma[2] and Christian Reich[3]
[1]Takeda Pharmaceuticals, Cambridge MA USA; [2]Odysseus Data Services, Cambridge MA USA; [3]QuintilesIMS, Cambridge MA USA

Abstract

In this study, we determine the rate of non-response to treatment with ALK inhibitors in patients of Non-Small Cell Lung Cancer (NSCLC). We also developed a predictive algorithm for identifying patients at high risk of treatment non-response to these agents. Both investigations are conducted as an OHDSI network study, i.e. the R developed on a local database is distributed for execution against a suite of four US database assets remotely using the ARACHNE study execution tool.

Introduction

Targeted therapy is now standard-of-care (SOC) for subsets of patients whose tumors harbor oncogenic alterations, exemplified by Non-Small Cell Lung Cancer (NSCLC) mutations. EGFR mutations are the most 'common' de novo mutations and are effectively treated using TKI inhibitors, with patients routinely achieving a significant response rate (ORR 60-70%)[1] and a median progression free survival (mPFS) of 9.5 months [2]. At the same time variety of ALK mutations (amplification, fusions: EML4 -ALK, RANBP2-ALK, NPM-ALK, STRN-ALK, KIF5B-ALK) is still not fully investigated. These patients have significantly lower response rates on TKI therapy.
As an area of unmet need, Takeda is interested in developing novel therapeutics intended to benefit NSCLC patients with less common mutations. To support these efforts, Takeda is interested in characterizing, in NSCLC patients, the real-world rate of treatment non-response to ALK-inhibitors, which are increasingly being used in NSCLC as well. This information is used to develop a predictive modeling algorithm for identifying patients at higher risk of treatment non-response to these agents. The role of the algorithm would be to aid in identifying patients likely to benefit from novel therapeutics through clinical parameters, thereby complementing genomic/biomarker based approaches. Such an approach has potential applications to characterizing population needs, clinical study recruitment, and treatment decision-making for patients with NSCLC.

Study overview

This study developed a risk model and used this model to predict health outcome for patients with locally advanced and metastatic NSCLC, constituting a retrospective, observational, new-user cohort study. We used OHDSI PatientLevelPrediction package to develop risk model and to predict health outcome for patients with locally advanced and metastatic NSCLC, constituting a retrospective, observational, new-user cohort study.

Study Design

All subjects in the database were included who meet the following criteria (fig.1):

Inclusion criteria:

- Exposure to at least 1 of ALK inhibitors (crizotinib, ceritinib or alectinib)
- At least 182 days of observation time prior to the index date
- AT least 18 years of age at the first diagnose date (diagnose is locally advanced or metastatic NSCLC)
- Absence of other cancer during the wash-out period
- At least 2-month gap in treatment after the previous regimen was allowed prior to the index date.

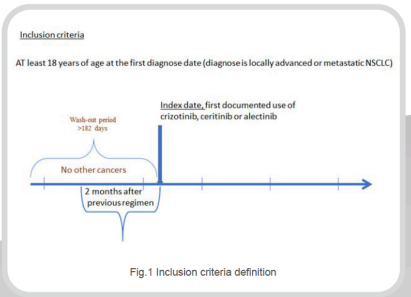


Fig.1 Inclusion criteria definition

Outcome definition

The primary outcome for each treated patient is response or nonresponse to treatment. Patients were assumed to be a non-responder if:

- A new line of therapy started after at least 60 days gap
- Treatment with another standard chemotherapy was initiated
- Surgical or radiotherapeutic procedures were administered

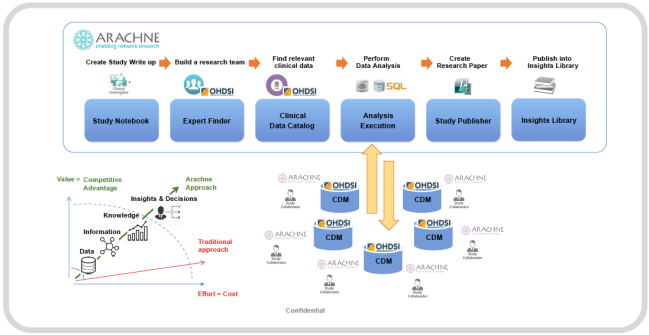
If none of these occurred within a window of observation the patient was considered a responder. As the data set had privacy limitations on death information, mortality couldn't be considered.
Response was observed in three different observations windows (fig 2.):

- at any time after index date (observation window 1)
- ≤ 6 months after the index date (observation window 2)
- ≤ 1 year after the index date (observation window 3)

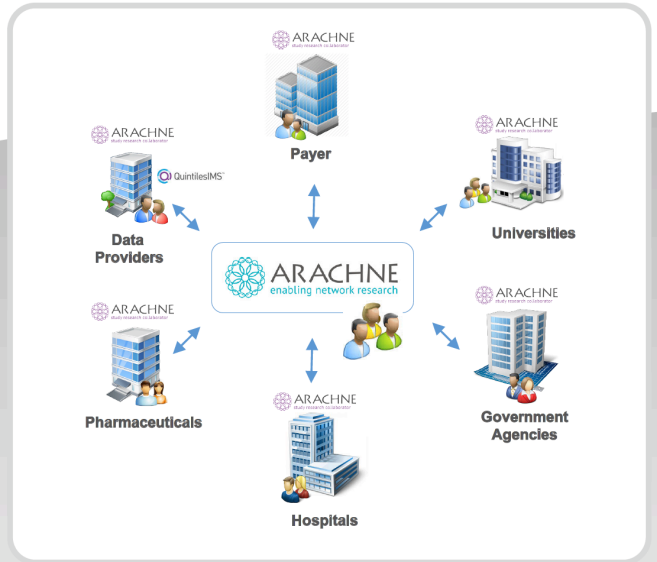
For feasibility and development of the study code, we used a commercial US Claims database.

Conducting Distributed Network Study

We are planning to execute the study and demonstrate results for a number of US data assets derived from payer-based claims, provider-based claims and EHR systems.
To execute a study in a federated data network, several tools need to in place. First, an extensive library of analytical methods need to be available to implement a wide variety of tasks and study designs. Second, a tool is needed that can distribute analytical code to a partner in the data network and can retrieve the analysis results. It is important that the data custodian has full visibility and control in this process, patient privacy is guaranteed and communication between sender and receiver is secured.



The ARACHNE Research Network platform has been developed by Odysseus Data Services (Odysseus) to perform this task and used to conduct this study e.g. collaborate on study protocol between multiple parties, share analytical code and analysis results, annotate and share final insights.



As a next step, we are planning to post the OHDSI call for collaborators to participate in a public network study.

Results

The following shows the preliminary results of generated from the database used for development.
473 patients could be identified in the development database, which were distributed into responders and non-responders as following:

- observation window 1: 187 responders vs 286 non-responders
- observation window 2: 296 responders vs 177 non-responders
- observation window 3: 233 responders vs 240 non-responders

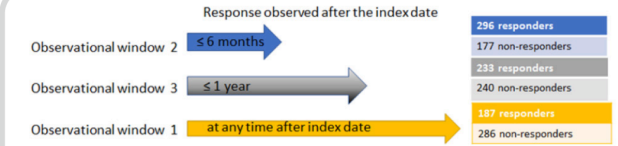


Fig 2. Observational windows definition and patient distribution within them

Models used:

Naive Bayes, Gradient Boosting Machine, Random Forest and Neural Network were trained for each observation window. The preliminary results for the predictive models were as following:

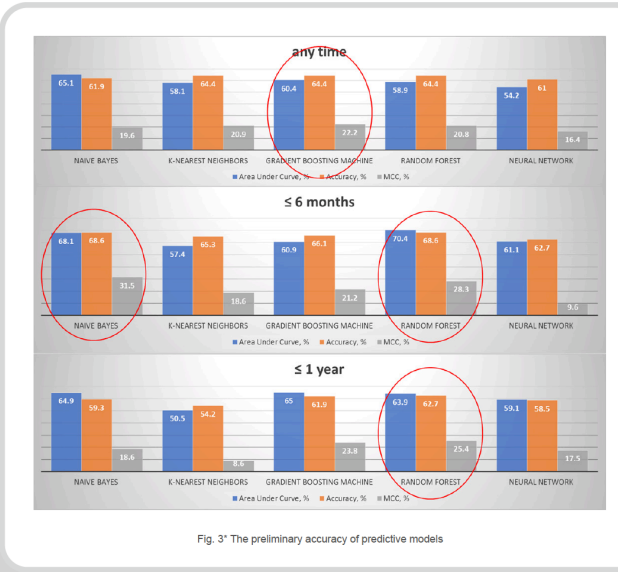


Fig. 3* The preliminary accuracy of predictive models

*Where MCC stands for Matthews Correlation Coefficient. Most accurate approaches are highlighted with Red circles.
For a first window (any time) Gradient Booster Machine works the best, for the less than 6 month window naïve Bayes and Random Forest works better, and for less than 1 year window the best is Random forest based model.

Conclusion

It is possible to predict treatment outcome for NSCLC patients, but the sample size was rather small to reach conclusions. In particular, the fact that different methods generated exactly the same accuracy (fig 3.) can be attributed to them probably generating the same effective model. We therefore want to expand the study using the larger OHDSI network. We also intend to expand our modeling efforts to include other targeted therapies, such as EGFR inhibitors.